

HYDROMAGNETIC MODEL
FOR THE SOLAR GENERAL CIRCULATION

by

Peter Augustus Gilman

B.A., Harvard College
(1962)

M.S., Massachusetts Institute of Technology
(1964)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1966

Signature of Author
Department of Meteorology, April 26, 1966

Certified by
Thesis Supervisor

Accepted by
Chairman, Departmental Committee on Graduate Students

Hydromagnetic Model for the Solar General Circulation

by Peter Augustus Gilman

Submitted to the Department of Meteorology on April 26, 1966
in partial fulfillment of the requirement for the degree of Doctor
of Philosophy

Abstract

Observations of phenomena on the surface of the sun of widely varying time and space scales are reviewed. Some of the physical and mathematical models for the larger, more slowly changing solar features, namely the solar cycle, large scale magnetic regions, and the differential rotation, are also surveyed. Motivated by Ward's sunspot displacement statistics, and Bumba and Howard's synoptic charts of solar magnetograms, we propose that a plausible alternative to these models could result from suitable hydromagnetic generalizations of the equations and models which have proved fruitful in understanding the earth's atmospheric general circulation. On the basis of recent work by Veronis, it is suggested that the granulation and supergranulation scale motions in the sun may produce available potential energy for larger scale motions.

To see the possible consequences of this energy source, the problem is reduced to the study of large scale adiabatic disturbances in an inviscid, perfectly electrically conducting hydromagnetic circumpolar vortex. The equations governing the mechanics, electromagnetics, and thermodynamics of such motions are scaled, and expanded in powers of the Rossby number, observed to be much less than unity. The scaling isolates quasi-heliotropic, hydrostatic, potential energy converting modes on a " θ plane". Due to the horizontal magnetic field, the potential vorticity is no longer conserved. Vertical magnetic fields can be produced, but their feedback on the motions and horizontal fields is excluded. Thus the system in this simple form cannot complete a dynamo cycle.

The scaled equations are perturbed about a steady axially symmetric zonal flow and zonal (toroidal) magnetic field. Changes in the initial state are inferred from products of perturbation quantities. These include the growth in the meridional plane of axially symmetric circulations and magnetic fields. The energetics of the system are examined. Bounds are placed on the complex phase velocities of unstable normal mode disturbances for both continuous and two-layer zonal flows and magnetic fields. For flows with vertical and horizontal shear, these bounds are the same as found by Pedlosky (1964a) for the nonmagnetic case. For the two-layer case, bounds are also placed on the phase velocities of neutral waves.

The stability of two-layer purely baroclinic flow (no horizontal shear) in a uniform zonal magnetic field is studied. The minimum vertical shear needed for instability no longer depends on the β effect or the static stability, but rather is determined by the zonal field strength. Short waves are destabilized by the magnetic field, long waves stabilized. All unstable waves convert available potential energy to kinetic energy of the disturbances, part of which is in turn converted into disturbance magnetic energy. Nonmagnetic changes in the initial state are similar to those of Phillips (1954). A single celled meridional (poloidal) field is produced. Properties of the solutions are compared to solar observations. At gas densities $\sim 10^{-4}$ gm/cm³ a $\sim 35^\circ$ K equator-pole temperature difference is enough to give baroclinically unstable disturbances in a zonal field of 100 gauss. The eddy vertical fields produced from this magnitude zonal field are ~ 1 gauss and wave number of about six around a latitude circle, consistent with the magnetograms.

The stability study is generalized to include a parabolic flow profile in the upper layer. Singularities may be present in the equations even when there are no extrema in the potential vorticity, when the magnetic field is increased beyond a certain strength. Below this strength, power series solutions show that, just as in the nonmagnetic case, horizontal shear renders shorter waves unstable. Reynolds stresses are seen to transport momentum up the gradient, while smaller Maxwell stresses oppose them. This net upgradient momentum transport could maintain the solar differential rotation. The perturbations show a tilt upstream away from the maximum of zonal flow, most pronounced in the vertical fields, suggestive of the tilted patterns in Bumba and Howard's charts.

Suggestions are made for improvements and generalizations of the model, and for other studies.

Thesis Supervisor: Victor P. Starr
Title: Professor of Meteorology

Dedication

To my father and mother, who for so many years have given unselfishly their moral and material resources toward the educational advancement of my brothers, my sister and me.

And to Susie, who is to be mine and I hers.

Acknowledgments

In am privileged to have been one of the many students of Professor Victor P. Starr. To me, as to all of the others, he has given so generously his wisdom and his time. He is a true teacher: always patient and kind to those inexperienced and just starting out; always tolerant and respectful of the student's own point of view; always willing to listen; never condescending or belittling, always encouraging and inspiring.

I have also benefited from countless discussions with many other members of the M.I.T. community and surrounding institutions. It would be impossible to acknowledge all of those who contributed, but I would like to mention some of those who helped to clarify and improve on specific aspects of the thesis, in particular Professors Jule Charney, Edward Lorenz and Joseph Pedlosky, and Drs. Terry Williams, George Veronis, and Fred Ward.

Dr. Ward kindly gave me his unpublished sunspot statistics for five solar cycles in tabular form (he will soon have nine cycles compiled).

Dr. Williams and Mr. Tony Gordon read the manuscript, and caught many errors, for which I am most grateful.

The computational work was done on the IBM 7094 at the M.I.T. computation center. All of the programming was done by two very able programmers, Mrs. Judy Copeland in the early stages, and Miss Judy Roxborough for the later work, particularly Chapter 9.

Miss Isabel Kole did her usual first-rate job in drafting all of the figures, and Miss Marie-Louise Guillot persevered through the typing of the two drafts, producing an excellent final copy.

I wish also to thank the Ford Foundation and the National Aeronautics and Space Administration for fellowship and traineeship support during all of my years at M.I.T., and the National Center for Atmospheric Research for providing an interesting six weeks in Boulder in July and August 1965.

Finally, I thank my office colleagues Mike Wallace and Bob Dickinson for providing a stimulating place in which to work, and many moments of levity.

Table of Contents

Chapter	Page
1. The physical state of the observable sun	1
Introduction	1
Layers of the solar atmosphere	3
Granulation	4
Supergranulation	6
Sunspots	7
Active regions	9
Solar cycle	13
Proper motion of spots: Differential rotation and eddies	14
Large scale magnetic regions	21
Temperature structure	26
2. Brief history of theories of solar phenomena	28
The granulation and supergranulation	28
Sunspots, magnetic regions, and the solar cycle	30
Differential rotation	37
3. A possible alternative approach	44
4. Formulation of the equations governing large scale hydromagnetic flow	56
Physical and scaling assumptions	56
Unscaled equations	64
Scaled equations; their expansion in the Rossby number	69
Boundary conditions and energy conversions	77

5. Equations for perturbations about an initially steady state	84
6. Integral theorems	94
Relation between stresses and symmetric state required for instability	94
Necessary conditions for instability	96
Bounds on phase speeds and growth rates of unstable waves	102
7. Formulation of a two-layer model; integral theorems	107
Perturbation and energy equations	107
Integral theorems	112
8. Stability of two-layer baroclinic flow in a uniform zonal magnetic field	121
Derivation of the eigenvalue equation	121
Stability criteria	124
Structure of disturbances and energy conversions	139
Changes in initial zonally symmetric state	145
Application to the sun	151
9. Stability of two-layer baroclinic flow with horizontal shear in a uniform magnetic field	156
Derivation of the eigenvalue equation	156
Stability criteria, phase velocities, growth rates ..	168
Structure of disturbances; stresses, energy conversions, and changes in initial state	175
Application to the sun	186

190	10. Criticisms and possibilities for further work
197	Bibliography
203	Appendix
211	Biographical note
212	Publications by the author

Tables and Figures

Table 1.	Scaling assumptions	Page 63
Figure 1.	Energy balance for scaled equations	90
2.	Levels for variables in two layer model	108
3.	Schematic stability diagrams for two layer baroclinic flow	128
4.	Numerically evaluated stability diagrams for two-layer baroclinic flow, for various magnetic field strengths	134
5.	Growth rates of unstable disturbances for various channel widths	136
6.	Phase velocities of unstable disturbances for various channel widths	137
7.	Phase structure of most unstable waves	142
8.	Curves of marginal stability for various hori- zontal shears	170
9.	Phase velocities of marginally stable waves for various horizontal shears.....	171
10.	Phase velocities and growth rates of various unstable waves	174
11.	Phase structure of an unstable wave as a func- tion of y	176
12.	Relative amplitudes of the same unstable wave as in Fig. 11.....	179
13.	Eddy heat transports of same unstable wave as in Fig. 11.....	180
14.	Reynolds and Maxwell stresses of same unstable wave as in Fig. 11.....	182
15.	Mixed stresses and poloidal field producing processes of same unstable wave as in Fig. 11.....	185

Appendix Figures

A-1	Mean zonal motion of sunspots	Page 204
A-2	Correlation of longitudinal and latitudinal spot motions	205
A-3	Covariance of longitudinal and latitudinal spot motions	206
A-4	Root-mean square longitudinal and latitudinal spot motions	207
A-5	Mean meridional spot motions	208
A-6	Synoptic chart of large scale magnetic regions at activity maximum	209
A-7	Synoptic chart of large scale magnetic regions at activity minimum	210

1. The physical state of the observable sun

Introduction

The sun is dynamically an extremely active celestial body, and without doubt the most extensively studied. But while most types of solar events have been observed in great detail, theoretical explanations and especially numerical models for these events are in most cases incomplete. This is quite understandable, since the observations show that solar events are highly complex phenomena. It is my hope that in this thesis I can demonstrate how some of the theoretical models which have been useful in understanding the dynamics of the earth's atmosphere may be fruitfully carried over, with suitable generalization to include hydromagnetic effects, to the solar problem. The effort will be directed toward developing equations and models for the very large scale, slowly changing features of the sun, specifically those which have much larger horizontal dimensions than sunspots, and which have observable lifetimes longer than a solar rotation. But before concentrating on the larger more slowly changing features, to the exclusion of others, it is important to set the stage by describing briefly the physical state of the observable sun on a much wider range of time and space scales.

Observations of the sun consist primarily of measurements of its electromagnetic radiation, although particle fluxes are also measured, especially since the advent of artificial satellites. From these radia-

tion measurements not only can one infer approximately the local intensity, temperature, composition and state of ionization, but one can also estimate the local magnetic fields and fluid velocities. The magnetic fields are found by analyzing spectral lines with pronounced Zeeman splitting, while fluid velocities are inferred by two means: from the Doppler shifting of the spectral lines, and from the movement of identifiable solar features. However, in addition to the usual difficulties encountered in spectrographic and photographic measurements of distant objects, there are several other limitations to these techniques. For example, although recent attempts to measure the transverse Zeeman effect have been made, (Severny, 1964) the great bulk of Zeeman splitting measurements have been of the easier to measure longitudinal component, which yields only the line of sight magnetic fields. Thus one obtains different components of the fields at different locations on the solar disk. Similar difficulties arise in the interpretation of Doppler shift measurements, since they, too, yield only the line of sight velocity. In tracking the movement of solar features, it is often difficult to resolve what is proper motion of the feature from its development or decay. Furthermore, it is not clear how well the movement of a solar feature reflects the motion of the gas surrounding it. And finally, in using any of the above mentioned techniques, discerning the mean level to which the measurements refer usually is an uncertain process. But despite these difficulties, it is apparent that from the measurements a great deal may now be said about the magnetic structure and motion which exists in the solar atmosphere.

Layers of the solar atmosphere

Before discussing some of the observations, it is pertinent to state some general properties of the solar atmosphere. The observable region of the sun is conventionally divided up into three rather loosely defined layers: the photosphere, chromosphere, and corona. We shall be describing observations primarily of the innermost layer, the photosphere, but some properties of the higher layers will also be included, since many photospheric features are known to extend to much higher levels.

The photosphere is that layer which extends inward from the solar limb (the edge of the solar disk) to the greatest observable depths. Its thickness is only ~ 400 km, but it is the direct source of nearly all the solar radiation. Although there is some variation according to the height convention used, the temperature at the limb is about 4500°K , increasing to around 8000°K at a depth of 400 km. Gas pressures inferred for this layer range from 10^2 dynes/cm² at the top, to 10^5 dynes/cm² at the bottom, (here radiation pressures are negligible by comparison) while densities are between 10^{-8} and 5×10^{-7} gm/cm³. In the upper 300 km, the optical depth reaches unity, below which it increases much more rapidly (Minnaert, 1953).

The chromosphere is the lowest layer outside the solar disk. Its depth is usually taken as about 20,000 km, through which the temperature increases from a minimum of $\sim 4500^{\circ}\text{K}$ at the bottom to $\sim 10^6$ $^{\circ}\text{K}$ at the top. Above this level is the corona, reaching to several solar radii, in which temperatures are in excess of 10^6 $^{\circ}\text{K}$ (Van de Hulst, 1953).

The principle constituents in photospheric and adjacent layers are fairly well known. By number, approximately 92% of the photospheric gas is atomic hydrogen, 8% helium, with less than 0.1% made up of a large number of heavier elements. In the photosphere, hydrogen and helium are predominantly in the neutral state. Below the photosphere, the hydrogen should be 50% ionized at temperatures $\sim 17,000^{\circ}\text{K}$, while helium should be half singly ionized near $25,000^{\circ}\text{K}$, and half doubly ionized around $80,000^{\circ}\text{K}$ (Schwarzschild, 1958). Both constituents are also primarily ionized above the lowest 6,000 km or so in the chromosphere (Van de Hulst, 1953). The ionization of hydrogen and helium below the photosphere is chiefly responsible for the rapid increase in opacity with depth.

Granulation

Since nuclear energy production in the sun is confined to the core, there must be, in the mean, no net flux divergence of heat at any point in the outer layers. However, locally, over relatively short time periods, deviations from this mean requirement will occur, and intensity fluctuations in the spectra will be observed. These fluctuations are observed to have time periods from a few minutes to many solar rotations, and to range in horizontal scale from a few hundred kilometers (the resolution limit of present instruments) up to the solar radius itself. Most of these fluctuations are obviously associated with motion fields of the same scale; some are also linked to magnetic field configurations of like dimensions.

The smallest and shortest lived of these fluctuations represents a non-steady small scale convection, known as the "granulation". This granulation covers the whole solar globe almost uniformly at all times, giving it a mottled appearance in high resolution photographs. Its existence has been known for several decades, but significantly better pictures of it were obtained around 1958 from Project Stratoscope's balloon borne telescopes. From these photographs, Schwarzschild (1959, 1961) found the average diameter of the granulations to be ~ 700 km, and the average life time (defined as twice the time for the autocorrelation to drop to 1/2) to be ~ 8 minutes. The root-mean-square temperature fluctuations were found to be $\pm 92^{\circ}\text{K}$ (Schwarzschild and Bahng 1961). Summarizing earlier work of several authors, Schwarzschild (1961) described the typical line of sight velocities from Doppler shift measurements as having a minimum of 300 m/sec at optical depth 0.1, rising to 2-3 km/sec a few hundred km above and below that level. In the levels above the minimum velocities, the motions appear to be penetrating the layer where the temperature actually increases with height (gravitationally highly stable).

More recently, Evans and Michard (1962a, b); Evans, Main, Michard and Sevajean (1962); and Leighton, Noyes and Simon (1962) found the fluctuations had a pronounced periodic component, with periods of 200-300 seconds, and with root-mean-square velocities ~ 0.4 km/sec at low levels, ~ 0.8 km/sec at high levels. Their dimensions were somewhat larger than found by Schwarzschild. Little or no evidence was found for magnetic field structures associated with the brightness and velocity fluctuations on this scale.

Supergranulation

In addition to convective motions with horizontal diameters $\sim 10^3$ km, much evidence has recently been found for convection with much larger time and space scales. The doppler shifts for these motions have been observed primarily near the limbs. Near the center of the disk, the more intense granulation scale vertical motions apparently prevent observation of the larger modes. This fact has led most observers to the conclusion that these larger scale motions are quasi horizontal in nature.

The first measurements of these motions were apparently made by Hart (1954, 1956), who detected horizontal velocities ~ 0.3 km/sec with horizontal correlations across $3-7 \times 10^4$ km. More extensive, detailed measurements by Evans and Michard (1962a,b); Leighton, Noyes and Simon (1962), and Simon and Leighton (1964), now characterize the flow as quasi-cellular in nature, with horizontal outflow velocities from the centers of cells ~ 0.5 km/sec. The average diameters of the cells is calculated to be 32,000 km, and the average lifetime 20 hours (Simon and Leighton 1964). In addition, there are some faint indications of rising motion at the cell centers, and sinking at the edges, ~ 0.1 km/sec (ibid.).

Despite the difficulties of observing it near the center of the disk, this "super granulation," as it is now called, appears, like the granulation, to be spread rather uniformly over the globe, requiring about 5,000 cells to cover the observable hemisphere. Unlike the granulation, however, it seems to have associated with it several other phenomena. For example, Simon and Leighton (1964) have shown that the

bright spots in the well known chromospheric "network" of brightness patterns seen in calcium lines tend to occur at the boundaries of the supergranulation cells. They showed further that vertical magnetic fields also seem to be concentrated at these boundaries.

Finally, the spectrographic information seems to indicate (Leighton, Noyes and Simon 1962) that the granulation and supergranulation are not just different parts of a smooth spectrum of scales of motion, but rather represent distinct phenomena, with comparatively little energy present in intermediate wavelengths and periods. The fact that the supergranulation has associated with it magnetic and additional spectral effects, while the ordinary granulation apparently does not, would tend to support this conclusion.

Sunspots

The observational features discussed so far are fairly homogeneously spread over the solar globe. There are many others, however, which show marked preference for certain latitudes and longitudes. Most of these appear to be associated in some way with the solar cycle. The most familiar of these is undoubtedly the sunspot. A sunspot has basically two characteristics which distinguish it from the surrounding granulation and supergranulation. It has a much lower temperature, and it possesses a strong vertical magnetic field. A typical spot may be only 40% as bright as its surroundings, and possess a magnetic field of up to about 3000 gauss (Abetti 1957), compared to background fields of 1-100 gauss (Bumba and Howard 1965a,b). The spot field is not entirely vertical,

as Severny (1964) has shown. Though darker than its surroundings, it is far from quiescent. Turbulence is known to be present (e.g. de Jager 1964). There is also usually a systematic outflow toward the edges of the spot, a typical magnitude being 1 km/sec, known as the Evershed effect. There is evidence for inflow at higher levels, and for rotational motion around the spot, the latter being of smaller magnitude than the Evershed velocities (Abetti, 1957). The rotational motion, at least at high levels where there is inflow, is usually in the sense given by the turning of the inflow due to the Coriolis force.

Sunspots often occur in groups. Usually in such groups one or two spots are much larger than the others. Almost all groups have bipolar magnetic fields, most with fairly well defined lines separating the regions of opposite polarity. Most groups are elongated in shape, with the long axis oriented nearly east-west. The leading end of the group (closest to the west limb or right edge of disk) is almost always closer to the equator than the trailing end. The major spot at the leading end of a group virtually always has opposite polarity in the northern hemisphere as compared to the southern.

Sunspot and sunspot group areas vary considerably. The smallest single spot (sometimes called a "pore") is of order one millionth the area of the solar disk, while the largest group is about 3000 times that in size. More than half the spot areas are between 30 and 300 millionth's (Ward 1965b). Thus the dimensions of a typical spot are comparable to that of the supergranulation. The larger spots usually show a more

complex structure, possessing a dark central region, called the umbra, surrounded by a lighter, annular shaped region known as the penumbra. Spot lifetimes vary widely, with the larger spots and spot groups tending to live longer. Small single spots may last only a day or two, while large spot groups are known to remain in some identifiable form for a few solar rotations.

Active regions

Well developed spots virtually always have associated with them numerous other hydromagnetic and thermal structures. Taken together with the sunspots, the sum total of these features constitute what is known as an "active region" or "center of activity". The principal features observed in an active region are the so-called faculae (and plages)¹, filaments (called prominences when seen projecting from the limb) and flares, as well as fairly strong (~ 100 gauss) magnetic fields exterior to the spots themselves.

The plages are large bright blotchy or stringy areas usually near sunspots, though they may occur in areas which have no spots. In general, their brightness, as seen in Ca II and H α emission lines, is 5-15% greater than the surrounding granulation (Abetti, 1957), representing a temperature excess of $\sim 100^{\circ}\text{K}$. The larger plage regions can extend up

¹Some astrophysicists distinguish faculae as those bright areas observable in white light, seen only near the limb, plages as observable in H α of hydrogen, K of calcium II, seen all over the disk, but this convention is apparently not universal. For example, Kiepenheuer (1953) calls them all faculae. The differences seem immaterial for this work, and so will not be stressed.

to 10^5 km in at least one direction. Their raised appearance when viewed near the limb suggests that they are largely chromospheric phenomena. The plages have much fine structure, which, like the chromospheric network which stretches over the whole globe, in which they are imbedded, is strongly correlated with the supergranulation patterns. The faculae and plage regions in the neighborhood of sunspots have also been shown (Howard 1959, Leighton 1959) to be the seat of quasivertical magnetic fields of magnitude 10-100 gauss. These field regions may also extend up to 10^5 km in one horizontal direction, and their fine structure is well correlated with that of the plage regions. Typically, the plage regions last much longer than the sunspots with which they are associated, appearing before them, and remaining often several rotations after the spots have disappeared. However, the fine structure within the plage regions shows much shorter time variations, on the scale of the supergranulation.

There is another kind of magnetic structure often present in an active region, known as a "filament", or as a "prominence", if it is observed extending from the limb. These are primarily chromospheric phenomena. They consist of twisted or looped magnetic fields and may take on many shapes. Some resemble fan-jet like fountains, some, strands of rope. They may change markedly in a few hours, or they may endure for long periods of time. The average lifetime of a filament is 2-3 rotations. They generally are very thin compared to their length. Typical dimensions are: length, 200,000 km; height, 50,000 km; thickness, 10,000 km (Abetti, 1957). These structures are found near all spot groups, and may form arches between neighboring spot groups. About one third of those spot

groups which contain a spot $>$ 200 millionths the solar disk area show filaments with a vortex like structure in the chromosphere (Kiepenheuer 1953).

Finally, in addition to the relatively long lived faculae and plages, there are brightness fluctuations of very short duration, known as flares. These have dimensions comparable to large sunspots, but lifetimes of only one hour or less. They occur quite often, at the rate of a few hundred per rotation.

From the observations, it is also possible to say a lot about the evolution of a typical active region, although there are, of course, considerable variations in individual cases. A concise chronology of this process has been given by Kiepenheuer (1953), although some more recent work, particularly Bumba and Howard (1965a), has elucidated the early parts of the evolution in greater detail. Basically, it is characterized by a rapid growth and expansion to maximum development, followed by a slower decay.

In summary, a new active region always appears in an area of pre-existing weak background magnetic field and plage, but paradoxically the magnetic flux in the new region does not seem to be an amplification of this background field (Bumba and Howard 1965a). At first the rather round brightness elements in the chromospheric network become elongated and oriented systematically, rather like iron filings in the influence of a magnet. Simultaneously a bright facular region appears, rapidly becoming oriented approximately east-west, but with the western or leading edge somewhat closer to the equator. By the second day, the first

spot is found, close to the west end. The faculae brighten close to the spot, and the chromospheric brightness and magnetic field fine structure expands rapidly (~ 250 m/sec - Bumba and Howard 1965a) with individual strands grouping themselves around the facular region. It very often appears up to this stage that the magnetic polarity of the developing active region is unbalanced locally. By the fifth day a second spot has usually formed near the eastern edge, and several smaller spots have appeared between the leading and following spots. More filaments appear in the chromosphere, and the first flares appear. On the limb, fountain-like spot prominences are evident. The spot group becomes most developed, with large penumbrae, by about the 11th day, but the brightness and extent of the facular and plage region continue to grow, exceeding 150,000 km in horizontal dimension.

During this development of the active region, the spots which are formed appear in the ring of faculae surrounding a supergranule. The fully developed groups show that larger spots appear to be occupying the space originally held by one or two supergranules, while the smaller spots are still in the intersupergranular spaces.

By one rotation, most spots in the group, except the large leading spot which developed first, have disappeared. The facular region has further increased, while flares have subsided. There is usually a well developed, stable filament present on the poleward side of the facular region, connected to the remaining spot, forming an angle of $\sim 40^\circ$ with the meridian. The chromospheric fine structure has aligned itself with the filament. By two rotations, all the spots are gone

and the facular region is decreasing, but the filament continues to grow, while turning toward an east-west alignment. By four rotations the faculae have dissolved completely, while the chromospheric fine structures are relatively unchanged. The filament has reached maximum length and is almost east-west. By five rotations, the fine structure blends into its surroundings. The filament shrinks and appears to migrate poleward at an irregular rate. Finally, beyond six rotations, the filament approaches a large group of filaments at the pole, and is eventually incorporated into it.

It is notable that many of the evolutionary features described above are also reflected in changes in the corona.

The solar cycle

The rate at which new active regions appear, the intensity to which a typical active region develops, and the most favored location for its development, vary in a quasi cyclic way, known as the solar or sunspot cycle. At any given time in the cycle, spots appear in latitude belts of width 15° - 20° in each hemisphere, which never extend poleward of about 40° . At the beginning of a new cycle, spots appear at the higher latitudes. As the cycle progresses, new spots appear preferentially closer and closer to the equator in each hemisphere. Near the end of the cycle, spots will form within 2° - 3° of the equator. The time between successive reoccurrences of spots at high latitudes is variable, but generally falls between 9 and 13 years. Often a new cycle will begin before the previous one is finished.

As measured by sunspot numbers and areas, at the start of a new cycle sunspot activity grows rapidly, reaching a maximum in about three years, after which it decreases at a slower rate until the end of the cycle. These intensity changes are also reflected in the rate at which flares occur, and in the intensity and extent of the faculae and filaments. Of course, these other features, usually being part of an active region, also occur preferentially in lower latitudes as the cycle progresses, but they are not so restricted as the spots.

By studying many successive cycles, some further interesting effects become evident. Differences in intensity of activity between hemispheres are seen in some cycles. But more importantly, the leading spots in spot groups in a new cycle are seen to have the opposite polarity from those in the previous cycle, a phenomenon first discovered by Hale in 1913. Thus a complete cycle in the magnetic field really comprises two sunspot cycles.

Proper motions of spots: Differential rotation and eddies

In addition to their complex structure and variation with the solar cycle, individual sunspots and active regions are also observed to have significant proper motions. Sunspots and sunspot groups show translation both in longitude and latitude. In addition, within sunspot groups, divergence of individual spots from one another is usually observed during early stages of group development, while convergence is often seen as the group decays. In general, all during the lifetime of

the spot group, positions of individual spots and their areas are continually fluctuating. When one attempts to interpret these motions, however, some difficulties immediately arise. For example, one would like to assume that the translation of a spot or spot group traces fairly well the motions of the surrounding gas which have much larger horizontal scale than the spots. However, since these groups have finite size and flow fields of their own, they will probably interact with the flow in which they are imbedded in a more complicated manner. Furthermore, since they are known to have considerable vertical extent, even if they do follow the flow, one must ask to what mean level this flow corresponds. If the larger spots extend deeper into the sun, their movement will presumably reflect the flow at a deeper level.

In addition to these ambiguities, there are further problems in tracking sunspot groups. For example, the Greenwich Observatory, which has made the most extensive measurements of spot motions, proceeds by charting the positions of the "center of gravity" of a group on successive days. (The center of gravity is found by a linear weighting of the sunspot areas according to position, in the same way as one would find the center of gravity of a two-dimensional heterogeneous distribution of mass). This technique will thus produce spurious motions in the direction of developing parts of the group (and away from decaying parts). Ward, who has made by far the most comprehensive statistical analysis of spot motions (Ward, 1964, 1965a,b) has also attempted to estimate the import of some of these obscuring effects. His results,

along with some of the earlier work, will be discussed in the following paragraphs.

The fundamental contribution of all earlier work (reviewed, for example, in Goldberg 1953) was to show that the sun does not rotate at its surface like a solid body, but rather possesses a differential rotation, the angular velocity being a maximum at the equator. This result was first established by Carrington in 1863. Since that time, a great deal of work has been done to refine the result with further sunspot data. Newton and Nunn (1951) used recurrent spot (spots seen on more than one rotation) displacements obtained at the Greenwich Observatory for the six cycles between 1878 and 1944, from which they derived for the rotation the empirical formula (in degrees per day)

$$14.38^{\circ} - 2.77^{\circ} \sin^2 \phi$$

where ϕ is the latitude. Thus the period is about 25 days at the equator, 27 days at 35° latitude. This result agreed with Carrington's and other earlier work very well. Newton and Nunn also found no measurable variation in the rotation with alternate cycles, or between hemispheres. There is evidence from trackings of filaments that the rotation continues to decrease toward higher latitudes beyond the sunspot zones, with the period extending to perhaps 30 days.

There have also been numerous attempts to measure the differential rotation by spectroscopic means, (again see Goldberg 1953 for review) whose results confirm its existence, but show poorer agreement

with each other. This should be expected, however, since the spectroscopic determinations represent averages over much shorter periods of time. In any case, systematic errors of several kinds seem to make up the difference compared with the spot determinations. That there is fair agreement between spectroscopic and spot measurements of the differential rotation indicates that, at least for the mean longitudinal component, the spots do in fact trace the motion rather well.

Some attempts were made, both by tracking other solar features such as filaments and by observing spectra coming from different levels, to determine the rotation as a function of height. These are summarized, for example, in de Jager (1959), and seem to indicate a tendency for faster rotation rates in the chromosphere and corona but the magnitude is not large enough to be free of doubt on statistical grounds. Ward's results may shed further light on this question.

In addition to the differential rotation, some calculations have been made, primarily from spot displacements, attempting to show the existence of a mean meridional motion (e.g. Tuominen, 1955; Klyakotkov 1958; Chistyakov, 1960). However, as discussed by Ward (1964, 1965a), and in later paragraphs here, none of these results appear to have statistical significance.

As previously mentioned, the most extensive statistical analysis of sunspot displacements has been made by Ward (1964, 1965a,b). To date, he has completed analysis of five cycles (private communication) including the period 1905-1954. All of his results are based on observations made

at the Greenwich Observatory. Newton and Nunn used only recurrent single spots, which are generally spots only of the largest area, representing a total of less than one thousand observations. By comparison Ward has used spots and groups of all areas and lifetimes, which, for the five cycles completed so far, represent a collection of almost fifty thousand observations. His major results are presented in graphical form in the Appendix. Ward's statistics easily reconfirm the shape of the differential rotation as found by Newton and Nunn and others previously, (Figure A-1). However, the rotation obtained from all spots by Ward is at every latitude somewhat larger than that found from the recurrent spots alone. Furthermore, grouping the spots into four categories according to area yielded, in general, a slower rotation rate for the larger spots at all latitudes except very near the equator. As Ward points out, then, if one were to assume that the larger spots extended further into the sun, these results would indicate a rotation rate increasing with height, in agreement with the tendency seen in other solar features. The magnitude of this effect, however, is not so large as to render it conclusive.

Although this result raises interesting possibilities, it is far overshadowed in immediate importance by another of Ward's findings. This is the systematic correlation of longitudinal and latitudinal spot displacements. Ward found that in each 5 degree latitude belt in each hemisphere, the correlation was such that movement of the spots toward the west limb (in the direction of the rotation) tended to be accompanied by movement toward the equator (Figure A-2). The amount of this cor-

relation is seen to have a maximum of about 0.32 in the 25 - 30° latitude belt, dropping off to 0.08 in the 0 - 5° belt, and also decreasing slightly toward higher latitudes. The 5% limits, as indicated in Figure A-2, clearly show that this correlation is statistically highly significant at all latitudes.

If one interprets the spot displacements as reasonable tracers of the flow in which the spots are imbedded, then this correlation has important physical consequences. The covariance of the spot motions (Figure A-3), which is just the correlation multiplied by the standard deviations, represents the horizontal eddy momentum flux, and with the correlation taking the sign found by Ward, this flux would be directed toward the equator in each hemisphere. Thus, as Ward points out, and as discussed further by Starr and Gilman (1965a,b), this flux could be the mechanism by which the differential rotation is maintained against frictional and magnetic torques. The magnitude of the flux is certainly large enough, since, if it were cut off, and all other torques remained, solid body rotation would be achieved in just a few rotations.

There are, however, some objections to this interpretation of the correlation. For example, Leighton (unpublished manuscript) argues that since almost all spot groups' axes are tilted such that the leading end is toward the equator, even random fluctuations in the spot areas and positions within the group will produce such a correlation. Leighton estimates that this effect could explain a large part or all of the covariance. Ward (unpublished; also Ward 1965b) agrees that the effect is undoubtedly present but feels Leighton overestimates its magnitude. He notes

that single spot motions are correlated with the same sign, but somewhat smaller magnitude than all spots together. Ward takes the difference between correlations as a measure of the effect Leighton proposes, and concludes that it amounts to only 20-30% of the total.

Just as with the mean longitudinal motion, Ward finds (Figure A-4), that the root mean square longitudinal and latitudinal displacements vary inversely with the area. For all four area categories the longitudinal daily displacements are about double those in latitude, being 0.8 deg/day and 0.4 deg/day, respectively, for all spots together.

Finally, Ward calculated the mean latitudinal displacements (Figure A-5) and found that even for the five cycles together, only the value in the $0-5^{\circ}$ latitude belt is statistically different from zero (equatorward) at the 5% level. Thus Tuominen's (1955) values, being based on a much smaller data sample, must be considered highly doubtful.

The basic picture presented by Ward's results, then, is one in which the general circulation is not simply made up of symmetric motions, but rather one in which there are also large quasi horizontal eddies, playing a fundamental part in the maintenance of the average symmetric flow, i.e., the differential rotation. If one follows this view, then, the instantaneous zonal flow could look quite different from the average, with relatively narrow meandering currents present. These currents, to provide the proper momentum flux could possess "tilted troughs" analogous to those in the upper level westerlies, first discussed by Starr (1948). This intriguing similarity to the dynamics of the earth's atmospheric general circulation is the motivation for a considerable part of the theoretical studies undertaken in later sections.

Large scale magnetic regions

As discussed earlier, the active regions always grow in locations of pre-existing "background" fields and plages. These background fields are also of interest in themselves, since they too exhibit dynamic properties.

The first reports of such a "general" solar magnetic field were made by Hale around the time he discovered sunspot fields. He claimed to have found a field strength 50 gauss near the poles. However, later measurements by others failed to confirm his result. At that time, instrument resolution allowed only fields greater than ~ 30 gauss to be detected, and conclusive evidence for a general field had to await the development of the solar magnetograph by H.D. and H.W. Babcock in the early 1950's. They reported (Babcock and Babcock, 1955) a unipolar field ~ 1 gauss poleward of about 55° . They also found large regions of mixed polarity at lower latitudes, far more numerous than sunspots. These areas, which they named Bipolar Magnetic Regions (BMR's), also had fields typically of a few gauss, with regions of one polarity having horizontal dimensions much larger than spots or spot groups. The magnetograms also suggested that these fields are primarily radial, since the line of sight field strengths drop off near the edge of the disk.

In 1959, H.D. Babcock reported that he had observed a reversal of polarity in the weak polar field. This reversal took place near the maximum of solar activity, with the southern pole reversing in the middle of 1957, the northern waiting until November 1958. Since then, several authors have shown that the polar fields have retained their new polarities, but that often the field is difficult to detect.

Recently Bumba and Howard, (1965b), in a companion paper to their work on active regions, published an extensive synoptic study of the large scale solar fields. Their work was done with an improved magnetograph, giving 23" angular resolution as opposed to 70" for the original Babcock instrument (the whole solar disk subtends an angle of about 30'). The synoptic charts made were "global", each representing a whole solar rotation. The field value, which represents the average field over the area subtended by the aperture, had a threshold of ± 2 gauss. Isogauss contours were drawn for 2 gauss, and in varying increments, up to 25 gauss. Two typical such charts, one representing the fields near the maximum of solar activity, the other near activity minimum, are reproduced in the appendix (Figures A-7, A-8).

Several general features of these charts are immediately evident. Some of them have been discussed in Starr and Gilman (1965b). Just as with active regions and sunspots, the area covered by observable fields (> 2 gauss) varies according to the solar cycle. Near the maximum activity of the cycle studied (the year 1959), about half the disk was occupied by integrated fields greater than 2 gauss. By contrast, near activity minimum (around 1964) only about 5-10% of the disk showed such fields. The patterns near minimum are much more fragmented than near maximum, but the regions containing a single polarity, at least at low latitudes, appear to be comparable in size.

Equatorward of about 40° , the fields are generally bipolar, with the polarity varying in a quasi-regular way along a latitude circle.

One might assign a typical longitudinal wave number of 6 or so to these variations. Although they are most evident in the charts drawn near activity maximum, some indications of their presence is seen also in charts from near minimum.

Poleward of 40° , the fields, though still possessing large scale variations, are almost entirely unipolar, of opposite sign in the two hemispheres. This high latitude effect is clearly discernible only fairly near activity maximum, as one would expect, since radial fields are more difficult to see near the poles than in lower latitudes. These unipolar regions are usually quite elongated, sometimes stretching for 100° longitude or more. Many of the unipolar regions seem to be extensions of a particular region of like polarity in the bipolar region nearer the equator, while others seem to be connected to several such regions. Each unipolar magnetic region which could be observed clearly showed that the region leading it of opposite polarity had the same general shape, but with the poleward extension much weaker in strength. Often a quiescent filament separated the two regions.

Virtually all of the magnetic regions, regardless of the phase of solar activity, show a characteristic tilted shape, as a function of latitude, upstream and away from the equator. This tilting occurs both in the bipolar regions at low latitudes, and the unipolar regions in high latitudes, and tends to follow the parabolic shape of the differential rotation.

Another feature, which is peculiar to this solar cycle, although

it has happened before, is the much greater area covered by magnetic regions in the northern than in the southern hemispheres. This effect is very clear near activity maximum, but is much less pronounced near minimum. It reflects the greater solar activity of the northern hemisphere observed for the cycle.

Finally, just as with smaller, active regions, the areas occupied by the stronger large-scale fields, both at low and high latitudes, are also usually the seat of plages.

The way in which some of the features typically found on the synoptic charts evolve was also studied by Bumba and Howard, by means of close examination of charts for successive rotations. They found that many large features change greatly over a few rotations, but the main characteristics of the patterns could usually be traced for about a year (~ 13 rotations). Just as with the active regions, discussed earlier, particular features in the large scale fields developed more rapidly than they decayed. Also, generally speaking, the larger was the feature, the longer was its lifetime. They attributed many of the movements and expansion of features to the "stretching and twisting" of the differential rotation, but many could not be so related. In some cases, the change of features was too fast to be ascribed to the differential rotation, and occasionally features moved across the equator. In general, addition and cancellation of vertical fields was taking place. That is, fields of like polarity often approached each other and merged, while fields of opposite polarity did not appear to move together. Occasionally certain certain areas would entirely lose observable fields.

The unipolar magnetic regions at higher latitudes seemed to develop from "complexes of activity" or several neighboring active regions. The trailing parts of these regions appeared to contribute the major part of the flux. Lifetimes of the unipolar magnetic regions were difficult to distinguish during the years of high activity, but during less active years, the unipolar regions typically took 2 or 3 rotations to form, whereafter they could be followed for six or seven rotations. The tails of these unipolar regions typically drifted toward the pole. This process seemed to be responsible for the production and subsequent reversal of the polar field observed by the Babcocks. However, the amount of flux reaching the pole appeared to be too large, suggesting that some other mechanisms must also be at work near the pole.

In concluding our description of these large scale magnetic regions, it is interesting to point out some recent work by Wilcox and Ness (1964, 1965), which indicates that these features have extensions not just into the chromosphere and corona but probably also into the interplanetary magnetic field at least as far out as the earth itself.

Just as with Ward's results for spot motions, several of the properties of the magnetic fields just described should strike familiar chords for the dynamic meteorologist. In particular, the large horizontal scale of the fields, their tilted structure, and relatively long life compared to the rotation period, all suggest they might be hydromagnetic manifestations of the "tilted trough" often seen in the upper level jet streams in the earth's atmosphere and suggested by Ward's spot correlations. This possibility has been discussed in some detail by Starr and Gilman (1965b),

and will be examined more fully in Chapter 3. Added to Ward's results, this interpretation of the field measurements comprises the second major motivation for the theoretical calculations made in later sections.

Temperature structure

Most calculations of the temperature structure in the solar atmosphere are based on observations of the so-called "limb darkening". An observer looking at the sun receives radiation from shallower layers at the limb than at the center of the disk, since the optical distance the light must travel from the same geometrical depth is greater at the limb. Since the brightness is less at the limb, the temperature must increase with depth. Also, since the light is redder at the limb, the opacity of these layers must be greater for shorter wavelengths. To actually calculate a temperature profile with depth, however, requires additional assumptions about the composition and state of ionization in order to estimate the opacity as a function of depth. A typical calculation of this type is that of Pagel (1961). From this calculation and many others, it has been deduced, as stated earlier, that a temperature minimum exists in the upper photosphere. Pagel, for example, finds it to lie at an optical depth of 7×10^{-3} at a wavelength of 6190 \AA , with an effective temperature of $4300 \text{ }^\circ\text{K} \pm 100 \text{ }^\circ\text{K}$. The vertical temperature gradients from such calculations are on the order of $1 \text{ }^\circ\text{K}/\text{km}$.

In addition to these vertical temperature gradients, it is of considerable interest to know whether larger scale horizontal temperature

gradients exist in photospheric layers. The presence of large bright plages in the vicinity of active regions suggests that such gradients may exist, at least in the chromosphere. There is also the possibility of temperature differences between equator and pole, and in principle, such differences could be estimated by comparing the limb darkening at different latitudes. This has been done using various techniques, by many authors, but, unfortunately, the results are not conclusive. Measurements indicating poles both hotter and colder than the equator have been reported, as well as several failures to find any measurable differences. Beckers (1960, 1962) has reviewed most of the earlier work and conducted further measurements of his own, on certain temperature sensitive Fraunhofer lines. He found the pole warmer than the equator by about 60°K , with a temperature minimum around 45° latitude. By plotting results of earlier work from other authors and his own as a function of the solar cycle, he concluded that the equator to pole difference might be a function of the solar cycle, with the poles much colder than the equator near activity minimum, and slightly warmer near activity maximum. However, recent measurements by Mulders and Slaughter (1965) at the new activity minimum have failed to confirm Beckers' suggestion, as they could find no observable difference.

At present, then, it is really not known for sure whether such equator to pole differences exist. However, present instruments cannot do better than about $\pm 20^{\circ}\text{K}$, so temperature differences less than this amount cannot be precluded. Beckers (private communication) has stated that increased accuracy in these measurements by photoelectric techniques should be forthcoming soon. In any case, these measurements do not preclude such temperature differences existing at deeper levels in the sun. This possibility will be discussed further in Chapter 3.

2. Brief history of theories of solar phenomena

In the previous chapter, an admittedly bewildering variety of solar phenomena were briefly described. From even this cursory sketch, it is obvious that all theoretical models of necessity have been, and will continue to be, highly oversimplified, both as regards the kinds of physical processes included, and the time and space scales covered. Let us now examine some of these models. Just as with the observational material, the emphasis will be directed toward theories for the larger scale, longer lived phenomena, from which chapter 3 will attempt to present a possible alternative approach. Discussion of theories relevant to the smaller scale phenomena, which form a large subject in themselves, will be limited to a few rather general remarks.

The granulation and supergranulation

Theoretical models for the granulation and supergranulation are seriously complicated over studies of convection in a liquid by several additional physical factors present on the sun. Among the obvious are rotation, magnetic fields, compressibility, radiative transfer, and ionization. In addition, of course, the solar convection has finite amplitude and is turbulent. Without attempting to review the literature on these complicating effects here, let it suffice to say that work is being done to elucidate the influence of each of them on the convective process.

It is generally agreed that the existence of the convection layer is due to the high opacity of these layers, which arises from the partial

ionization of hydrogen just below the photosphere, and of helium of somewhat deeper levels. For radiative transfer to carry all the energy produced in the core through these layers would require superadiabatic temperature gradients. These layers would then be unstable to perturbations, and convective overturnings would develop. Throughout most of the convective layer the gas is so opaque to radiation that almost all the heat transport outward is by the convection. The thickness of this convective layer is, of course, uncertain, but on the basis of stellar structure theory for stars of solar mass and luminosity, it is generally believed to be one to two tenths of the solar radius. Considering reasonable estimates of temperatures in these layers, this thickness represents a great many scale heights.

As mentioned earlier, the observations seem to indicate that the granulation and supergranulation are distinct phenomena. Theories for the supergranulation are only in the suggestion stage, the principal one being that because of their large size, they are due to the ionization of helium deeper down (Simon and Leighton 1964). Theories of the granulation are split between those which assume the observed granules are actual convective cells, (Schwarzschild 1961) and those which assume they are acoustic-gravity waves propagating from the convective layer through a gravitationally stable layer above (Whitney 1963).

Some work has been done on the effect of the granulation and supergranulation on magnetic fields, largely by Parker (1963a,b) and Clark (1965). Parker showed that those eddy motions with lifetimes shorter than, or comparable to, the advective time scale, tended to give r.m.s.

fields amplified over the general fields, but by a factor less than ten. The fields thus produced fall short of energy equipartition with the motion. Somewhat larger scale eddy motions, with lifetimes longer than the advective scale but shorter than the magnetic diffusion time, tended to concentrate fields into sheets or filaments. The supergranulation is probably of this type, concentrating the fields at the edges of the cells. Here equipartition of energy is approached in the areas of concentrated fields. Finally eddy motions with longer lifetimes than the magnetic diffusion time, tended to exclude the field from the region of motion. Clark (1965) concentrated on the effect of the supergranulation on vertical magnetic fields, finding the convective modes which were capable of maintaining steady vertical fields at the cell boundaries.

Sunspots, magnetic regions, and the solar cycle

Sunspot formation is conveniently discussed in the context of the solar cycle, the former being but one manifestation of this quasi-periodic redistribution of magnetic flux. The great majority of solar astrophysicists now believe that the dynamics of this redistribution is confined to a relatively thin outer shell of the sun, probably not extending deeper than the bottom of the convective layer. This belief is based on the assumption that the convective layer plays an essential role in the dynamo maintenance of the field. Furthermore, maintaining a field throughout the star would require tremendous kinetic energies in the interior, due to the high density there.

Among the early theories of the solar cycle, reviewed, for example, in Cowling (1953), probably the best known are those of V. Bjercknes (1926, 1937) and H. Alfven (1945a,b). Bjercknes envisioned that two rings of strong magnetic field, one of each polarity, toroidal about the axis of rotation, were carried in a symmetric meridional circulation which extended to undetermined depth. The magnetic field in each ring was supposedly maintained by strong circular motion in meridional planes around the toroidal field. When one ring was near the surface, local instabilities could cause loops or bumps in the ring to develop, which then protruded into the visible photosphere as bipolar sunspot groups. The beginning of a new solar cycle would be marked by the appearance of a new toroidal ring near the surface in higher latitudes.

Alfven, on the other hand, pictured the formation of sunspots as due to the propagation of hydromagnetic "whirl rings" from the core to the surface along the flux lines of the sun's "general" magnetic field, which Alfven believed (as did many others at that time) was basically a dipole field extending throughout the sun. Since it took longer for these hydromagnetic disturbances to reach the surface near the equator than in high latitudes, spots would appear first nearer the poles, marking the beginning of a new cycle. However, enough strong objections have been raised to both Bjercknes' and Alfven's theories (again, see Cowling 1953 for review), which, together with more recent and detailed magnetic field data, make them only of historical interest now.

Present theories of the maintenance of the solar magnetic field center about hydromagnetic dynamo theory. This branch of theoretical

physics is directed at a somewhat more general problem, namely, how magnetic fields in fluid masses can be maintained by internal motions against dissipation by Joule heating. Most work along this line considers motions in rotating, spheroidal, self gravitating fluid bodies, for the obvious reason that most of the major examples in nature (the sun and many magnetic stars, the earth and some other planets) are of this type.

This field has a long and distinguished list of contributors, most of whose work is reviewed in Elsasser (1956). Among the early papers, work by Cowling (1934) (later extended by Bachus and Chandrasekhar 1956) and Ferraro (1937) are of particular importance. Cowling showed that it is not possible to maintain magnetic fields in meridian planes (usually called poloidal fields) which are steady and symmetric about the axis of rotation by motions in meridian planes which are also steady and symmetric. Bachus and Chandrasekhar (1956) extended this theorem to encompass all steady axisymmetric fields and motions, i.e. to include symmetric zonal fields (usually called toroidal) and flows. In brief, steady, axisymmetric dynamos are not possible. Consequently all subsequent work has been concentrated on nonaxisymmetric and/or unsteady models.

Ferraro (1937) was the first to show that in a perfectly conducting fluid mass rotating non-uniformly the only configuration of magnetic field which results in a steady state is one in which the field lines are everywhere parallel to the streamlines. Needless to say, if the fluid has finite conductivity, from Bachus and Chandrasekhar's work, this state cannot remain steady unless the field is everywhere uniform and is completely enclosed by perfectly conducting walls.

With these results known, the major effort was then directed toward finding steady nonaxisymmetric dynamos. This search was formulated in a systematic manner by Bullard and Gellman (1954). The approach was to assume a steady motion field which looks favorable for giving a dynamo, and then to try and find a magnetic field configuration which can be maintained by it (described by a stationary eigenvalue). So far, however, this approach has failed to produce convergent series for a dynamo solution.

The next step was then to look for nonaxisymmetric dynamos which are steady only in some statistical sense, a much less restrictive requirement. Due to the complexity of this problem, virtually all work of this type up to now has been forced to be of a rather qualitative nature. Several physical processes, usually suggested from observed patterns, often occurring on widely different scales, are put together in sequence to provide enough feedbacks to maintain finite amplitude fields. But from this technique it cannot, of course, be proven that in an actual model the sequence would proceed in the manner suggested.

The main elements in this succession of physical events have been put forth, with some variations and alternatives, by Alfven (1950, 1961), Dungey and Loughhead (1954), and Parker (1955a,b), and Elsasser (1956). More specific application of some parts of this approach to the sun were made by Babcock (1961) and Leighton (1964).

As outlined in Elsasser (1956), there are basically three prerequisites for a system to act as a hydromagnetic dynamo. The first is that the motions must have large enough linear scale so that the advective

time scale associated with them is short compared to the decay time of a current due to ohmic dissipation (i.e., the magnetic diffusion time). In other words, the conductivity must be sufficiently high. Otherwise the feedback mechanisms will not have time to be effective. The second is that, since from Cowling's theorem and its extension, dynamos cannot have symmetry in one dimension, the dynamo must possess some mechanism for keeping a low degree of symmetry. An obvious possibility is rotation, through the action of the coriolis forces on the motion.¹ Lastly, the dynamo needs an internal source of energy to generate the motion needed to maintain the fields. This energy source is without doubt a combination of potential and internal energy, arranged in some way as to make it "available" for conversion into kinetic energy of motion. The mode of this conversion therefore must be some form of convection.

Most dynamo schemes then start with a magnetic field confined to meridian planes (i.e., a poloidal field). This field is stretched out into the zonal direction by nonuniform rotation, principally by the horizontal shear. When the toroidal field produced in this manner becomes strong enough, some form of instability is postulated which will bring flux loops to the visible surface. This instability may take the form of a "kink" formed in a toroidal magnetic flux tube due to twisting of the flux lines about a zonal axis (Alfven 1950, 1961; Dungey and Loughhead 1954), a rising up of portions of flux tubes due to "magnetic buoyancy"

1. We know from the Taylor column theorems, however, that the rotation cannot be too large, or else two dimensional motion will again result.

of the tube (Parker, 1955a), or simply of convective motions on the scale of the granulation and supergranulation dragging up flux loops with them (Parker 1955b, Elsasser 1956), or some combination of these three effects. During this rising up process, it is further required that some twisting of the flux loops about a vertical axis take place, so that there will be loops again in the meridional plane. This would be a natural consequence of the kink instability; the influence of rotation on the convective motions might also produce it.

Once the new vertical flux loops in meridian planes are produced, still another process is needed to carry flux to the pole. In the case of the sun, this flux must be of the proper sign to reverse the sense of poloidal field from which the process began. This new poloidal field is then again stretched out into a toroidal field, eventually changing the sign of the original toroidal field, and the whole sequence is repeated. The means by which this poleward migration is effected has been alternatively postulated to be meridional "drift currents" (e.g. Babcock 1961), and diffusion by the supergranulation motions (Leighton 1964).

Parker (1955b) added a new element to the dynamo scheme by considering the hydromagnetics of a plane of fluid instead of a spherical shell. He found that, given small scale twisting motions producing vertical flux loops, when horizontal shear in the flow was present, unstable "migratory dynamo waves" could be produced, which traveled toward the direction of increasing zonal flow. If the conductivity of the fluid was high enough, these waves grew exponentially. Parker suggested that this migratory property might be related to the migration toward the equator of the zone of sunspot activity as the solar cycle progresses.

Babcock (1961) used essentially the dynamo sequence described above (except for Parker's migratory dynamo waves) to explain the solar magnetic features, designating the bipolar magnetic regions as the required vertical flux loops. He made rough calculations of the magnitudes of the various processes, using as a basis the observed total polar flux, from which he found that the times required for the various processes to act fitted reasonably well into the solar cycle.

More recently, Leighton (1964) has taken a somewhat different view of the processes affecting the vertical fluxes. He assumes that the sunspots and sunspot groups are the precursors to new unipolar and bipolar magnetic regions. The supergranular motions, by a kind of random walk process, are supposed to erode the sunspots, diffusing smaller flux tubes away from them. These tubes are then spread out to form the larger scale magnetic regions. In the presence of the differential rotation, the areas of diffused fields become tilted in a manner not unlike the observed fields. If a doublet source is used, representing a bipolar sunspot group, and if this doublet is initially placed with its axis tilted to the meridian in accordance with the observations, the diffusion of the field results in the flux from the trailing portion of the doublet extending further toward the pole than does that from the leading part. Leighton identifies this effect with the formation of the large unipolar regions poleward of 40° discussed in Chapter 1.

As mentioned earlier, Leighton further shows that the cumulative action on many of these spot groups should lead to the production of a poloidal field, which will fit into the rest of the quasicyclic dynamo

process outlined earlier. If the spots originated from an initial sub-surface toroidal field, as is assumed in the dynamo, then the poloidal field produced by Leighton's mechanism will cancel the original poloidal field and form a new one of the opposite sign, as is desired. With the proper value of the diffusion coefficient for the magnetic field, Leighton also can fit his calculations fairly well to the timing of the solar cycle.

Differential rotation

Early ideas on the origin of the solar differential rotation stemmed largely from work by von Zeipel (1924a,b) and Eddington (1925). Von Zeipel was considering the problem of radiative equilibrium in a rotating star. He assumed that the composition of the star was still constant on a level surface, as it would be in the nonrotating case. He then found that in order for the star to satisfy statically the conditions of radiative and gravitational (hydrostatic) equilibrium, the nuclear energy production e within the star must at every point depend on the rate of rotation Ω and the local density ρ in a form

$$e \sim 1 - \Omega^2 / 2\pi\rho G$$

where G is the universal gravitational constant. He also found that under these conditions the effective surface temperature of the star (the temperature of a black body with the same radiative flux) varied with latitude on a level surface, being warmest at the pole and coldest at the equator. The density was a function of latitude in such a way that the pressure on a level surface was constant.

Eddington (1925) argued that it was physically unreasonable to expect the nuclear energy production within the star to obey such a peculiar law (which in general even requires a sink of energy in outer layers). Instead, it seemed much more likely that its conditions would be violated, in which case the pressure would no longer be constant on a level surface. In this case, the horizontal pressure gradient would cause fluid to move along meridians, i.e., a system of meridional circulatory currents would be set up. Since these meridional motions would be acted upon by the coriolis force, zonal currents would be produced, and in general a state of nonuniform rotation would result. Eddington implied that this might even be the means by which the observed differential rotation of the sun was maintained. The idea of meridional currents maintaining the differential rotation was also suggested around the same time by Bjercknes (1926) (see also Bjercknes, 1937) in connection with his theory of the solar cycle, mentioned briefly earlier. Bjercknes postulated that the interior of the sun might have a higher angular velocity than observed on the equator at the surface, which through a meridional circulation could serve as a source of supply for the differential rotation.

However, subsequent calculations (e.g., Opik 1951, Cowling 1953) have shown that Eddington's currents would be much too small ($\sim 10^{-9}$ cm/sec) to play any role in the angular momentum balance even over a time as long as the present assumed age of the sun. Furthermore, it is difficult to see how symmetric meridional motions could put the maximum of angular momentum at the equator at the surface. If the interior were

rotating faster than every point on the surface, it would exert a net torque due to frictional coupling on the surface layer, an imbalance which certainly could not last over the whole age of the sun.

The possibility that the differential rotation is an equilibrium condition in a rotating star was further examined by Schwarzschild (1947). He considered a star in which the core was in convective equilibrium¹ and in solid rotation due to the turbulent mixing, surrounded by an outer shell in radiative equilibrium in which viscosity was negligible. In both the radial and meridional directions, the pressure gradient, gravitational and the centrifugal forces were assumed to be in static balance. The radiative (convective, in the core) flux divergence of heat was balanced by nuclear energy production. In addition to the vanishing of the temperature at the surface, all the variables were required to be continuous at the interface between the convective core and the radiative outer shell. Schwarzschild claimed to have found by numerical integration a solution which contained an equatorial acceleration, actually twice as large as that observed. This solution stood for almost two decades, until Roxburgh (1964) showed that an approximation made by Schwarzschild in satisfying the interface continuity condition on the pressure was not valid. Roxburgh solved the problem without this approximation, and showed that the rotation was a function of radius only, so that on any concentric spherical shell, the rotation was sensibly constant.

1. From more recent stellar structure theory, stars of solar mass and luminosity are now thought to have cores in radiative equilibrium.

The work of Schwarzschild and Roxburgh did not include viscous dissipation, so no energy source was necessary to maintain the motion. Work on the steady state symmetric balance problem in which viscous effects are included has been done by Randers (1938, 1942), Schwarzschild (1942), Krogdahl (1944), and recently by Kippenhahn (1963). Randers early work relied on the release of potential and internal energy by Eddington currents arising as a result of the lack of satisfaction of the earlier mentioned condition on nuclear energy production found by von Zeipel. However, as we have already said, the estimated velocities in such meridional currents are far too small to be effective. His later work (Randers, 1942) pertained to nonuniform rotation in stars with convective cores, since as already stated at the time the sun was thought to fall into that class. His argument, however, if valid for convective cores, would be equally valid for convective envelopes, as in the case of the sun. Randers believed that in convection on a rotating sphere, the motion and therefore the heat transport would tend to take place preferentially along the axis of rotation, i.e., parcels of gas would move as closely as possible along surfaces of constant angular momentum. Parcels trying to move along the angular momentum gradient would be restrained by the centrifugal force. Randers then argued that the excess polar heating brought about in this way would produce a mean meridional circulation, which he claimed could be balanced only by an increased centrifugal force and consequently increased angular velocity in equatorial regions. However, Randers contention that convective motions along the gradient of angular momentum are suppressed can hold only for modes axisymmetric about the axis of rotation. In fact, we now know (Chandrasekhar (1953) that at least the onset of convection should be inhibited near

the poles rather than near the equator. Randers did correctly show that as a consequence of his polar heating, the angular velocity should decrease with height above the equatorial plane (a kind of thermal wind), but it does not necessarily follow from this that the angular velocity decreases with increasing latitude.

Krogdahl (1942) looked into the problem of maintenance of the differential rotation against viscosity in a more mathematical way, and showed that in autobarotropic stars (in which the pressure and density are functionally related) the only possible steady state is one of uniform rotation. Essentially the same result was found also by Schwarzschild (1942). Krogdahl showed that if the star were instead baroclinic, non-uniform rotation could in general exist in the equilibrium state. In the baroclinic star, the isobaric and equipotential surfaces would not in general, coincide, so that there would be pressure forces and consequently accelerations in the meridional plane. This would again give a set of meridional currents, which, when turned by the coriolis force, could give a differential rotation. (However, he did not find an actual solution which gave the equatorial acceleration.) Krogdahl felt that uniform rotation itself directly gave rise to the force to maintain a differential rotation. But it appears that all such a direct effect could produce would be the Eddington currents, which, as we have already said, are much too small. If one considers the influence of rotation on the cellular convection more carefully than was done by Randers, however, there appears to be another way in which the baroclinic state required by Krogdahl might be obtained. This possibility will be discussed in the next chapter.

An attempt to include more sophisticated effects of the granular convection was made by Kippenhahn (1963). He reasoned that since the only steady solution for a barotropic rotating star when the viscosity (which could be molecular or eddy viscosity) was uniform and isotropic was one of solid rotation, it might be better to assume an anisotropic eddy viscosity. He justified this assumption partially on the grounds that since the convective turbulence is driven by gravity, it has a preferred direction. With such an assumption, it was no longer necessary to require the star to be baroclinic to obtain nonuniform rotation. He therefore assumed the vertical temperature gradient to be adiabatic (with uniform temperature in the horizontal). In this case, since pressure force has no curl, taking the curl (only the zonal component contributes) of the equations of motion gives a balance between just the inertial and the viscous terms. Kippenhahn examined this balance for a spherical envelope, taking an anisotropic viscosity which distinguished the radial (vertical) from the lateral directions. With this anisotropy, uniform rotation was not a solution. In addition, the solution always contained symmetric meridional motions. If the viscosity in the radial direction is assumed smaller than in the lateral directions, the meridional circulation produced gives rising motion at the poles and sinking at the equator. For larger radial than lateral viscosity, the meridional currents circulate in the opposite sense. For small radial viscosity, he shows that an equatorial acceleration will result. However, in his approximate formulation, the influence of the derived meridional circulation on the distribution of angular velocity is neglected. Kippenhahn stated that for the sun, his approximations to linearize the problem do not hold too well, but he felt that solution of the nonlinear balance equations would still give an equatorial acceleration.

Although the convective layer may well possess an effectively anisotropic viscosity, it seems difficult to say on either physical or mathematical grounds which component should be the strongest. In Kippenhahn's solution, if the radial viscosity were assumed larger than the lateral, apparently an equatorial "deceleration" would result. Moreover, it is not inconceivable that eddy viscosities in such thermally driven turbulence could be effectively negative (transporting momentum up, rather than down the gradient).

At this point, the meteorological reader should be struck by the similarity in reasoning between some of the theories of the differential rotation of the sun, and some of the earlier theories for the maintenance of the westerlies in the earth's atmospheric general circulation. In particular, the complete reliance on axisymmetric motions, except possibly for turbulence parameterized on the basis of mixing length concepts, stands out. Modern theories of the terrestrial general circulation have of necessity gotten away from these ideas, to recognize the fundamental role played by a particular kind of thermally driven turbulence, namely the cyclones and anticyclones. And as we saw earlier in this chapter, hydromagnetic dynamo theory has had to consider the unsteady, asymmetrical motions as well. Possibly, then, it is time to consider asymmetric, unsteady theories for the solar differential rotation. The statistical work of Ward (1964, 1965a) appears to lend considerable support to this idea. It is even tempting to think that the differential rotation problem and the dynamo problem might be considered as one. The theoretical analysis presented in the remaining chapters appears to fortify this conjecture.

3. A possible alternative approach

It should be evident from Chapter 2 that our theoretical understanding of large scale hydromagnetic and hydrodynamic processes on the sun is far from complete. For one thing, there seems to have been too much emphasis placed on the possible (and in several cases, speculated) role of axisymmetric processes in accounting for the differential rotation. Other difficulties are also apparent. Most of the models are energetically incomplete. For example, in the hydromagnetic dynamo theory for the solar magnetic fields (e.g. Babcock, 1961) the differential rotation is assumed to be given, but the energy source required for its maintenance is not specified. In addition, other motions may be postulated to bring about desired configurations in the magnetic field, but again the energy sources for such motions are often only vaguely suggested, or not specified at all.

A similar energetic incompleteness also pervades theories of the differential rotation. None of these theories attempts to take into account the possibility of work being done on the motion by electromagnetic forces. It is generally assumed that such forces are too small to significantly affect the motion, but it does not appear that this can be unequivocally proven from the observations.

A further possible weakness of many of the theories is their often crucial reliance on parametrized diffusion processes to bring about the desired effects. The differential rotation theory of Kippenhahn (1963) and the large scale magnetic field model of Leighton (1964), both discussed

in Chapter 2, are examples. In Kippenhahn's work, which depends on the eddy viscosity being anisotropic, as we pointed out, it is just as easy to obtain an equatorial "deceleration" as "acceleration", according to the type of anisotropy assumed, and there appears to be little independent physical justification for picking the anisotropy which gives the right answer. In Leighton's work, the value chosen for the magnetic diffusivity is really justified primarily by how well the resulting diffusive transport of vertical magnetic fields fits into the timing of the solar cycle.

But aside from the uncertainty as to the proper value of the eddy viscosity or magnetic diffusivity to be used, reliance on such schemes may fundamentally oversimplify the physical nature of the system. That is, countergradient transport processes will not be allowed. It is now well known that, such processes commonly occur in geophysical fluid systems. For example, countergradient horizontal transports of momentum predominate in maintaining the westerlies in the earth's lower atmosphere, and countergradient horizontal transports of heat are observed in the lower stratosphere, just to name two.

In the light of the above remarks, it seems desirable, as an alternative approach for modelling the differential rotation and large scale magnetic phenomena, to formulate and examine a dynamical system in which motions need not be axisymmetric, where hydromagnetic forces may influence the motions (and also motions may change the magnetic fields), and where parameterized diffusion does not necessarily predominate in determining the character of the motions and magnetic fields. By this I do not mean to argue that diffusive processes are unimportant, or even necessarily less important than other effects, but simply that other possible mechanisms should be examined.

A beginning in this direction has been made by Starr and Gilman (1965a,b). As mentioned in Chapter 1, they discussed (Starr and Gilman 1965a) the energetic implications of Ward's spot correlation statistics. They showed that the countergradient transport of momentum implied by Ward's correlations effects a conversion of energy from the implied horizontal eddy motions to the mean zonal flow, and suggested that this process could be responsible for maintaining the differential rotation against dissipation by friction. They also speculated on the nature of the required energy source for the horizontal eddy motions, suggesting baroclinic conversion of eddy available potential energy, and nonlinear interactions with the smaller scale convective motions, as possibilities.

In Starr and Gilman (1965b), the energetic considerations were extended to include some magnetic effects. The principal point was that horizontal Maxwell stresses, the magnetic analog of horizontal Reynolds stresses, might under certain circumstances act as a significant brake on the differential rotation, thereby opposing the action of the Reynolds stresses. They further showed how both conversions could be taking place as a result of the same hydromagnetic disturbance. This disturbance would be a wavelike distortion of the mean zonal flow which took on the characteristic "tilted trough" pattern so commonly seen on upper air weather maps, first suggested by Starr (1948), and now well known to be primarily responsible for the horizontal eddy flux of momentum in the earth's atmosphere. In the hydromagnetic analog of this disturbance, a tilted structure would appear not only in the stream lines, but also in the magnetic flux lines. These flux lines would originally be part of the subsurface

toroidal field, which, due to the high electrical conductivity of the solar gases, would tend to follow the fluid particles, resulting in the tilted shape. Starr and Gilman put forth Howard's (Bumba and Howard 1965b, discussed in Chapter 1 and reproduced in the appendix) solar magnetograms of the line of sight fields, showing patterns of large scale, tilted in the desired sense, as observational evidence for this kind of disturbance. They further speculated that the energy taken from the differential rotation by the Maxwell stresses, which goes into "eddy" magnetic energy, might be one of the principal energy sources for some of the solar magnetic activity. No further light was shed, however, on a possible energy source for the eddy motions themselves. However, it is obvious that whatever its nature, it must be able to supply energy at a greater rate than the Maxwell stresses are taking it away.

With the addition of the work of Starr and Gilman (1965a,b) to the previous theories for the differential rotation and solar cycle phenomena, it appears that there are basically two possible directions for further research to take, both of which merit considerable attention.

On the one hand, if one rejects Ward's interpretation of the spot correlations, as does Leighton (see Chapter 1), then one really must require that the granular and supergranular motions be responsible not only for the large scale magnetic field patterns, but also for maintaining the differential rotation. It may in fact be possible for these motions to accomplish both these tasks, but present theory of turbulent convection in a basic flow with shear is not developed nearly enough to tell.

On the other hand, if one accepts Ward's interpretation of the spot correlations, the work of Starr and Gilman follows quite naturally, and one is then led to the conclusion that the source of energy for the large scale eddy motions and zonal flow is most likely to be of comparable scale. If this is so, then the obvious source is the same as in the terrestrial atmospheric case, namely available potential energy associated with large scale horizontal density variations. It is the possible consequences of this kind of energy source which we wish to examine in this thesis.

In the case of the earth's atmosphere, we know that the required horizontal density variations are set up as a result of differential heating by the sun. It is obvious that no analogous differential heating by an external source takes place on the sun. Furthermore, as discussed in Chapter 1, the attempts to measure meridional temperature gradients by means of differential limb darkening have proved inconclusive. But these two considerations do not necessarily rule out differential heating by an internal source, namely the convection itself.

In considering this possibility, the work of Lorenz (1955) is helpful. Lorenz defined "available potential energy" as the difference between the total potential energy present and the minimum value which could be reached by adiabatic redistributions of mass. In order for the available potential energy to be positive, the stratification of the fluid must not be both exactly horizontal and gravitationally stable. If the stratification is horizontal but gravitationally unstable, then the potential energy available can go only into ordinary convective motions. If the stratification is gravitationally stable but not exactly horizontal, then the available

potential energy can go only into Hadley cell motions or unstable baroclinic waves (i.e., cyclones and anticyclones). If the stratification is gravitationally unstable and not horizontal, some combination of these motions should result.

In the case of the sun's convective layer, it is obvious that available potential energy due to a gravitationally unstable stratification is being released. But due to the sun's rotation and finite amplitude effects this may not be the whole picture. Chandrasekhar (1952, 1953) has shown that the onset of convection of a liquid between parallel horizontal planes is inhibited by the vertical component of rotation (the critical Rayleigh number increases with the Taylor number). Recently Veronis (to be published) has studied steady finite amplitude convection in a liquid with free boundaries, using a truncated fourier series model. He shows that in general the horizontally averaged vertical temperature gradient depends upon both the Rayleigh and Taylor numbers. The implication for the solar case from this albeit vastly simpler system is that, since the normal component of rotation is a function of latitude, the Taylor number would be a function of latitude, so that meridional temperature gradients, and therefore, in general, density gradients, could be set up. Thus the convection, responding to one kind of available potential energy, namely that due to a gravitationally unstable density stratification, could conceivably produce the other kind, that due to horizontal density variations. This could be the baroclinic atmosphere that Krogdahl (1944) was searching for, as discussed in Chapter 2. Unfortunately, it is virtually impossible to estimate the magnitude of this effect for the sun, due to the uncertainty of many of the parameters involved.

Even if the above scheme could produce significant horizontal density variations, a further theoretical problem would necessarily remain, were it not for the additional results of Veronis. That is, if the convective layer remained gravitationally unstable everywhere, it would be very difficult in regard to the equations to deal with baroclinic overturnings, since the unstable convective modes would still be present. To get at the baroclinic effects separately, we need to deal with a fluid layer that, over time periods long compared with the convective time scales, is gravitationally stable. In the light of some further results of Veronis, it may not be unreasonable to assume this, for at least some levels in the solar convection layer. Veronis found that over a wide range of Taylor number and supercritical Rayleigh number, the mean vertical temperature gradient was actually gravitationally stable through a large part of the interior of the fluid, with highly unstable layers occurring at the upper and lower boundaries. Possibly, then, the time averaged vertical temperature gradient in the interior of the convection layer on the sun is also stable. (If so, this would be still another example in which a transport process, in this case vertical heat transport, could not be represented meaningfully as parameterized diffusion).

Parenthetically, we should remark that in stellar structure theory, it is generally assumed that convective layers have adiabatic temperature gradients. The implications for stellar models of stable gradients in such layers are probably not very great, since the stable regions are not very thick compared to the stellar radius, and because the deviations from an adiabatic gradient are not large.

So far, then, we have suggested that the small scale convection could produce a layer which was gravitationally stable over times long compared to the convective time scales, and that within this layer large scale meridional temperature gradients could be established due to the varying normal component of rotation. The available potential energy associated with these meridional temperature gradients is usually called mean zonal available potential energy. To get from this energy source to potential energy converting baroclinic or cyclone waves, however, one further difficulty must be considered. The difficulty is in how long zonal, as opposed to meridional, temperature gradients can exist in the sun in the face of the eddy conduction of heat. These zonal temperature gradients, which represent what is usually called eddy available potential energy, are essential to the growth of baroclinic waves.

To illustrate the difficulty most clearly, we turn to a setting seemingly far removed from the solar problem, namely the laboratory. Energy converting disturbances that can grow from zonal available potential energy have been studied extensively in the laboratory, by Hide (1958) and Fultz (1961) and many others. These studies comprise the so-called rotating dishpan and rotating annulus experiments. In these experiments, well known now to most dynamic meteorologists, a cylindrical or annular shaped container is filled with liquid (usually water). The container is then allowed to rotate at a controllable rate about its axis of symmetry, and by means of thermal reservoirs exterior to the container, a temperature difference is established between the center and edge, in the case of the cylinder, or between the inner and outer side walls, in the case of the annulus.

The motions of the liquid which result in response to this differential heating are then examined for different rotation rates and horizontal temperature contrasts. Depending on these, and other parameters which are fixed once the fluid is chosen (e.g., the thermal conductivity and viscosity) several different modes of motion may result. The simplest is the so-called Hadley regime, in which the motions are axisymmetric, consisting of a meridional cell (with rising motion at the warm wall and sinking at the cold wall) with an associated zonal flow produced from the meridional flow by the coriolis force. For other rotation rates and temperature contrasts, the so-called Rossby regime, results, in which the motion is asymmetric, consisting of horizontal waves or meanders on a zonal flow. In this case, the isotherms also have a wavelike structure, and are positioned relative to the streamlines in such a way that warm liquid is being carried toward the cold wall, and vice versa. Within the waves, eddy available potential energy is being converted to kinetic energy, through the rising of liquid in the warm parts of the wave, and sinking in the cold parts. For still other conditions, more complicated regimes may be found.

One of the requirements for the existence of energy converting wavelike motions of the Rossby regime described above is that the molecular thermal conductivity of the liquid be relatively low. That is, it should be low enough so that the zonal temperature perturbations are not virtually destroyed before the eddy heat transport and energy conversion processes have a chance to take place. This requirement is easily met for most fluids used in the laboratory. It is also satisfied

fairly well for the analogous situation in the earth's atmosphere. That is, the potential energy converting cyclone waves are observed to grow much faster than radiation or eddy heat conduction can dissipate them.

In the solar case, on the other hand, since the convection is so vigorous, it may be more difficult to maintain such gradients. That is, Veronis' work implies that along a given latitude circle, the convection will tend to make the mean vertical temperature structure the same, even if it is gravitationally stable. Consequently zonal temperature gradients would tend to be wiped out. If this effect is strong enough, then the only kind of large scale motion which could result would be of the Hadley regime type. However, considering the possibility of nonlinear interactions between the small and large scales of motion, to expect zonal gradients to last may not be completely unreasonable. To see how the two scales of motion might interact and compete, it would be interesting to extend the annulus experiments, imposing a temperature gradient between top and bottom as well as at the inner and outer walls. On the sun, one might be tempted to cite the plage and facular regions as observational evidence for zonal temperature gradients and hence eddy available potential energy, but these temperature fluctuations may be specifically due to the presence of the magnetic fields, as discussed by Simon and Leighton (1964).

In summary, then, we propose to reduce the problem of the solar general circulation to consideration of the dynamics of a gravitationally stable baroclinic circumpolar vortex, which is confined to a relatively thin spherical shell in the convection layer. To include hydromagnetic

effects, we shall assume with most previous authors that there exists a subsurface mean toroidal magnetic field, everywhere parallel to the mean zonal flow. From both a thermal and a magnetic point of view, then, the real seat of the large scale dynamics will be somewhere in the interior of the convection layer, rather than near its upper boundary to which most of the observations previously discussed refer. Nonetheless these observations could all fit fairly well into such a picture. The meridional temperature gradients, not conclusively seen at the surface, could instead be present in the deeper layer. Veronis' work would suggest this since the temperatures in his convection model (and almost all others) are fixed at the top and bottom boundaries. Furthermore, it is only in the interior of the convecting layer that the stable mean vertical temperature gradients can occur. The fact that the observed large scale magnetic fields at the surface are primarily vertical is also consistent. These could be just the vertical parts of closed or helical flux loops whose lower horizontal parts are associated with the toroidal field in the interior layer, with the upper parts being in the chromosphere or corona. The sunspot fields could be similarly anchored. The horizontal Maxwell stresses discussed by Starr and Gilman (1965b), then, would be distortions of the toroidal field, and as such would be confined to the interior layer. The motions of the sunspots and their correlations, could also be evidence for the presence of horizontal eddies and their associated Reynolds stresses which are really produced in the deeper layer. The differential rotation itself could easily be supposed to extend at least this deep, if not all the way to the bottom of the convective layer.

In order to study this baroclinic hydromagnetic circumpolar vortex mathematically, however, many further assumptions, some of them physically not very realistic, will have to be made. These will be discussed in detail with the formulation of the equations in Chapter 4. The principal effort in later chapters will be directed toward the stability properties of this vortex with respect to normal mode disturbances, the structure of unstable disturbances that can be produced, and their energy conversion properties. As such, it constitutes a direct generalization of work that has been done on the nonmagnetic baroclinic vortex problem by Phillips (1954), Charney and Stern (1962), and particularly by Pedlosky (1964a,b). The ultimate goal of this investigation will be to see whether the dynamics of large scale disturbances on such a hydromagnetic vortex include processes that would maintain an equatorial acceleration and at the same time act as a dynamo to simulate the solar cycle. While the results that follow do not answer this question completely in the affirmative, they indicate a good possibility of such a regime.

4. Formulation of the equations governing large scale hydromagnetic flow.

Physical and scaling assumptions

Several assumptions are needed to simplify the physics of our problem. Some of these assumptions agree quite well with the actual conditions in the solar convection layer. Others exclude physical effects which are undoubtedly quantitatively important in the actual dynamics, but which do not appear to add any qualitatively crucial processes. We shall assume then,

1. Local thermodynamic equilibrium exists through the layer. This should hold very well in the interior of the convection layer.

2. No nuclear energy production anywhere in the layer. This holds very well throughout the convection layer.

3. The medium is a perfect gas, which consists of partially ionized hydrogen and helium, rendering it very highly electrically conducting, and very opaque to radiation. These should hold quite well. However, the thermodynamic effects of partial ionization will be neglected. That is, the equation of state for the system will be written for an unionized gas, and the heat of ionization will be omitted from the thermodynamic equation. Obviously, the thermodynamic effects would make a quantitative difference in the dynamics, (e.g., the available potential energy for a given temperature distribution would be changed), but they do not seem to introduce any important new phenomena.

4. The thickness of the layer is small enough so that gravity may be taken as a constant, and so that divergence of radii from the center of the sun may be neglected. The self-gravitation of the region will also be neglected.

In addition to these assumptions about the physical nature of the system, we also must make several assumptions about the time and space scales of the hydromagnetic disturbances to be described by the equations. These restrictions encompass the well known "filtering" approximations (e.g. Thompson 1961) of dynamic meteorology, but go a step further to allow for hydromagnetic effects.

To retain continuity with the nonmagnetic problem, the notation of Charney and Stern (1962) and Pedlosky (1964a,b) will be largely retained. The scaling arguments for the nonmagnetic aspects will be parallel to those works, though admittedly some of the assumptions are less supported by observational evidence in the solar case.

If we accept the spot displacements as measures of the large scale flow, then the horizontal scale of the typical disturbance in this flow must be larger than typical sunspot or sunspot group dimensions. Let us take this scale, denoted by L , to be $L \sim 10^5$ km. L of course, is not the characteristic wavelength of the disturbance, but rather the distance over which the perturbation variables are changed by an amount equal to their typical values. That is, the inverse of the wave number k or $2\pi/\lambda$, where λ is the wavelength, would be a reasonable measure of the scale. Taking $L \sim 10^5$ km corresponds on the sun to about six complete wavelengths around a latitude circle in middle or low latitudes, which fits well with the observed magnetograms cited earlier.

Letting a be the radius of the sun, we note then that $L/a \ll 1$ since $a = 7 \times 10^5$ km. This allows us to use a local cartesian rather than spherical coordinate system, a considerable simplification. The constraint due to the variation in the normal component of rotation with latitude (the so-called β effect) remains, but the variation becomes linear in the meridional coordinate. Obviously, then, we are not able to handle very well disturbances which extend nearly from the pole to the equator.

In the terrestrial case, the horizontal flow is conventionally assumed to be geostrophic. That is, to the lowest order there is a balance between horizontal pressure gradient and Coriolis forces. Inertial and viscous forces are generally assumed to be of higher order. These approximations are well supported by observations of the actual dynamics of the earth's atmosphere. In the solar case, we cannot, of course, measure either the pressure gradient or the viscous forces directly. However, if our abstraction of the solar general circulation problem to the consideration of a baroclinic circumpolar vortex has any validity, we would still expect the horizontal pressure forces to play a key role.

As for the viscous forces, if one were to argue on the basis of mixing length theory, using the granulation velocities and dimensions discussed in Chapter 1, one could claim that viscous forces due to small scale vertical stresses are actually as strong or stronger than coriolis forces associated with the large scale flow. However, since the granulation and supergranulation are thermally driven, mixing length arguments applied to mechanical transport processes, e.g., transport of momentum, may not be applicable. In any case, the question of the magnitude of

the magnitude of the viscous forces can be argued from another point of view. That is, they cannot dissipate the kinetic energy of the mean flow any faster than the energy sources renew it. This renewal time has been estimated by Ward and by Starr and Gilman from the conversion of eddy to zonal kinetic energy, as from one to several rotations. Starr and Gilman (1965b) pointed out, the inclusion of horizontal Maxwell stresses may increase the renewal time appreciably. Therefore, at least to be consistent with the interpretation of spot displacements as reflecting large scale motions, we should, as in the terrestrial case, assume that the viscous forces are of higher order.

The inertial forces can be estimated from Ward's data. They will be small compared to the coriolis force, and the geostrophic balance (or, more appropriately, heliostrophic) will be complete in the absence of electromagnetic forces, if the time scale of the disturbances is the advective scale L/U , where U is the characteristic horizontal velocity, and if the Rossby number $Ro = U/2\Omega L \ll 1$, where Ω is the angular rotation rate of the sun. At least in sunspot zones $\Omega \sim 3 \times 10^{-6} \text{ sec}^{-1}$, and from Ward's data, $U \sim 50 \text{ m/sec}$, so that $Ro \sim 10^{-1}$.

When electromagnetic forces are included, an additional assumption is needed to give a heliostrophic balance, namely, that the Alfvén velocity $M/(4\pi\rho)^{1/2}$, where M is the characteristic horizontal magnetic field strength and ρ the density, is of the same size or less than U . We shall use the parameter $P = M^2/4\pi\rho U^2$, the ratio of typical

horizontal magnetic to horizontal kinetic energy in the system. Thus if we assume $P \lesssim 1$, the motion will be heliostrophic to lowest order. Near the top of the convection layer, say for $\rho \sim 3 \times 10^{-7} \text{ gm cm}^{-3}$, we would have to take $M \lesssim 10$ gauss. For a deeper interior layer, such as we are assuming, for which, say, $\rho \sim 10^{-9}$, we are allowed $M \lesssim 200$ gauss, which is the same order as is generally assumed to exist at such depths. Although the assumption that $P \lesssim 1$ appears to be ad hoc, it is supported a posteriori by the results in Chapters 8 and 9.

Given that the motion is quasi-heliostrophic, and that $L/a \ll 1$, we may easily show that the motion will also be horizontally nondivergent to lowest order. Therefore to lowest order we shall be able to represent the horizontal velocities by a stream function. Because the motion is quasi-heliostrophic, gravity waves will effectively be "filtered out" of the system (e.g. Thompson 1961).

Since there are no known magnetic monopoles, the magnetic field is always nondivergent in three dimensions. But, as a consequence of the heliostrophic and horizontally nondivergent nature of the motion, we can show that the magnetic field will also be horizontally nondivergent (to the present order of approximation). This important simplification will be demonstrated when the hydromagnetic equations are derived. As a consequence, we can also represent the horizontal magnetic field to lowest order by a "stream" function. In this case, the streamlines are simply the magnetic flux lines.

Since the system is gravitationally stable, we shall assume the horizontally averaged pressure and density to be in hydrostatic balance. For the perturbation pressure and density to be hydrostatically related, it is sufficient to assume that the vertical scale of the disturbances, D , is much smaller than the horizontal scale L ; i.e. $\delta = D/L \ll 1$. The assumption that $P \lesssim 1$ is then enough to keep the magnetic fields from entering into the balance. Since $L \sim 10^5 \text{ km}$ all disturbances except those which extend through the entire convective layer will be hydrostatic.

For simplicity, the perturbations will be assumed to be adiabatic. Therefore the specific entropy, or equivalently the potential temperature (see (4-33)), of a parcel will be conserved. Intuitively, this does not seem like a very good assumption, since there is so much small scale vertical mixing. That is, as discussed in Chapter 3, the maintenance of eddy available potential energy may be difficult. We do not know how large an effect this will be, but we shall assume that it is not so strong as to completely wipe out the perturbations. In order to be potential energy releasing, the particle trajectories in the perturbations must have slopes that are less inclined to the horizontal than are the potential temperature surfaces. That is, we must have

$$\frac{w}{U} \div - \frac{\partial/\partial y \ln \theta}{\partial/\partial z \ln \theta} \lesssim 1$$

where w is a typical vertical velocity, θ the potential temperature, and y and z are horizontal and vertical coordinates. Since the motion is horizontally quasi-nondivergent, the numerator is of order δR_0 . The denominator may be estimated from the thermal wind relation. That is

$\frac{\partial}{\partial y} \ln \theta \sim 2 \Omega U / g D$, where g is gravity. If we let

$K = D \frac{\partial}{\partial z} \ln \theta$ be a measure of the static stability, then we

have that $\epsilon \equiv 4 \Omega^2 L^2 / g D K \geq 1$ for potential energy releasing motions. For the terrestrial troposphere, $K \sim 10^{-1}$ gives

$\epsilon \approx 1$, but for the sun, since $4 \Omega^2 L^2 / g D \lesssim 10^{-4}$, K need be no larger than 10^{-4} . That is, the solar mean vertical temperature gradients need not be nearly so much less than the adiabatic as their atmospheric counterparts. For simplicity, ϵ is assumed to be a constant in time and space.

It is necessary primarily on mathematical grounds to limit the vertical extent of the disturbances according to the "quasi-Boussinesq" assumption (Charney and Stern, 1962). This assumption does not rule out effects of compressibility altogether, but rather excludes those disturbances whose vertical scale is much greater than the scale height $S = RT/g$, where T is the temperature. That is, letting $D/S = \mu$, we assume that $r \mu \lesssim 1$, where $r = c_v / c_p$ is the ratio of specific heats. On the sun, at the depths we are considering, $S \sim 2000$ km, so it is difficult to say how bad a physical approximation this is, but neither baroclinic stability theory nor convection theory have proceeded very far without it.

A summary of the nondimensional numbers which define the approximations made above is presented in Table 1.

The filtering approximations customarily used in dynamic meteorology effectively screen out sound waves (with the hydrostatic approximation) and

Table 1. Scaling Assumptions

$$Ro = \frac{U}{2\Omega L} \ll 1$$

$$P = \frac{M^2}{4\pi\rho U^2} \lesssim 1$$



heliostatic

$$K = D \frac{\partial}{\partial z} \ln \theta \ll 1$$

$$E = \frac{4\Omega^2 L^2}{gDK} \gtrsim 1$$



potential energy
converting

$$\frac{L}{a} \ll 1$$

horizontally
nondivergent
(if heliostatic)

$$S = \frac{D}{L} \ll 1$$

hydrostatic

$$r\mu = \frac{rD}{S} \lesssim 1$$

quasi-Boussinesq

gravity waves (with the geostrophic approximation). In magneto-hydrodynamics, a similar procedure is often used to filter out electromagnetic waves. It is sufficient that all the characteristic velocities of the system, i.e., the flow, Alfvén, sound, and gravity wave velocities, be nonrelativistic, a requirement that is easily satisfied for the sun. When these conditions hold, there is no relativistic correction to the density, displacement currents are negligible, and advection of charge by the flow does not contribute significantly to the total current. In addition, the body force due to finite charge densities in an electric field may be neglected. (See Hide and Roberts (1962), or Elsasser (1956) for the detailed scaling arguments).

Unscaled equations

Incorporating all of the physical assumptions made above, we may now write down the unscaled equations governing inviscid, adiabatic, perfectly conducting hydromagnetic flow. Let us define \underline{u} to be the velocity vector, p the pressure, α the specific volume, \underline{E} the electric field, \underline{H} the magnetic field, \underline{J} the electric current density, ϕ the solar latitude, x, y, z the local zonal, meridional and vertical directions, and R the gas constant/unit mass. Cgs and electromagnetic units are used, and since the material is not ferromagnetic, the magnetic permeability is unity. Asterisks are used to distinguish dimensional variables. Then the equation of motion is given by

$$\frac{d\underline{u}^*}{dt^*} = -2\underline{\Omega} \times \underline{u}^* - \alpha^* \text{grad } p^* - g \underline{k} + \alpha^* \underline{J}^* \times \underline{H}^* \tag{4-1}$$

The equation of mass continuity is

$$\frac{d\rho^*}{dt^*} + \rho^* \operatorname{div} \underline{u}^* = 0 \quad (4-2)$$

The thermodynamic equation is

$$\frac{d}{dt^*} \ln \theta^* = 0 \quad (4-3)$$

The equation of state is

$$p^* = \rho^* R T^* \quad (4-4)$$

The relevant Maxwell's equations are

$$\operatorname{curl} \underline{H}^* = 4\pi \underline{J}^* \quad (4-5)$$

$$\operatorname{curl} \underline{E}^* = -\frac{\partial \underline{H}^*}{\partial t} \quad (4-6)$$

and

$$\operatorname{div} \underline{H}^* = 0 \quad (4-7)$$

Since the electrical conductivity is infinite, the conduction current precisely balances the induction current. That is

$$\underline{E}^* = -\underline{u}^* \times \underline{H}^* \quad (4-8)$$

We can eliminate consideration of the current density \underline{J}^* and the electric field \underline{E}^* from the problem, by substituting from (4-5) into (4-1), and from (4-8) into (4-6). Thus we have

$$\frac{d\underline{u}^*}{dt^*} = -2\underline{\Omega} \times \underline{u}^* - \alpha^* \operatorname{grad} p^* - g\underline{k} + \frac{\alpha^*}{4\pi} (\operatorname{curl} \underline{H}^*) \times \underline{H}^* \quad (4-9)$$

and

$$\frac{\partial \underline{H}^*}{\partial t^*} = \text{curl} (\underline{U}^* \times \underline{H}^*) \quad (4-10)$$

Equation (4-10) is a mathematical expression of the well known fact that when the conductivity of the medium is infinite, the magnetic field lines are attached to fluid particles. This may be seen more clearly if we note the vector identity

$$\begin{aligned} \text{curl}(\underline{U}^* \times \underline{H}^*) &= \underline{U}^* \text{div} \underline{H}^* - \underline{H}^* \text{div} \underline{U}^* \\ &\quad + (\underline{H}^* \cdot \text{grad}) \underline{U}^* - (\underline{U}^* \cdot \text{grad}) \underline{H}^* \end{aligned} \quad (4-11)$$

From (4-7), the first term on the right in (4-11) vanishes. Using (4-11), we may write (4-10) as

$$\frac{d}{dt^*} \underline{H}^* = (\underline{H}^* \cdot \text{grad}) \underline{U}^* - \underline{H}^* \text{div} \underline{U}^* \quad (4-12)$$

From (4-12), it is evident that the field strength following a certain particle of fluid can be changed only by either twisting the lines of force by shearing motion (the first term on the right in (4-12)), or by squeezing together or pushing apart of field lines by divergent motions (the second term). In both of these processes the field lines are dragged with the fluid as it moves.

We may equivalently argue that the total magnetic flux through an element of area made up of a small group of fluid particles is conserved as these particles move and the element is distorted.

In order to nondimensionalize the equations in terms of the parameters defined earlier, it is convenient to separate the vertical from horizontal components, and to expand some of the vector products by means of appropriate vector identities. Thus we define

$$\underline{u}^* = \underline{v}^* + w^* \underline{k} \quad (4-13)$$

where

$$\underline{v}^* = u^* \underline{i} + v^* \underline{j} \quad (4-14)$$

$$\underline{H}^* = H^* \underline{i} + h^* \underline{k} \quad (4-15)$$

where

$$\underline{H}^* = H_x^* \underline{i} + H_y^* \underline{j} \quad (4-16)$$

We also split the operators

$$\text{curl} = \nabla^* \times + \partial/\partial z^* \underline{k} \times \quad (4-17)$$

$$\text{div} = \nabla^* \cdot + \partial/\partial z^* \underline{k} \cdot \quad (4-18)$$

$$\text{grad} = \nabla^* + \partial/\partial z^* \underline{k} \quad (4-19)$$

The horizontal equation of motion becomes

$$\begin{aligned} \left(\frac{\partial}{\partial t^*} + \underline{v}^* \cdot \nabla^* + w^* \frac{\partial}{\partial z^*} \right) \underline{v}^* &= -2\Omega \sin \phi \underline{k} \times \underline{v}^* - 2\Omega \cos \phi w^* \underline{i} \\ &\quad - \alpha^* \nabla^* p^* \\ &\quad + \frac{\alpha^*}{4\pi} \left[(\nabla^* \times \underline{H}^*) \times \underline{H}^* + h^* \frac{\partial \underline{H}^*}{\partial z^*} - \nabla^* \frac{h^{*2}}{2} \right] \end{aligned} \quad (4-20)$$

while the vertical equation is

$$\left(\frac{\partial}{\partial t^*} + \underline{V}^* \cdot \underline{\nabla}^* + w^* \frac{\partial}{\partial z^*}\right) w^* = 2\Omega \cos\phi \underline{V}^* \cdot \underline{z} - \alpha^* \frac{\partial p^*}{\partial z^*} - g + \frac{\alpha^*}{4\pi} \left(\underline{H}^* \cdot \underline{\nabla}^* h^* - \frac{\partial}{\partial z^*} \frac{H^{*2}}{2} \right) \quad (4-21)$$

The mass continuity equation (4-2) becomes

$$\left(\frac{\partial}{\partial t^*} + \underline{V}^* \cdot \underline{\nabla}^* + w^* \frac{\partial}{\partial z^*}\right) \rho^* + \rho^* \underline{\nabla}^* \cdot \underline{V}^* + \rho^* \frac{\partial w^*}{\partial z^*} = 0 \quad (4-22)$$

The equation for the horizontal magnetic field may then be written as

$$\left(\frac{\partial}{\partial t^*} + \underline{V}^* \cdot \underline{\nabla}^* + w^* \frac{\partial}{\partial z^*}\right) \underline{H}^* = \underline{H}^* \cdot \underline{\nabla}^* \underline{V}^* + h^* \frac{\partial}{\partial z^*} \underline{V}^* - \underline{H}^* \underline{\nabla}^* \cdot \underline{V}^* - \underline{H}^* \frac{\partial w^*}{\partial z^*} \quad (4-23)$$

while the vertical magnetic field equation becomes

$$\left(\frac{\partial}{\partial t^*} + \underline{V}^* \cdot \underline{\nabla}^* + w^* \frac{\partial}{\partial z^*}\right) h^* = \underline{H}^* \cdot \underline{\nabla}^* w^* - h^* \underline{\nabla}^* \cdot \underline{V}^* \quad (4-24)$$

Here use has again been made of the vector identity (4-11), and (4-7).

Equation (4-7) becomes

$$\underline{\nabla}^* \cdot \underline{H}^* + \frac{\partial h^*}{\partial z^*} = 0 \quad (4-25)$$

Finally, the thermodynamic equation (4-3) is written as

$$\left(\frac{\partial}{\partial t^*} + \underline{V}^* \cdot \underline{\nabla}^* + w^* \frac{\partial}{\partial z^*}\right) \ln \theta^* = 0 \quad (4-26)$$

The equation of state (4-4) remains the same.

Scaled equations; their expansion in the Rossby number

Anticipating the horizontal quasi-nondivergent nature of the flow and the magnetic fields, it is helpful to split these horizontal fields into irrotational and nondivergent parts. That is, we define

$$\underline{V}^* = \underline{V}_\psi^* + \underline{V}_\sigma^* \quad (4-27)$$

where $\underline{V}_\psi^* = \underline{k} \times \nabla^* \psi^*$; $\underline{V}_\sigma^* = \nabla^* \sigma^*$ (4-28)

ψ^* is then the stream function for the flow, and σ^* is the velocity potential.

Similarly, we define

$$\underline{H}^* = \underline{H}_\chi^* + \underline{H}_\delta^* \quad (4-29)$$

where

$$\underline{H}_\chi^* = \underline{k} \times \nabla^* \chi^*; \quad \underline{H}_\delta^* = \nabla^* \delta^* \quad (4-30)$$

Using the scaling parameters defined earlier, we may define non-dimensional independent variables for the system as

$$x^*, y^* = Lx, Ly; \quad z^* = Dz; \quad t^* = \frac{L}{U} t \quad (4-31)$$

The nonmagnetic dependent variables may be nondimensionalized in the following way

$$\left. \begin{aligned} \underline{V}_\psi^* &= U \underline{V}_\psi \\ \underline{V}_\sigma^* &= Ro U \underline{V}_\sigma \\ w^* &= \delta Ro U w \end{aligned} \right\} \begin{aligned} p^* &= p_s^* (1 + \mu k \epsilon Ro p) \\ \rho^* &= \rho_s^* (1 + K \epsilon Ro \rho) \\ \alpha^* &= \alpha_s^* (1 + K \epsilon Ro \alpha) \\ \ln \theta^* &= \ln \theta_s^* + K \epsilon Ro \theta \end{aligned} \quad (4-32)$$

where the subscript, s , denotes a horizontal average.

The log of the potential temperature, which is proportional to the specific entropy of the gas, is related to the pressure and specific volume according to the formula

$$\ln \theta^* = r \ln p^* + \ln \alpha^* + \text{const.} \quad (4-33)$$

Then from (4-32) and (4-33) it immediately follows that

$$\theta = r \mu p + \alpha \quad (4-34)$$

As stated earlier, the horizontally averaged pressure and density are in hydrostatic balance. That is,

$$-\frac{\partial p_s^*}{\partial z} - \rho_s^* g = 0 \quad (4-35)$$

To obtain guidance in scaling the magnetic field, we should examine equation (4-24) for the vertical magnetic field. When this equation is made nondimensional, the left hand side is at most of order Ro . However, the right hand side, since one term contains the vertical motion and the other the horizontal divergence, will be at most of order Ro^2 . Therefore, when h and the other dependent variables are expanded in powers of Ro , as will be done in the next section, the terms first order in Ro would give simply

$$\left(\frac{\partial}{\partial t} + \underline{v}_\perp \cdot \nabla \right) h = 0 \quad (4-36)$$

That is, the vertical magnetic field attached to the quasi-horizontally moving particles will be conserved. Thus if the vertical fields are

initially zero everywhere, they will remain zero. In our perturbation study the flow and magnetic fields initially must satisfy Ferraro's law (c.f. Chapter 2) in order to be steady, i.e., they must be coplanar. Since, in general, the flow will have vertical shear, the vertical magnetic fields must therefore initially vanish everywhere, so that, from (4-36), to first order in Ro , they will remain zero. But from (4-25), this implies that the horizontal divergence of the horizontal field is zero to lowest order. Consequently, just as with the horizontal velocity field, the horizontal magnetic field can to lowest order be represented by a "stream" function. In this case, the "streamlines" represent the magnetic flux lines. In view of the assumption of infinite conductivity, this is a reasonable result. That is, since the field lines will be attached to fluid particles, fluid motions which are horizontally non-divergent to lowest order should tend to produce field patterns which are also horizontally nondivergent.

From the above reasoning, then, we can nondimensionalize the magnetic fields as follows

$$\begin{aligned} \underline{H}_x^* &= M \underline{H}_x \\ \underline{H}_y^* &= Ro M \underline{H}_y \\ h^* &= \delta Ro M h \end{aligned} \tag{4-37}$$

With the scaling equations (4-32) and (4-37), and the relations (4-33), (4-34), and (4-35), then, we may nondimensionalize equations (4-20)-(4-26), respectively, in the form

$$\begin{aligned}
 R_0 \left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla + R_0 w \frac{\partial}{\partial z} \right) (\underline{V}_\psi + R_0 \underline{V}_\sigma) = \\
 - \sin \phi \frac{k}{\rho} \times (\underline{V}_\psi + R_0 \underline{V}_\sigma) + R_0 \delta \cos \phi w \underline{i} \\
 - (1 + K \epsilon R_0 \alpha) \nabla p + (1 + K \epsilon R_0 \alpha) P_s R_0 \left\{ (\nabla \times \underline{H}_x) \right. \\
 \left. \times (\underline{H}_x + R_0 \underline{H}_y) + R_0 h \frac{\partial}{\partial z} (\underline{H}_x + R_0 \underline{H}_y) - \delta^2 \nabla \frac{h^2}{2} \right\}
 \end{aligned} \tag{4-38}$$

$$\begin{aligned}
 \delta^2 R_0^2 \left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla + R_0 w \frac{\partial}{\partial z} \right) W = \\
 \delta \cos \phi (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \underline{i} - \alpha_s^* (1 + K \epsilon R_0 \alpha) \frac{\partial}{\partial z} \frac{p}{\alpha_s^*} - \alpha \\
 + (1 + K \epsilon R_0 \alpha) P_s R_0 \left\{ \frac{\partial}{\partial z} (\underline{H}_x + R_0 \underline{H}_y)^2 \right. \\
 \left. + R_0 \delta^2 (\underline{H}_x + R_0 \underline{H}_y) \cdot \nabla h \right\}
 \end{aligned} \tag{4-39}$$

$$\begin{aligned}
 R_0 \rho_s^* K \epsilon \left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla \right) \rho + R_0 \rho_s^* (1 + R_0 K \epsilon \rho) \nabla \cdot \underline{V}_\sigma \\
 + R_0 \frac{\partial}{\partial z} \left(P_s^+ (1 + R_0 K \epsilon \rho) w \right) = 0
 \end{aligned} \tag{4-40}$$

$$\begin{aligned}
 R_0 \left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla + R_0 w \frac{\partial}{\partial z} \right) (\underline{H}_x + R_0 \underline{H}_y) \\
 = R_0 \left\{ (\underline{H}_x + R_0 \underline{H}_y) \cdot \nabla (\underline{V}_\psi + R_0 \underline{V}_\sigma) + R_0 h \frac{\partial}{\partial z} (\underline{V}_\psi + R_0 \underline{V}_\sigma) \right. \\
 \left. - R_0 (\underline{H}_x + R_0 \underline{H}_y) \nabla \cdot \underline{V}_\sigma - R_0 (\underline{H}_x + R_0 \underline{H}_y) \frac{\partial w}{\partial z} \right\}
 \end{aligned} \tag{4-41}$$

$$R_0^2 \delta \left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla + R_0 w \frac{\partial}{\partial z} \right) h = R_0^2 \delta \left[(\underline{H}_\chi + R_0 \underline{H}_\gamma) \cdot \nabla w - R_0 h \nabla \cdot \underline{V}_\sigma \right] \quad (4-42)$$

$$R_0 \left(\nabla \cdot \underline{H}_\gamma + \frac{\partial h}{\partial z} \right) = 0 \quad (4-43)$$

$$R_0 \left\{ \frac{\left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla \right) \mu r p}{1 + \mu k \epsilon R_0 p} + \frac{\left(\frac{\partial}{\partial t} + (\underline{V}_\psi + R_0 \underline{V}_\sigma) \cdot \nabla \right) \alpha}{1 + k \epsilon R_0 \alpha} \right. \\ \left. \frac{w}{\epsilon} + R_0 w \frac{\partial \theta}{\partial z} \right\} = 0 \quad (4-44)$$

Consistent with the scaling, the sine of the latitude is expanded about a fixed latitude, i.e.

$$\sin \phi = \sin \phi_0 + \frac{L}{a} (y - y_0) \cos \phi_0 \quad (4-45)$$

Defining

$$\beta = \frac{L}{a} \frac{\cos \phi_0}{R_0} \quad (4-46)$$

we may write (4-43) as

$$\sin \phi = \sin \phi_0 + R_0 \beta (y - y_0) \quad (4-47)$$

Then all the dependent variables are expanded in powers of the Rossby number, i.e. in the form

$$X = \sum_{n=0}^{\infty} X^{(n)} R_0^n \quad (4-48)$$

where X represents any of the variables. The so-called "0'th order" equations, representing the geostrophic and hydrostatic balances, are then immediately found from (4-38) and (4-39) to be

$$\sin \phi_0 \underline{k} \times \underline{V}_\psi^{(0)} = -\nabla p^{(0)} \quad (4-49)$$

and

$$\frac{\partial}{\partial z} \frac{p^{(0)}}{\alpha_s^*} = \frac{\alpha^{(0)}}{\alpha_s^*} \quad (4-50)$$

From (4-49) and the definition for \underline{V}_ψ (4-28) we see that the stream function $\psi^{(0)}$ is given by

$$\psi^{(0)} = \frac{p^{(0)}}{\sin \phi_0} \quad (4-51)$$

If we remind ourselves that δ and Ke are much less than unity, and note that $\beta \sim 1$, then we may write the first order equations as

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \underline{V}_\psi^{(0)} \cdot \nabla \right) \underline{V}_\psi^{(1)} + \beta (y - y_0) \underline{k} \times \underline{V}_\psi^{(0)} + \sin \phi_0 \underline{k} \times \underline{V}_\sigma^{(0)} + \sin \phi_0 \underline{k} \times \underline{V}_\psi^{(1)} \\ = -\nabla p^{(1)} + P_s (\nabla \times \underline{H}_X^{(0)}) \times \underline{H}_X^{(0)} \end{aligned} \quad (4-52)$$

$$\nabla \cdot \underline{V}_\sigma^{(0)} + \frac{1}{\rho_s^*} \frac{\partial}{\partial z} (\rho_s^* w^{(0)}) = 0 \quad (4-53)$$

$$\left(\frac{\partial}{\partial t} + \underline{V}_\psi^{(0)} \cdot \nabla\right) \underline{H}_\chi^{(0)} = \underline{H}_\chi^{(0)} \cdot \nabla \underline{V}_\psi^{(0)} \quad (4-54)$$

$$\left(\frac{\partial}{\partial t} + \underline{V}_\psi^{(0)} \cdot \nabla\right) \theta^{(0)} + \frac{w^{(0)}}{\epsilon} = 0 \quad (4-55)$$

Taking the vertical curl, i.e. $\underline{k} \cdot (\nabla \times)$, of (4-52), using the definitions of \underline{V}_ψ and \underline{H}_χ , and (4-53), we may obtain the vorticity equation (here, the superscripts (o) have been dropped for convenience).

$$\left(\frac{\partial}{\partial t} + \underline{V}_\psi \cdot \nabla\right) \nabla^2 \psi + \beta \frac{\partial \psi}{\partial x} - \frac{\sin \phi_0}{\rho_s^*} \frac{\partial}{\partial z} \rho_s^* w = \rho_s (\underline{H}_\chi \cdot \nabla) \nabla^2 \chi \quad (4-56)$$

From (4-32), (4-34), (4-49), (4-50) and (4-51), and the fact that $\kappa \ll 1$, we may show that

$$\theta^{(0)} \approx \frac{\partial \psi^{(0)}}{\partial z} \sin \phi_0 \quad (4-57)$$

Taking the gradient of (4-57) would give us the nondimensional form of the so-called "thermal wind" equation. Using (4-57), we may write (4-55) in the form

$$\left(\frac{\partial}{\partial t} + \underline{V}_\psi \cdot \nabla\right) \frac{\partial \psi}{\partial z} + \frac{w}{\epsilon \sin \phi_0} = 0 \quad (4-58)$$

Equations (4-54), (4-56) and (4-58), together with appropriate boundary conditions, form a closed system. Equation (4-54) is actually redundant in vector form, since only one of the two components is needed

to relate χ to ψ . Thus we need only

$$\left(\frac{\partial}{\partial t} + \underline{v}_\psi \cdot \nabla\right) \frac{\partial \chi}{\partial x} = \underline{H}_\chi \cdot \nabla \frac{\partial \psi}{\partial x} \quad (4-59)$$

For energy computations, however, it is useful to retain (4-54).

Several properties of this set of equations should be noted. We can see that with the presence of the horizontal magnetic fields, the potential vorticity will no longer be conserved. It can be changed by vertical electric currents, if these currents vary along the direction of the horizontal magnetic fields -- which they must, since the current loops must close. As one might expect, the nonconservation of potential vorticity will make it difficult to obtain a useful necessary condition for instability of the vortex, such as was found in the nonmagnetic case by Charney and Stern (1962). This difficulty will be demonstrated in greater detail in Chapter 6.

Although the horizontal electric currents will, in general, be much larger than the vertical currents, they do not enter the horizontal equations of motion to this order. This is because that part of the horizontal component of the electromagnetic body force which arises from the coupling of the horizontal currents and the vertical magnetic field is too small.

Just as in the nonmagnetic case, the vertical motion provides the primary link between the mechanical and thermodynamic processes in the system, acting to keep it in quasi-heliostrophic and hydrostatic balance. The manner of this linkage, however, is modified in several respects by the vertical electric currents, as will be seen more clearly in later chapters.

The vertical magnetic fields do not assume a function comparable to the vertical motion. In fact, they do not enter into the system at all to first order. However, by looking at the second order equation for the vertical field, it can be seen that vertical fields can be produced by the vertical motion. That is, from equation (4-42), collecting terms which are second order in R_0 gives

$$\left(\frac{\partial}{\partial t} + \underline{V}_\psi^{(0)} \cdot \nabla \right) h^{(0)} = \underline{H}_x^{(0)} \cdot \nabla w^{(0)} \quad (4-60)$$

Thus vertical magnetic fields can be produced by the tilting of horizontal fields into the vertical by the vertical motion. In the first approximation, this calculation can be made independently of the other second order equations, since all the required input is of lower order.

The fact that the vertical fields can be separated out in this way obviously represents a considerable mathematical simplification of the problem, while at the same time retaining a means of producing them. The system we have arrived at is a system in which no account is taken of the feedback of the vertical fields on the processes which produced them. While this is probably acceptable in a perturbation study, for long term integrations, it would be a serious drawback, because the system could not act as a dynamo in the usual sense. This will be discussed in the next section.

Boundary conditions and energy conversions

If our scale analysis to obtain the system (4-54), (4-56), (4-58) is consistent, the total energy should be conserved, except for fluxes

across the boundaries. By appropriate choice of boundary conditions, we can eliminate these fluxes. Since we are using a local cartesian coordinate system, we represent a closed latitude circle by taking all variables to be cyclic in λ . Virtually any additional choice of boundary condition is subject to criticism on physical grounds, since the convective layer on the sun does not have any well defined boundaries. Consistent with the scale analysis, however, we should limit the extent of the disturbances in the meridional direction so that they do not cover the whole spherical shell. We therefore introduce rigid, perfectly conducting vertical walls, at $y = 0, l$.

As for boundary conditions at the top and bottom of the layer, about the only guidance we have is the theoretical desirability of isolating the layer energetically from its surroundings. That is, we would prefer to exclude the possibility that work may be done on the layer under consideration by forces originating from adjacent strata. This can be achieved in several ways. One is to bound the layer by rigid, perfectly conducting horizontal walls, taking the vertical velocity and vertical magnetic field to be zero at these walls. If we extend the upper boundary to infinity, we can require that the kinetic and magnetic energy density, and vertical energy flux vanish at infinity. The same could apply to the lower boundary, although the resulting bounds on the stream function would have to be stronger, due to the increasing density.

These boundary conditions would effectively limit the dynamics to a "rotating channel". The closing of electric current loops would be brought about by skin currents in the side walls, and in the top and bottom, if they

are rigid. Such a channel model is obviously not too realistic a representation of the solar problem, but the degree of success attained with such models in simulating the earth's general circulation (e.g. Phillips 1956) suggests that useful results might also be obtained in the solar case.

Mathematically, the side wall boundary conditions are that

$$\frac{\partial \psi}{\partial x}, \frac{\partial \chi}{\partial x}, \frac{\partial^2 \langle \psi \rangle}{\partial y \partial t} = 0 \quad y=0, l \quad (4-65)$$

where the angular brackets denote an average with respect to x . The condition on $\langle \psi \rangle$ is found by averaging the zonal equation of motion at $y=0, l$, and is needed to eliminate the time dependence of the stream function at the walls (Phillips, 1954).

If rigid, perfectly conducting horizontal top and bottom boundaries are assumed at $z = z_1$, and $z = z_2$, the condition is simply that

$$w = 0, h = 0 \quad \text{at} \quad z = z_1, z_2 \quad (4-66)$$

or, from (4-58)

$$\left(\frac{\partial}{\partial t} + \underline{v}_\psi \cdot \nabla \right) \frac{\partial \psi}{\partial z} = 0 \quad (4-67)$$

If the top boundary is extended to infinity, then we must require

$$\begin{aligned} \lim_{z_2 \rightarrow \infty} \int \rho_s^* \left[\left(\frac{\partial \psi}{\partial x} \right)^2 + \left(\frac{\partial \psi}{\partial y} \right)^2 \right] dx &= 0 \\ \lim_{z_2 \rightarrow \infty} \int \rho_s^* P_s \left[\left(\frac{\partial \chi}{\partial x} \right)^2 + \left(\frac{\partial \chi}{\partial y} \right)^2 \right] dx &= 0 \\ \lim_{z_2 \rightarrow \infty} \int \rho_s^* \psi w^{(0)} dx &= 0 \end{aligned} \quad (4-68)$$

Using these boundary conditions, and appropriate integration by parts, we may write the energy balance equations for the system, from equations (4-54), (4-56), and (4-58). From (4-58), after multiplying by $\rho_s^* \epsilon \sin^2 \phi_0 \frac{\partial \psi}{\partial z}$, we may obtain in this way the available potential energy equation

$$\frac{\partial}{\partial t} \int \rho_s^* \frac{\epsilon \sin^2 \phi_0}{2} \left(\frac{\partial \psi}{\partial z} \right)^2 = - \int \rho_s^* \sin \phi_0 w \frac{\partial \psi}{\partial z} \quad (4-69)$$

where the integral is understood to be over a volume of the channel of unit length in the X direction.

Similarly, from (4-56), after multiplying by $\rho_s^* \psi$, we may obtain the kinetic energy balance equation for the region

$$\frac{\partial}{\partial t} \int \rho_s^* \frac{|\nabla \psi|^2}{2} = \int \rho_s^* \sin \phi_0 w \frac{\partial \psi}{\partial z} + \int \rho_s^* P_s \nabla^2 \chi (\nabla \psi \times \nabla \chi) \quad (4-70)$$

Finally, from (4-54), after taking the scalar product with $\rho_s^* P_s \underline{H}_\chi$ we obtain the magnetic energy balance equation

$$\frac{\partial}{\partial t} \int \rho_s^* P_s \frac{|\nabla \chi|^2}{2} = - \int \rho_s^* P_s \nabla^2 \chi (\nabla \psi \times \nabla \chi) \quad (4-71)$$

From adding (4-69)-(4-71), it is evident that the total energy is conserved, i.e.,

$$\frac{\partial}{\partial t} \int \rho_s^* \left[\frac{\epsilon \sin^2 \phi_0}{2} \left(\frac{\partial \psi}{\partial z} \right)^2 + \frac{|\nabla \psi|^2}{2} + P_s \frac{|\nabla \chi|^2}{2} \right] = 0 \quad (4-72)$$

We note that the integral $\int \rho_s^* \sin \phi_0 w \frac{\partial \psi}{\partial z}$ appears in (4-69)

and (4-70) with opposite signs, while the integral

$$\int \rho_s^* P_s \nabla^2 \chi (\nabla \psi \times \nabla \chi) \quad \text{appears in (4-70) and (4-71).}$$

The first integral represents the conversion from available potential to kinetic energy, through the rising of relatively warm and the sinking of relatively cold gas. Alternatively this could have been written as the work done by diverging motions against the pressure. The second integral gives the conversion from kinetic to magnetic energy as the work done by the motion against the horizontal electromagnetic body force. Thus we have a system in which, even in a single disturbance, conversion of energy from available potential to kinetic, whence from kinetic to magnetic energy, is possible. Unfortunately, as intimated earlier, this is not enough to allow the system to act as a dynamo. This can be seen by considering equation (4-54). We may write it in component form as

$$\begin{aligned} \frac{\partial}{\partial x} \frac{\partial \chi}{\partial t} &= \frac{\partial}{\partial x} \left(\frac{\partial \chi}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x} \frac{\partial \chi}{\partial y} \right) \\ \frac{\partial}{\partial y} \frac{\partial \chi}{\partial t} &= \frac{\partial}{\partial y} \left(\frac{\partial \chi}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x} \frac{\partial \chi}{\partial y} \right) \end{aligned} \quad (4-73)$$

Partially integrating each of the equations gives

$$\begin{aligned} \frac{\partial \chi}{\partial t} &= \frac{\partial \chi}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x} \frac{\partial \chi}{\partial y} + D(y) \\ \frac{\partial \chi}{\partial t} &= \frac{\partial \chi}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x} \frac{\partial \chi}{\partial y} + E(x) \end{aligned} \quad (4-74)$$

where D is a function only of y , and E is a function only of x .

Equations (4-74) can be satisfied simultaneously only if

$$D(y) = E(x) = \text{constant.} \quad (4-75)$$

Without loss of generality, we may set this constant equal to zero, in which case (4-74) becomes

$$\frac{\partial \chi}{\partial t} = \frac{\partial \chi}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x} \frac{\partial \chi}{\partial y} \quad (4-76)$$

Equation (4-76) states that the magnetic flux is simply advected by the horizontal flow. But if we multiply (4-76) by χ , and integrate over the volume, applying the boundary conditions (4-65), we can easily show that the averaged square of the flux is an invariant, i.e.

$$\frac{\partial}{\partial t} \int \chi^2 = 0 \quad (4-77)$$

Suppose that magnetic dissipation is then added to the system, i.e., the electrical conductivity in the fluid is no longer infinite. In that case, for any reasonable mathematical representation for the diffusion of flux lines due to the finite conductivity (e.g. proportional to the vector Laplacian of the field strength), the time tendency of the average square of the flux will be negative. That is, (4-77) would become

$$\frac{\partial}{\partial t} \int \chi^2 < 0 \quad (4-78)$$

This result would prevent the system from attaining a statistically steady or quasi-periodic state in which dissipation of the fields by Joule heating

was balanced by conversions from kinetic energy. The magnetic energy present in any initial state would always be lost to heat at a faster rate than it could be resupplied. Thus the initial magnetic fields could not be maintained.

The invariance of the mean square horizontal flux with infinite conductivity, and its decay with finite conductivity, are consequences of the fact that our scaling restrictions have ruled out feedbacks from the vertical fields. These feedbacks are essential to regenerate the horizontal fields, i.e. to complete the dynamo cycle. Therefore, in long term integrations, such as we might eventually hope to use to simulate a solar cycle, these feedbacks from the vertical fields must be included. But it is encouraging that the system in its present form seems capable of producing the vertical fields needed for the feedbacks. Finally, it should be pointed out that time integrations of intermediate length, for which dissipation by Joule heating might be neglected, could still yield useful information about the growth and structure of the large scale disturbances.

5. Equations for perturbations about an initially steady state

We wish to study the stability properties, structure, and energetics of small disturbances on an initially steady hydromagnetic circumpolar vortex. Therefore we need to split the dependent variables into two parts, a zonally averaged part, (an average in \mathbf{x} , which is denoted by an angular brackets) and a deviation from that average. For the perturbation analysis to be valid, we must begin with a zonally averaged state that is steady in the absence of any perturbation. When magnetic fields are present, this state has to satisfy Ferraro's Law of Isorotation (c.f. Chapter 2), namely, that the magnetic flux lines and the streamlines must everywhere be parallel. Thus we wish to perturb an initially steady state in which we have only a zonal component of velocity, $\langle U \rangle = -\partial\langle\psi\rangle/\partial y$, and only a zonal (toroidal) component of magnetic field, $\langle H \rangle = -\partial\langle\chi\rangle/\partial y$, and a mean potential temperature field, $\langle\theta\rangle = \partial\langle\psi\rangle/\partial z \sin\phi_0$, which are functions of y and z . The changes in the initial flow, temperature gradients, and magnetic field as the disturbances grow can be examined with the aid of the second order equations for their time derivatives.

Expanding about this mean state, then, using primes to denote the deviations from the mean state, the vorticity, thermodynamic, and magnetic field equations for the perturbations may be written from (4-56), (4-58), (4-54) respectively, in the forms

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x}\right) \nabla^2 \psi' + \left(\beta - \frac{\partial^2 \langle U \rangle}{\partial y^2}\right) \frac{\partial \psi'}{\partial x} - \frac{\sin\phi_0}{\rho_s^*} \frac{\partial}{\partial z} \rho_s^* w' \\ = P_s \langle H \rangle \nabla^2 \frac{\partial \chi'}{\partial x} - P_s \frac{\partial^2 \langle H \rangle}{\partial y^2} \frac{\partial \chi'}{\partial x} \end{aligned} \quad (5-1)$$

$$\left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x}\right) \frac{\partial \psi'}{\partial z} - \frac{\partial \langle U \rangle}{\partial z} \frac{\partial \psi'}{\partial x} + \frac{w'}{\epsilon \sin \phi_0} = 0 \quad (5-2)$$

$$\left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x}\right) \frac{\partial \chi'}{\partial x} = \langle H \rangle \frac{\partial^2 \psi'}{\partial x^2} \quad (5-3)$$

with the redundant equation

$$\left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x}\right) \frac{\partial \chi'}{\partial y} = \frac{\partial \langle H \rangle}{\partial y} \frac{\partial \psi'}{\partial x} + \langle H \rangle \frac{\partial^2 \psi'}{\partial x \partial y} - \frac{\partial \langle U \rangle}{\partial y} \frac{\partial \chi'}{\partial x} \quad (5-4)$$

to be used for energy calculations.

The time rates of change of the initial meridional temperature gradient, the initial zonal flow, and the initial toroidal magnetic field are found by retaining the time derivatives of the zonally averaged variables and the products of perturbation quantities in the expansions of (4-54), (4-56), and (4-58), and then averaging the expanded equations with respect to \mathbf{x} , thereby eliminating all the terms that are linear in perturbations. In this way (analogous to Phillips, 1954) we obtain the tendency equations for the zonally averaged state in the form

$$\frac{\partial \langle \theta \rangle}{\partial t \sin \phi_0} = - \frac{\partial}{\partial y} \left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial z} \right\rangle - \frac{\langle w \rangle}{\epsilon \sin \phi_0} \quad (5-5)$$

$$\begin{aligned} \frac{\partial \langle U \rangle}{\partial t} = \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial y} \right\rangle - \frac{\rho_s^2}{\rho_s^*} \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \chi'}{\partial x} \frac{\partial \chi'}{\partial y} \right\rangle \\ - \frac{\sin \phi_0}{\rho_s^*} \frac{\partial}{\partial z} \rho_s^* \langle w \rangle \end{aligned} \quad (5-6)$$

$$\frac{\partial}{\partial t} \langle H \rangle = - \frac{\partial}{\partial y} \left\langle \frac{\partial \chi'}{\partial x} \frac{\partial \psi'}{\partial y} - \frac{\partial \chi'}{\partial y} \frac{\partial \psi'}{\partial x} \right\rangle \quad (5-7)$$

Expanding (4-60) we may also obtain an equation for the perturbation vertical magnetic field h' , which is given by

$$\left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x} \right) h' = \langle H \rangle \frac{\partial w'}{\partial x} \quad (5-8)$$

By retaining the products of perturbation quantities, and $\frac{\partial \langle h \rangle}{\partial t}$ in the expansion of (4-60) and then averaging with respect to x , we can get a tendency equation for the mean vertical magnetic field $\langle h \rangle$, which is given by

$$\frac{\partial \langle h \rangle}{\partial t} = \frac{\partial}{\partial y} \left\langle w' \frac{\partial \chi'}{\partial x} \right\rangle - \frac{\partial}{\partial y} \left\langle h' \frac{\partial \psi'}{\partial x} \right\rangle \quad (5-9)$$

The boundary conditions at the side walls, from (4-63) become

$$\frac{\partial \psi'}{\partial x}, \frac{\partial \chi'}{\partial x}, \frac{\partial \langle U \rangle}{\partial t} = 0, \quad y = 0, l \quad (5-10)$$

For rigid, perfectly conducting top and bottom boundaries, the conditions are, from (4-67) (or from 5-2))

$$\left(\frac{\partial}{\partial t} + \langle U \rangle \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \frac{\partial \langle U \rangle}{\partial z} \frac{\partial \psi'}{\partial x} = 0, \quad z = z_1, z_2 \quad (5-11)$$

If the top boundary is taken to infinity, from (4-68) the kinetic and magnetic energy densities, and the vertical energy propagation must vanish separately for the initial steady state and for the perturbations.

Using these boundary conditions, we can as before eliminate the boundary fluxes of energy. Each form of energy will now have a zonally symmetric and an eddy component, so that there will be six types to consider. For convenience, let the volume integrals of the available potential, kinetic, and magnetic energies be represented respectively by A , K , M , and the various conversion terms by the operator $(\ , \)$. Then the energy balance equations may be written as

$$\frac{\partial}{\partial t} \langle A \rangle = -(\langle A \rangle, A') + (\langle K \rangle, \langle A \rangle) \quad (5-12)$$

$$\frac{\partial}{\partial t} A' = (\langle A \rangle, A') - (A', K') \quad (5-13)$$

$$\frac{\partial}{\partial t} \langle K \rangle = -(\langle K \rangle, \langle A \rangle) + (K', \langle K \rangle) - (\langle K \rangle, M') \quad (5-14)$$

$$\frac{\partial}{\partial t} K' = -(K', \langle K \rangle) + (A', K') - (K', M') - (K', \langle M \rangle) \quad (5-15)$$

$$\frac{\partial}{\partial t} \langle M \rangle = (K', \langle M \rangle) \quad (5-16)$$

$$\frac{\partial}{\partial t} M' = (\langle K \rangle, M') + (K', M') \quad (5-17)$$

The six forms of energy are defined as (the integral is again understood to be over the channel volume with unit length in the x direction)

$$\begin{aligned} \langle A \rangle &= \int \rho_s^* \frac{\epsilon}{2} \langle \theta \rangle^2 & ; & \quad A' = \int \rho_s^* \frac{\epsilon \sin^2 \phi_0}{2} \left\langle \left(\frac{\partial \psi'}{\partial z} \right)^2 \right\rangle \\ \langle K \rangle &= \int \rho_s^* \frac{\langle U \rangle^2}{2} & ; & \quad K' = \int \frac{\rho_s^*}{2} \langle (\nabla \psi')^2 \rangle \\ \langle M \rangle &= \int \rho_s^* P_s \frac{\langle H \rangle^2}{2} & ; & \quad M' = \int \frac{\rho_s^*}{2} P_s \langle (\nabla \chi')^2 \rangle \end{aligned} \quad (5-18)$$

In equations (5-12)-(5-17) all the conversion terms are defined so that they are positive when there is a conversion of energy from the first form in the parenthesis to the second. The definitions of these conversions then are:

$$(\langle K \rangle, \langle A \rangle) = -\int \rho_s^* \langle w \rangle \langle \theta \rangle \quad (5-19)$$

$$(\langle A \rangle, A') = -\int \rho_s^* \epsilon \sin \phi_0 \left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial z} \right\rangle \frac{\partial \langle \theta \rangle}{\partial y} \quad (5-20)$$

$$(A', K') = \int \rho_s^* \sin \phi_0 \langle w' \frac{\partial \psi'}{\partial z} \rangle \quad (5-21)$$

$$(K', \langle K \rangle) = -\int \rho_s^* \left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial y} \right\rangle \frac{\partial \langle U \rangle}{\partial y} \quad (5-22)$$

$$(\langle K \rangle, M') = -\int \rho_s^* P_s \left\langle \frac{\partial \chi'}{\partial x} \frac{\partial \chi'}{\partial y} \right\rangle \frac{\partial \langle U \rangle}{\partial y} \quad (5-23)$$

$$(K', M') = -\int \rho_s^* P_s \left\langle \frac{\partial \psi'}{\partial x} \nabla^2 \chi' \right\rangle \langle H \rangle \quad (5-24)$$

$$(K', \langle M \rangle) = -\int \rho_s^* P_s \left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \chi'}{\partial y} - \frac{\partial \psi'}{\partial y} \frac{\partial \chi'}{\partial x} \right\rangle \frac{\partial \langle H \rangle}{\partial y} \quad (5-25)$$

To better visualize the energy conversion processes, we represent equations (5-12)-(5-17) by the energy flow diagram (analogous to Phillips 1956) pictured in Fig. 1. From the definitions (5-19)-(5-22), we see that a positive value of:

a) $(\langle A \rangle, A')$ arises from the meridional eddy flux of heat

from warm to cold latitudes, which is manifested by a phase difference between the perturbation streamlines and isotherms.

b) (A', K') is due to the rising of warm gas and sinking of cold gas at different longitudes.

c) $(K', \langle K \rangle)$ is due to the eddy flux of momentum by the horizontal Reynold's stresses against the gradient of mean zonal momentum, requiring the perturbation streamlines to tilt upstream away from the maximum of mean zonal momentum.

d) $(\langle K \rangle, \langle A \rangle)$ arises from the net sinking of warm gas and rising of cold gas due to the mean meridional motions.

The top part of Fig. 1, including only these four conversions, represents the energy cycle as it is known to exist in the earth's atmosphere, the four conversions having been shown from calculations based on extensive observations to be in the directions indicated by the arrows. The measurement of these conversions has recently been reviewed by Oort (1964). For the sun, it has been possible so far to estimate only

$(K', \langle K \rangle)$ which, as discussed in Chapter 3, was done by Starr and Gilman (1965a). Using sunspot displacement and correlation statistics of

ENERGY BALANCE

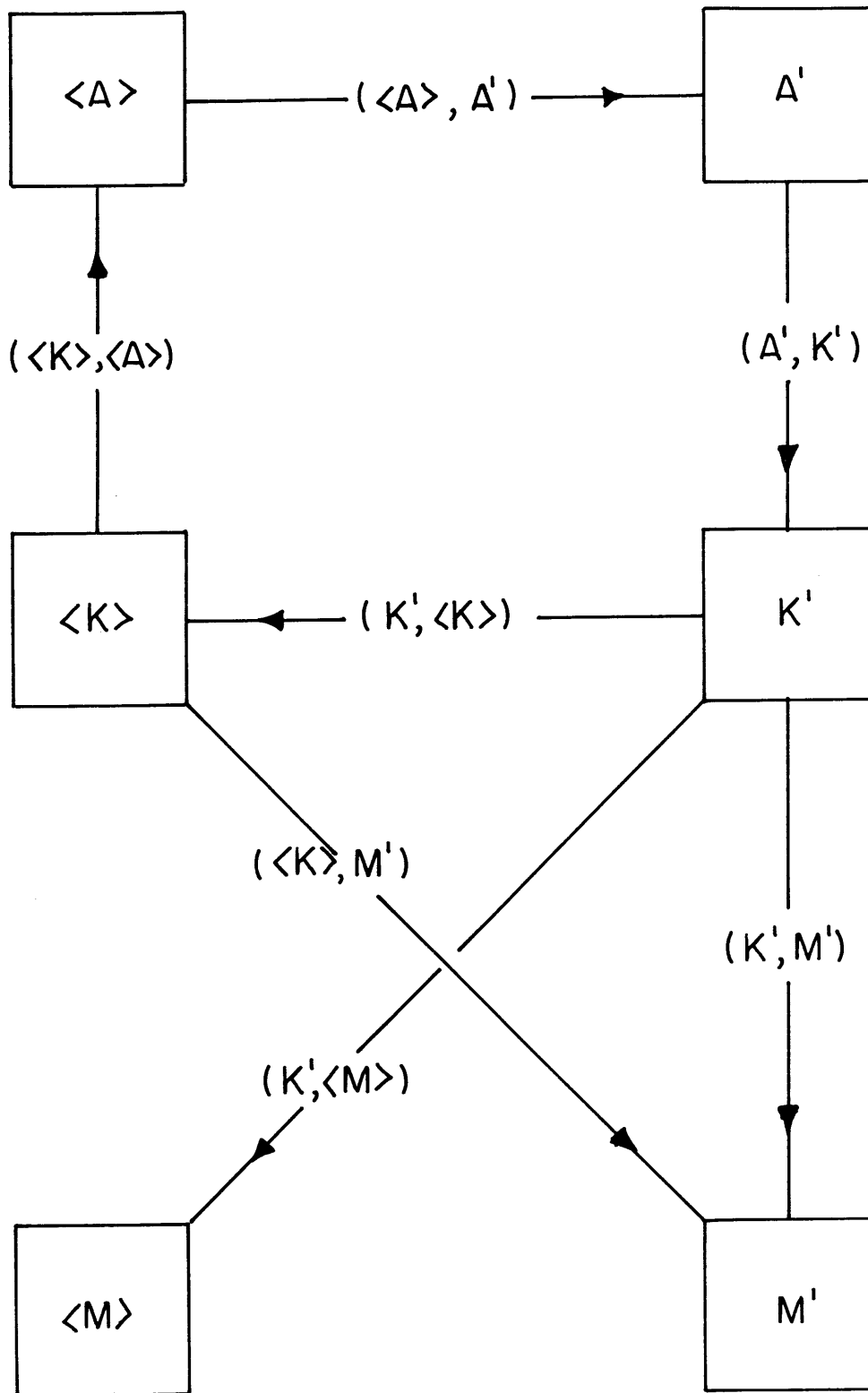


Fig. 1. Energy flow diagram for scaled system of equations. A, K, M, denote available potential, kinetic, and magnetic energies respectively. Angular brackets denote energy of zonally symmetric component; primes denote energy of eddy component. Arrows are in the direction of a positive value for conversion designated. For mathematical form of conversions, see equations (5-19)-(5-25), and also (7-16)-(7-22).

Ward (1964), they found a conversion from eddy to mean zonal kinetic energy just as in the earth's atmosphere, but requiring a smaller number of rotations over which dissipation must act to give balance than is needed for the earth.

Turning to the lower part of Fig. 1, the conversions between kinetic and magnetic energy, from the definitions (5-23)-(5-25) we see that a positive value of:

e) $(\langle K \rangle, M')$ is due to a transport of momentum by the horizontal Maxwell stresses in the direction of the gradient of momentum, requiring a tilt in the perturbation magnetic flux lines in the same sense as for the streamlines in c). That such tilts would probably be separate manifestations of the same hydromagnetic disturbance, and that the horizontal Maxwell stresses would then act as a brake on the differential rotation, was suggested by Starr and Gilman (1965b) (c.f. Chapter 3).

f) (K', M') arises from the work done by the horizontal eddy motions against that part of the electromagnetic body force which is due to the coupling between vertical perturbation currents and the mean zonal magnetic field. It thus requires a phase difference between the perturbation stream lines and perturbation magnetic flux lines.

g) $(K', \langle M \rangle)$ arises from the work done by the horizontal eddy motions against that part of the electromagnetic body force due to the coupling between the mean vertical electric current, associated with the mean toroidal field, and the horizontal eddy magnetic fields. This cross correlation of the eddy flow and magnetic fields, taking a functional form similar to the Maxwell and Reynolds stresses, might conveniently be

called a "mixed" stress. This conversion also requires a phase shift between the perturbation streamlines and flux lines.

In Chapter 2, we sketched the present state of hydromagnetic dynamo theory and its application to the solar cycle problem. It is possible in principle to discuss the physical processes comprising this theory in the context of energy balance equations like equations (5-12)-(5-25) and Fig. 1. If we were to do this, we would find that (5-19)-(5-25) do not contain all of the conversions present in the dynamo theory, and also that they contain a few not present there. Obviously, due to the scaling restrictions imposed in our system, we do not have conversions corresponding to the "kink instability" or "magnetic buoyancy" of the magnetic flux tubes, or to the dragging up of such flux tubes by the granulation or supergranulation scale motions. These have been replaced by a second order process, represented in equations (4-60) and (5-8), namely, the tipping of the large scale horizontal fields into the vertical by the large scale vertical motion. We have thus enlarged the horizontal scale on which the vertical flux loops are being produced. While this may be at variance with the actual mechanisms for generating such loops on the sun, as suggested by the observations, it has the considerable advantage of allowing us to deal with them concurrently with other aspects of the large scale dynamics since they have the same time and space scales. This will be particularly important when more advanced models are developed, in which the feedbacks of these vertical fields are included.

Our system also contains a higher order effect which corresponds to the "drift currents" postulated by Babcock (1961) to bring one branch of the vertical flux loops to the polar regions. This process is defined by the second term on the right in equation (5-9). The first term on the

right defines a process not explicitly included in dynamo theory, namely the production of mean poloidal fields by eddy vertical transports of eddy meridional fields.

To first order, our system does not contain a process analogous to the one conventionally used to begin the dynamo process, namely, the twisting of the mean meridional (poloidal) magnetic field into a mean toroidal field by the differential rotation. This is the reason for the absence of a connection between $\langle K \rangle$ and $\langle M \rangle$ in Fig. 1. It is present to second order, from an expansion of equations (4-41) and (4-42), for example, but obviously would be quite difficult to calculate.

The reason this process is not included to the first order is clearly the quasi-nondivergent nature of the motion and, as we showed in Chapter 4, consequently of the magnetic fields. If the assumption that $L/a \ll 1$, from which this property follows, is relaxed, requiring the use of spherical coordinates, the twisting of the poloidal field into a toroidal one will be present to first order. Since for $L/a \sim 1$, the advective time scale $L/U \sim 200$ days, so that to describe long period changes, such as the progression of the solar cycle, we would have to include this twisting process.

Finally, we point out that present dynamo theory does not include a conversion of the form (5-25). That is, the maintenance of the toroidal field is looked upon as a purely axisymmetric process, comprising the twisting from the poloidal field which we just discussed. While it is true that the conversion (5-25) cannot accomplish the maintenance alone, as implied from the discussion at the end of Chapter 4, it can still make a significant contribution to the total. When both processes are included, as they would be if $L/a \sim 1$ it is difficult to tell which would be the more important.

6. Integral theorems

In this chapter we prove several theorems concerning the stability of the hydromagnetic circumpolar vortex with respect to asymmetric disturbances. The techniques follow closely those of Pedlosky (1964a) for the nonmagnetic case.

Relation between stresses and symmetric state required for instability

The first theorem follows immediately from the energy balance equations (5-13), (5-15), and (5-17). By definition, the perturbation on the vortex will grow only if

$$\frac{\partial}{\partial t}(A' + K' + M') > 0 \tag{6-1}$$

But by simply adding equations (5-13), (5-15), and (5-17), using the definitions (5-18)-(5-25), dropping the angular brackets and primes in the conversion terms without confusion, we obtain

$$\begin{aligned} \frac{\partial}{\partial t}(A' + K' + M') = \int \rho_s^* \left[-\epsilon \sin \phi_0 \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial z} \frac{\partial \theta}{\partial y} \right. & \tag{6-2} \\ & + \left(\frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial y} - P_s \frac{\partial x}{\partial x} \frac{\partial x}{\partial y} \right) \frac{\partial U}{\partial y} \\ & \left. - P_s \left(\frac{\partial x}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial x}{\partial y} \frac{\partial \psi}{\partial x} \right) \frac{\partial H}{\partial y} \right] \end{aligned}$$

Therefore, for the perturbations to grow, we must have that

$$\begin{aligned} \int \rho_s^* \left[-\epsilon \sin \phi_0 \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial z} \frac{\partial \theta}{\partial y} + \left(\frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial y} - P_s \frac{\partial x}{\partial x} \frac{\partial x}{\partial y} \right) \frac{\partial U}{\partial y} \right. & \tag{6-3} \\ & \left. - P_s \left(\frac{\partial x}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial x}{\partial y} \frac{\partial \psi}{\partial x} \right) \frac{\partial H}{\partial y} \right] > 0 \end{aligned}$$

In the absence of a toroidal magnetic field, this of course reduces to Pedlosky's (1964a) result, namely

$$\int \rho_s^* \left[-\epsilon \sin \phi_0 \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial z} \frac{\partial \theta}{\partial y} + \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial y} \frac{\partial v}{\partial y} \right] > 0 \quad (6-4)$$

Comparing (6-4) with (6-3) can give us some insight into the role played by the magnetic field. Suppose that there is no horizontal shear in the toroidal field, i.e. $\partial H / \partial y = 0$. Then the last term in (6-3) vanishes. Consider, then, a disturbance which is baroclinically unstable, but barotropically stable, such as found, for example, by Pedlosky (1964b). For this kind of disturbance, in the nonmagnetic case, (6-4) will be satisfied by the first term giving a larger positive contribution than the second gives a negative one. That is, for the perturbations to grow, the conversion from zonal to eddy available potential energy through down-gradient eddy heat transport, must be greater than the resulting conversion from eddy kinetic to zonal kinetic energy, achieved through upgradient momentum transport by the Reynolds stress. When the magnetic field is added, the resulting Maxwell stress in the perturbation will tend to have the opposite sign to the Reynolds stress, since due to the infinite conductivity the magnetic flux lines should distort in the same way as the streamlines. Thus the conversion to mean zonal kinetic energy will be reduced. Consequently the Maxwell stresses should have a destabilizing effect on this kind of flow. This is borne out in Chapter 10 by calculations for a parabolic flow in a uniform field, the simplest hydromagnetic generalization of Pedlosky's (1964b) solution.

Conversely, it would appear from (6-3) that toroidal fields with no horizontal shear would, through the Maxwell stresses, tend to stabilize flows which were barotropically unstable but baroclinically stable.

Relation (6-3) would also appear to suggest the existence of an instability in which "available magnetic energy" associated with horizontal shears in the toroidal field would be released into kinetic energy and even potential energy. However, the bounds obtained on growth rates and phase velocities of unstable waves later in this chapter argue against this possibility.

Necessary conditions for instability

Charney and Stern (1962) and later, with some increase in generality, Pedlosky (1964a) derived necessary conditions for instability of a baroclinic circumpolar vortex to wavelike disturbances. These conditions involved the zonal flow and its potential vorticity. In particular, one condition was that if the vertical shear of the zonal flow vanishes on the top and bottom boundaries (or its kinetic energy density vanishes at the top if the top extends to infinity), the flow could be unstable if the potential vorticity gradient changed sign somewhere in the region. In this section we attempt to generalize this theorem to the magnetic case, showing that the same condition is no longer necessary. Unfortunately, the new necessary condition does not appear to have any simple or useful interpretation.

Between equations (5-1) and (5-2) we may eliminate the vertical velocity, forming the so-called potential vorticity equation, written in the form

$$\begin{aligned} & \left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{1}{\rho_s^*} \frac{\partial}{\partial z} \left(\rho_s^* \epsilon \sin^2 \phi_0 \frac{\partial \psi}{\partial z} \right) \right) \\ & + \frac{\partial q}{\partial y} \frac{\partial \psi}{\partial x} = P_s H \left(\frac{\partial^3 \chi}{\partial x^3} + \frac{\partial^3 \chi}{\partial x \partial y^2} \right) - P_s \frac{\partial^2 H}{\partial y^2} \frac{\partial \chi}{\partial x} \end{aligned} \quad (6-5)$$

where

$$\frac{\partial q}{\partial y} = \beta - \frac{\partial^2 U}{\partial y^2} - \frac{1}{\rho_s^*} \frac{\partial}{\partial z} \left(\rho_s^* \epsilon \sin^2 \phi_0 \frac{\partial U}{\partial z} \right) \quad (6-6)$$

is the gradient of potential vorticity for the flow.

Since all the coefficients in (6-5) are independent of x, t , we may write a separated solution in the form

$$\psi; \chi = \left[\tilde{\psi}(y, z); \tilde{\chi}(y, z) \right] e^{ik(x-ct)} \quad (6-7)$$

corresponding to a traveling wave with wave number $k = 2\pi/\lambda$, and (complex) phase velocity c , i.e. $c = c_r + i c_i$.

Substitution of the solutions (6-7) into (6-5) then gives

$$\begin{aligned} & (U-c) \left(-k^2 \tilde{\psi} + \frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{1}{\rho_s^*} \frac{\partial}{\partial z} \left(\rho_s^* \epsilon \sin^2 \phi_0 \frac{\partial \tilde{\psi}}{\partial z} \right) \right) \\ & + \frac{\partial q}{\partial y} \tilde{\psi} = P_s H \left(\frac{\partial^2 \tilde{\chi}}{\partial y^2} - k^2 \tilde{\chi} \right) - P_s \frac{\partial^2 H}{\partial y^2} \tilde{\chi} \end{aligned} \quad (6-8)$$

To complete the system, $\tilde{\Psi}$ may be related to $\tilde{\chi}$ by substituting from (6-7) into (5-3), giving

$$(U-c)\tilde{\chi} = H\tilde{\Psi} \quad (6-9)$$

Let us assume that $C_i \neq 0$. Then we may divide by $(U-c)$.

Using (6-9), and noting that

$$H \frac{\partial^2 \tilde{\chi}}{\partial y^2} - \frac{\partial^2 H}{\partial y^2} \tilde{\chi} = \frac{\partial}{\partial y} \left(H \frac{\partial \tilde{\chi}}{\partial y} - \frac{\partial H}{\partial y} \tilde{\chi} \right)$$

we may write the right hand side of (6-8) in the form

$$P_s \frac{\partial}{\partial y} \left[H^2 \frac{\partial}{\partial y} \left(\frac{\tilde{\Psi}}{U-c} \right) \right] - \frac{k^2 P_s H^2 \tilde{\Psi}}{(U-c)}$$

Then we may write (6-8) as

$$\begin{aligned} \frac{\partial^2 \tilde{\Psi}}{\partial y^2} + \frac{1}{\rho_s^*} \frac{\partial}{\partial z} \left(\epsilon \rho_s^* \sin^2 \phi_0 \frac{\partial \tilde{\Psi}}{\partial z} \right) - k^2 \tilde{\Psi} \\ + \frac{\partial \rho}{\partial y} \frac{\tilde{\Psi}}{U-c} = \frac{P_s}{U-c} \frac{\partial}{\partial y} \left(H^2 \frac{\partial}{\partial y} \frac{\tilde{\Psi}}{U-c} \right) - \frac{k^2 P_s H^2 \tilde{\Psi}}{(U-c)^2} \end{aligned}$$

(6-10)

In passing, we note that if we took $\epsilon = 0$ (infinite static stability) and $\partial \psi / \partial y = 0$, equation (6-10) would yield ordinary Alfvén waves. If the β effect were retained, some combination of Rossby and Alfvén waves would result. This case has been studied by Hide (1964). His results should be extendible to include static stability by keeping $\epsilon = \text{const} \neq 0$. This is planned for later work.

The side wall boundary conditions (5-10), after substituting from (6-7) and recalling the relation (6-9), become simply

$$\tilde{\Psi} = 0, \quad y = 0, l \quad (6-11)$$

while the top and bottom boundaries, if rigid, give, from (5-11)

$$(U-C) \frac{\partial \tilde{\Psi}}{\partial z} - \frac{\partial U}{\partial z} \tilde{\Psi} = 0, \quad z = z_1, z_2 \quad (6-12)$$

If the top is taken to infinity, the condition is

$$\lim_{z_2 \rightarrow \infty} \tilde{\Psi}_* \frac{\partial \tilde{\Psi}}{\partial z} = 0 \quad (6-13)$$

where the subscript asterisk denotes the complex conjugate.

If we now multiply (6-10) by $\rho_S^* \tilde{\Psi}_*$, and integrate by parts, applying the boundary conditions just stated, we have

$$\begin{aligned}
 & \int_0^l dy \int_{z_1}^{z_2} dz \rho_s^* \left(k^2 |\tilde{\Psi}|^2 + \left| \frac{\partial \tilde{\Psi}}{\partial y} \right|^2 + \epsilon \sin^2 \phi_0 \left| \frac{\partial \tilde{\Psi}}{\partial z} \right|^2 \right) \\
 &= - \int_0^l dy \frac{\epsilon \rho_s^* \sin^2 \phi_0}{U-C} \frac{\partial U}{\partial z} |\tilde{\Psi}|^2 \Big|_{z=z_1}^{z=z_2} \\
 &+ \int_0^l dy \int_{z_1}^{z_2} dz \rho_s^* \left(\frac{\partial g}{\partial y} + \frac{P_s H^2 k^2}{(U-C)^2} + \frac{P_s H^2 \left(\frac{\partial U}{\partial y} \right)^2}{(U-C)^4} \right) |\tilde{\Psi}|^2 \\
 &+ \int_0^l dy \int_{z_1}^{z_2} dz \rho_s^* \frac{P_s H^2}{(U-C)^2} \left| \frac{\partial \tilde{\Psi}}{\partial y} \right|^2 \tag{6-14} \\
 &- \int_0^l dy \int_{z_1}^{z_2} dz \rho_s^* \frac{P_s H^2 \frac{\partial U}{\partial y}}{(U-C)^3} \frac{\partial}{\partial y} |\tilde{\Psi}|^2
 \end{aligned}$$

From (6-14), it is evident that setting $H=0$ everywhere, we may recover the necessary conditions for instability found by Pedlosky (1964a). With the magnetic field present it is obvious that these conditions will no longer hold, so that instability could occur even if $\partial g / \partial y$ does not change sign. But unfortunately, rather little further information can be obtained from (6-14). To see this, let us separate real and imaginary parts of (6-14) which must separately equate. For the imaginary part,

we have

$$\begin{aligned}
 C_i \left\{ \int_0^l dy \frac{\epsilon \rho_s^* \sin^2 \phi_0}{|U-C|^2} \frac{\partial U}{\partial z} |\tilde{\Psi}|^2 \Big|_{z_1}^{z_2} + \int_0^l dy \int_{z_1}^{z_2} dz \frac{|\tilde{\Psi}|^2}{|U-C|^2} \frac{\partial q}{\partial y} \right. & \quad (6-15) \\
 + \int_0^l dy \int_{z_1}^{z_2} \rho_s^* P_s H^2 \left[2(U-C_r) \left(\frac{k^2 |\tilde{\Psi}|^2 + \left| \frac{\partial \tilde{\Psi}}{\partial y} \right|^2}{|U-C|^4} + \frac{2[(U-C_r)^2 + C_i^2] \left(\frac{\partial U}{\partial y} \right)^2}{|U-C|^8} \right. \right. \\
 \left. \left. - \frac{[3(U-C_r)^2 - C_i^2] \frac{\partial U}{\partial y} \frac{\partial}{\partial y} |\tilde{\Psi}|^2}{|U-C|^6} \right) \right] & \quad \left. \right\}
 \end{aligned}$$

Then if $C_i \neq 0$, the necessary condition for instability is that the expression in brackets vanish. This condition thus depends explicitly on the phase velocity of the disturbance, and also on its structure (from the term containing $\frac{\partial}{\partial y} |\tilde{\Psi}|^2$) which makes its usefulness limited.

We can, however, say more about disturbances of short wavelength. There is just one term within the brackets in (6-15) that is proportional to k^2 . It can be written as

$$2 k^2 \int_0^l dy \int_{z_1}^{z_2} dz \frac{\rho_s^* P_s H^2 (U-C_r) |\tilde{\Psi}|^2}{|U-C|^4} \quad (6-16)$$

We would thus expect the magnetic field to dominate in determining the stability criteria at short wavelength. In the specific cases studied in Chapters 8 and 9, we shall see that this is indeed the case.

In closing, we note that the real part of (6-15), which we omit here, tells us only, as we might expect, that the positive correlation between the zonal flow and potential vorticity gradient required for instability is not necessarily required in the magnetic case.

Bounds on phase speeds and growth rates of unstable waves.

In a manner very similar to that of Pedlosky (1964a), we may place bounds on the phase speeds and growth rates of the unstable waves. Beginning from equations (6-8) and (6-9), we introduce the transformation

$$\tilde{\Psi} = (U-C)G \quad (6-17)$$

Then, from (6-9), $\tilde{\chi}$ is related to the new function G according to

$$\tilde{\chi} = HG \quad (6-18)$$

Then substituting (6-17) and (6-18) into (6-8), rearranging some derivatives, we can obtain

$$\begin{aligned} \frac{\partial}{\partial y} \left[(U-C)^2 \frac{\partial G}{\partial y} \right] + \frac{1}{\rho_s^*} \frac{\partial}{\partial z} \left[(U-C)^2 \rho_s^* \sin^2 \phi_0 \frac{\partial G}{\partial z} \right] \\ - k^2 (U-C)^2 G + \beta (U-C) G = P_s \frac{\partial}{\partial y} \left[H^2 \frac{\partial G}{\partial y} \right] - k^2 P_s H^2 G \end{aligned} \quad (6-19)$$

The boundary conditions transform to

$$\begin{aligned} G = 0, \quad y = 0, l \\ \frac{\partial G}{\partial z} = 0, \quad z = z_1, z_2 \end{aligned} \quad (6-20)$$

We then multiply (6-19) by $\rho_s^* G_x$, and integrate over the channel cross-section, integrating by parts once, applying the boundary conditions (6-20) to eliminate boundary terms. If we define

$$\begin{aligned} \rho_s^* \left(\left| \frac{\partial G}{\partial y} \right|^2 + k^2 |G|^2 \right) &\equiv Q \\ \rho_s^* \epsilon \sin^2 \phi_0 \left| \frac{\partial G}{\partial z} \right|^2 &\equiv R \\ \rho_s^* |G|^2 &\equiv S \end{aligned} \tag{6-21}$$

then the result may be written in the form

$$\int (U-C)^2 (Q+R) = \int (U-C) \beta S + \int P_s H^2 Q \tag{6-22}$$

where again the integral is understood to be over the channel cross-section of unit length in the x direction. The last integral in (6-22), namely $\int P_s H^2 Q$, represents the only addition due to the magnetic field.

Now the real and imaginary parts of (6-22) must separately equate.

The imaginary part gives

$$C_i \left[\beta \int S - \int (U-C_r)(Q+R) \right] = 0 \tag{6-23}$$

since $C_i \neq 0$, we must have

$$\int (U-C_r)(Q+R) = \beta \int S \tag{6-24}$$

This is precisely the condition arrived at by Pedlosky (1964a) for the non-magnetic case. Therefore the phase velocities have the same bounds as in the nonmagnetic case, namely,

$$U_{\min} - \frac{\beta/2}{k^2 + \pi^2/l^2} \leq c_r \leq U_{\max} \quad (6-25)$$

where U_{\max} and U_{\min} are the maximum and minimum velocities in the flow, respectively.

The real part of (6-22) gives

$$\int [(U - c_r)^2 - c_i^2](Q + R) = \int (U - c_r) \beta S + \int P_s H^2 Q \quad (6-26)$$

If we follow exactly the procedure used by Pedlosky (1964a) for all the non-magnetic terms, we can easily show that

$$0 \geq \left[\left(c_r - \frac{U_{\max} + U_{\min}}{2} \right)^2 + c_i^2 - \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{\max} - U_{\min}}{2} - \left(\frac{U_{\max} - U_{\min}}{2} \right)^2 \right] \int (Q + R) + \int P_s H^2 Q \quad (6-27)$$

but since $\int P_s H^2 Q$ is positive definite,

$$0 \geq \left(c_r - \frac{U_{\max} + U_{\min}}{2} \right)^2 + c_i^2 - \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{\max} - U_{\min}}{2} - \left(\frac{U_{\max} - U_{\min}}{2} \right)^2 \quad (6-28)$$

Thus the complex phase velocity must lie inside the same semicircular area as in the nonmagnetic case. From (6-28), the bound on the growth rate is still

$$C_i^2 \leq \left(\frac{U_{max} - U_{min}}{2} \right)^2 + \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{max} - U_{min}}{2} \quad (6-29)$$

The inequality (6-27) suggests that in the magnetic case, the actual growth rates are probably less than in the nonmagnetic case. This is certainly true if the flow is barotropic. In this case, $R \equiv 0$, so that (6-27) may be reduced to

$$0 \geq \left(C_r - \frac{U_{max} + U_{min}}{2} \right)^2 + C_i^2 - \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{max} - U_{min}}{2} - \left(\frac{U_{max} - U_{min}}{2} \right)^2 + (P_s H^2)_{min} \quad (6-30)$$

Thus in the barotropic case, the square of the radius of the semicircle in which the complex phase velocity must lie is reduced by an amount

$$(P_s H^2)_{min}$$

due to the presence of the magnetic field. If this field is strong enough, all disturbances will be rendered neutrally stable, i.e., if

$$(P_s H^2)_{min} \geq \left[\left(\frac{U_{max} - U_{min}}{2} \right)^2 + \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{max} - U_{min}}{2} \right] \quad (6-31)$$

The inequality (6-30), of course, does not say that the magnetic field can never be destabilizing, since the bounds set are not necessarily

sharp. Examples of destabilization of barotropic flow by magnetic fields can be found in the literature, e.g. Stern (1963).

It is interesting to note that if we set $\beta = 0$, (6-30) has precisely the same mathematical form as the semicircle theorem proved by Howard (1961) for normal mode disturbances in two dimensional incompressible stratified shear flow. In that case, the gravitationally stable stratification is the stabilizing influence corresponding to the toroidal magnetic field in our case.

The fact that the bounds on the growth rates and phase velocities do not depend on the presence of the toroidal magnetic field suggests, although it does not prove, that a hydromagnetic circumpolar vortex of this type is not capable of producing unstable disturbances in which the impetus is the release of "available magnetic energy" associated with the horizontal shear in the toroidal field. From (6-31) we can see that this is true in the special case when $U_{max} = U_{min}$. If this also could be proved in the general case, it would have considerable implications for the theory of the solar cycle, since one conceivable mechanism for limiting the strength of the toroidal field would be ruled out.

In closing this chapter on integral theorems, it should be pointed out that Pedlosky (1964a) was able to find other bounds on growth rates of unstable waves by introducing a transformation of the form

$$\tilde{\Psi} = (U - c)^{1/2} K \quad (6-32)$$

In the magnetic case, this kind of transformation does not appear to be useful, since the magnetic terms become very complicated. The transformation (6-17) worked well because of the simplicity of the resulting relation (6-18).

7. Formulation of a two-layer model; integral theorems

Perturbation and energy equations

In general, to get actual solutions representing disturbances on a hydromagnetic vortex whose flow and magnetic fields are continuously varying functions is a very difficult problem. Even if, for example, equations (6-8) and (6-9) were simplified to study a vortex with purely baroclinic flow (i.e., one in which $\partial v / \partial y$ everywhere), and a uniform magnetic field, and which was incompressible, the resulting equation would be a confluent hypergeometric equation. If, to simplify further, the β effect were neglected, a Bessel's equation would result, but the order of the Bessel functions would depend on the magnetic field strength, becoming imaginary for fields above a certain strength. (This equation will be derived and discussed briefly in Chapter 10).

Because of the computational difficulties in dealing with such equations, it was decided first to remove some of the complications by examining instead a two layer model. This will truncate vertical variations as severely as possible without completely destroying the baroclinic character of the motion. All vertical derivatives are replaced by finite differences, but all horizontal derivatives remain continuous.

Following the usual construction of such layer models in meteorological studies (c.f. Thompson 1961), the various dependent variables are evaluated at the levels indicated in Fig. 2. The mean zonal flow U and toroidal field H are evaluated at two levels, labeled 1 and 3, as are the perturbation stream function ψ and magnetic flux χ .

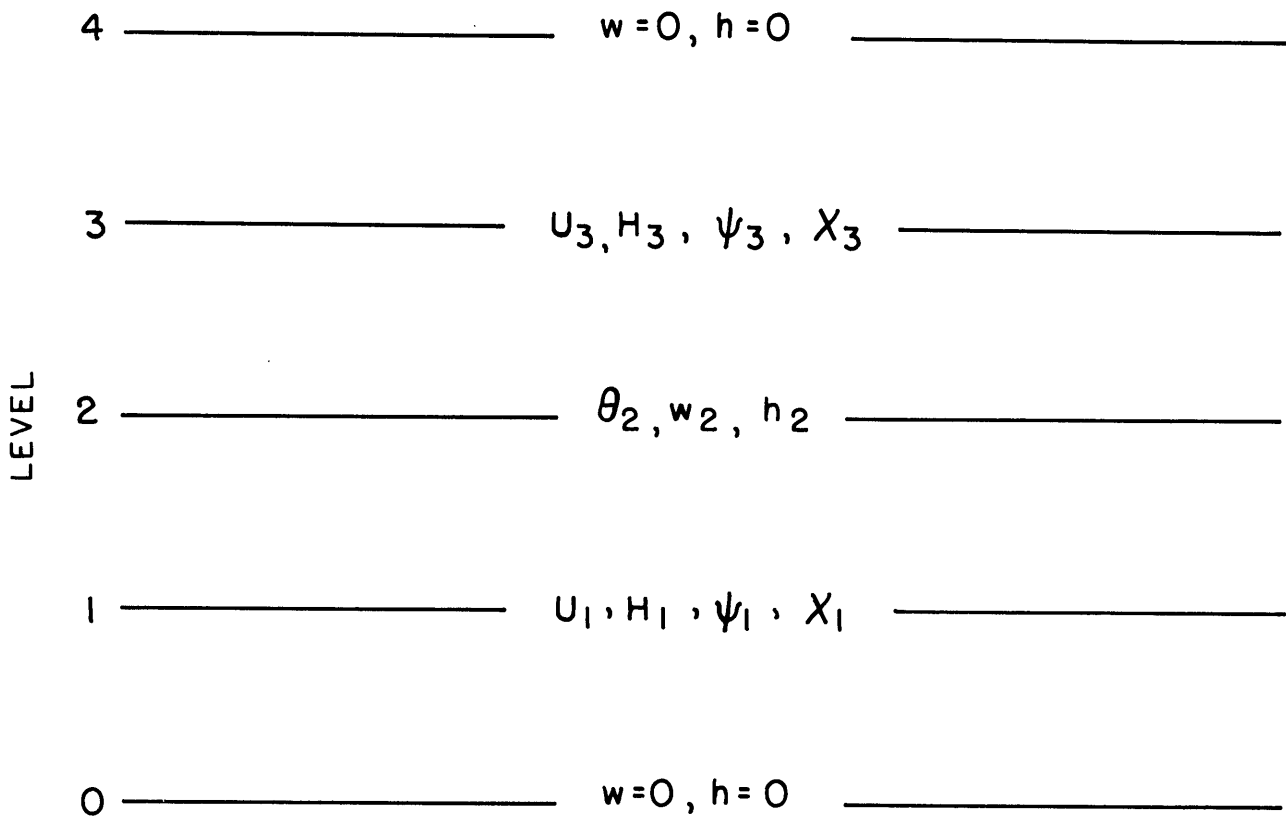


Fig. 2. Levels at which the various dependent variables are evaluated for the two-layer model.

The temperature θ , vertical motion w , and vertical magnetic field h are evaluated at one level, level 2. The assumed rigid perfectly conducting top and bottom require that w and h be zero at levels 0 and 4. The levels are spaced equally and the total nondimensional distance between top and bottom is assumed to be unity. The static stability, represented by ϵ , remains a constant. The fractional difference in mean density ρ_s^* of the two layers is assumed much less than unity.

Using the variables defined in Fig. 2 then, we may write the perturbation equations (5-1)-(5-4) and (5-8) in the form

$$\left(\frac{\partial}{\partial t} + U_1 \frac{\partial}{\partial x}\right) \nabla^2 \psi_1 + \left(\beta - \frac{\partial^2 U_1}{\partial y^2}\right) \frac{\partial \psi_1}{\partial x} - 2 \sin \phi_0 w_2 = P_s H_1 \nabla^2 \frac{\partial \chi_1}{\partial x} - P_s \frac{\partial^2 H_1}{\partial y^2} \frac{\partial \chi_1}{\partial x} \quad (7-1)$$

$$\left(\frac{\partial}{\partial t} + U_3 \frac{\partial}{\partial x}\right) \nabla^2 \psi_3 + \left(\beta - \frac{\partial^2 U_3}{\partial y^2}\right) \frac{\partial \psi_3}{\partial x} + 2 \sin \phi_0 w_2 = P_s H_3 \nabla^2 \frac{\partial \chi_3}{\partial x} - P_s \frac{\partial^2 H_3}{\partial y^2} \frac{\partial \chi_3}{\partial x} \quad (7-2)$$

$$\left(\frac{\partial}{\partial t} + \frac{U_1 + U_3}{2} \frac{\partial}{\partial x}\right) 2(\psi_3 - \psi_1) - (U_3 - U_1) \frac{\partial}{\partial x} (\psi_1 + \psi_3) + \frac{w_2}{\epsilon \sin \phi_0} = 0 \quad (7-3)$$

$$\left(\frac{\partial}{\partial t} + U_1 \frac{\partial}{\partial x}\right) \frac{\partial \chi_1}{\partial x} = H_1 \frac{\partial^2 \psi_1}{\partial x^2} \quad (7-4)$$

$$\left(\frac{\partial}{\partial t} + U_3 \frac{\partial}{\partial x}\right) \frac{\partial \chi_3}{\partial x} = H_3 \frac{\partial^2 \psi_3}{\partial x^2} \quad (7-5)$$

$$\left(\frac{\partial}{\partial t} + u_1 \frac{\partial}{\partial x}\right) \frac{\partial \chi_1}{\partial y} = \frac{\partial H_1}{\partial y} \frac{\partial \psi_1}{\partial x} + H_1 \frac{\partial^2 \psi_1}{\partial y \partial x} - \frac{\partial u_1}{\partial y} \frac{\partial \chi_1}{\partial x} \quad (7-6)$$

$$\left(\frac{\partial}{\partial t} + u_3 \frac{\partial}{\partial x}\right) \frac{\partial \chi_3}{\partial y} = \frac{\partial H_3}{\partial y} \frac{\partial \psi_3}{\partial x} + H_3 \frac{\partial^2 \psi_3}{\partial y \partial x} - \frac{\partial u_3}{\partial y} \frac{\partial \chi_3}{\partial x} \quad (7-7)$$

$$\left(\frac{\partial}{\partial t} + \frac{u_1 + u_3}{a} \frac{\partial}{\partial x}\right) h_2 = \frac{H_1 + H_3}{a} \frac{\partial w_2}{\partial x} \quad (7-8)$$

To eliminate extra primes and angular brackets, in the above equations and those that follow, we have written the perturbation variables (except w and h) in terms of the stream and magnetic flux functions, while the mean state variables are in terms of the potential temperature, zonal flow and toroidal magnetic field.

The tendency equations for changes in the initial state, equations (5-5)-(5-7) and (5-9), can similarly be written as

$$\frac{\partial}{\partial t} \frac{\theta_2}{\sin \phi_0} = - \frac{\partial}{\partial y} \left\langle (\psi_3 - \psi_1) \frac{\partial}{\partial x} (\psi_3 + \psi_1) \right\rangle - \frac{\langle w_2 \rangle}{\epsilon \sin \phi_0} \quad (7-9)$$

$$\frac{\partial}{\partial t} \frac{\partial u_1}{\partial y} = \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \psi_1}{\partial x} \frac{\partial \psi_1}{\partial y} \right\rangle - P_s \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \chi_1}{\partial x} \frac{\partial \psi_1}{\partial y} \right\rangle - 2 \sin \phi_0 \langle w_2 \rangle \quad (7-10)$$

$$\frac{\partial}{\partial t} \frac{\partial u_3}{\partial y} = \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \psi_3}{\partial x} \frac{\partial \psi_3}{\partial y} \right\rangle - P_s \frac{\partial^2}{\partial y^2} \left\langle \frac{\partial \chi_3}{\partial x} \frac{\partial \psi_3}{\partial y} \right\rangle + 2 \sin \phi_0 \langle w_2 \rangle \quad (7-11)$$

$$\frac{\partial H_1}{\partial t} = -\frac{\partial}{\partial y} \left\langle \frac{\partial \chi_1}{\partial x} \frac{\partial \psi_1}{\partial y} - \frac{\partial \chi_1}{\partial y} \frac{\partial \psi_1}{\partial x} \right\rangle \quad (7-12)$$

$$\frac{\partial H_3}{\partial t} = -\frac{\partial}{\partial y} \left\langle \frac{\partial \chi_3}{\partial x} \frac{\partial \psi_3}{\partial y} - \frac{\partial \chi_3}{\partial y} \frac{\partial \psi_3}{\partial x} \right\rangle \quad (7-13)$$

$$\frac{\partial \langle h_2 \rangle}{\partial t} = \frac{1}{2} \frac{\partial}{\partial y} \left\langle w_2 \frac{\partial}{\partial x} (\chi_1 + \chi_3) - h_2 \frac{\partial}{\partial x} (\psi_1 + \psi_3) \right\rangle \quad (7-14)$$

The boundary conditions (5-10) for the side walls become

$$\frac{\partial \psi_{1,3}}{\partial x}, \quad \frac{\partial \chi_{1,3}}{\partial x}, \quad \frac{\partial U_{1,3}}{\partial t} = 0 \quad y=0, l \quad (7-15)$$

Finally, the energy balance equations (5-12)-(5-17) of course have the same form as in the continuous case. The energy conversion terms (5-19) to (5-25) may be written as (analogous to Phillips 1956).

$$(\langle K \rangle, \langle A \rangle) = -\int_0^l \rho_s^* \langle w_2 \rangle \theta_2 dy \quad (7-16)$$

$$(\langle A \rangle, A') = -\int_0^l \rho_s^* E \sin \phi_0 \left\langle (\psi_3 - \psi_1) \frac{\partial}{\partial x} (\psi_3 + \psi_1) \right\rangle \frac{\partial \theta_2}{\partial y} dy \quad (7-17)$$

$$(A', K') = 2 \int_0^l \rho_s^* \sin \phi_0 \langle w_2 (\psi_3 - \psi_1) \rangle dy \quad (7-18)$$

$$(K', \langle K \rangle) = -\frac{1}{2} \int_0^l \rho_s^* \left\langle \frac{\partial \psi_1}{\partial x} \frac{\partial \psi_1}{\partial y} \right\rangle \frac{\partial U_1}{\partial y} dy - \frac{1}{2} \int_0^l \rho_s^* \left\langle \frac{\partial \psi_3}{\partial x} \frac{\partial \psi_3}{\partial y} \right\rangle \frac{\partial U_3}{\partial y} dy \quad (7-19)$$

$$(\langle K \rangle, M') = -\frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \chi_1}{\partial x} \frac{\partial \chi_1}{\partial y} \right\rangle \frac{\partial U_1}{\partial y} dy - \frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \chi_3}{\partial x} \frac{\partial \chi_3}{\partial y} \right\rangle \frac{\partial U_3}{\partial y} dy \quad (7-20)$$

$$(K', M') = -\frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \psi_1}{\partial x} \nabla^2 \chi_1 \right\rangle H_1 dy - \frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \psi_3}{\partial x} \nabla^2 \chi_3 \right\rangle H_3 dy \quad (7-21)$$

$$\begin{aligned} (K', \langle M \rangle) &= -\frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \psi_1}{\partial x} \frac{\partial \chi_1}{\partial y} - \frac{\partial \psi_1}{\partial y} \frac{\partial \chi_1}{\partial x} \right\rangle \frac{\partial H_1}{\partial y} dy \\ &\quad - \frac{1}{2} \int_0^l \rho_s^* P_s \left\langle \frac{\partial \psi_3}{\partial x} \frac{\partial \chi_3}{\partial y} - \frac{\partial \psi_3}{\partial y} \frac{\partial \chi_3}{\partial x} \right\rangle \frac{\partial H_3}{\partial y} dy \end{aligned} \quad (7-22)$$

Integral theorems

It is obvious that for the two level model, there is a relation between the perturbation stresses and the mean state required for the perturbations to grow that is analogous to (6-3). It has the form

$$\begin{aligned} \int_0^l \rho_s^* dy &\left[-\epsilon \sin \phi_0 \left\langle (\psi_3 - \psi_1) \frac{\partial (\psi_3 + \psi_1)}{\partial x} \right\rangle \frac{\partial \theta_2}{\partial y} \right. \\ &\quad + \frac{1}{2} \left\langle \frac{\partial \psi_1}{\partial x} \frac{\partial \psi_1}{\partial y} - P_s \frac{\partial \chi_1}{\partial x} \frac{\partial \chi_1}{\partial y} \right\rangle \frac{\partial U_1}{\partial y} \\ &\quad + \frac{1}{2} \left\langle \frac{\partial \psi_3}{\partial x} \frac{\partial \psi_3}{\partial y} - P_s \frac{\partial \chi_3}{\partial x} \frac{\partial \chi_3}{\partial y} \right\rangle \frac{\partial U_3}{\partial y} \\ &\quad - \frac{1}{2} P_s \left\langle \frac{\partial \chi_1}{\partial x} \frac{\partial \psi_1}{\partial y} - \frac{\partial \chi_1}{\partial y} \frac{\partial \psi_1}{\partial x} \right\rangle \frac{\partial H_1}{\partial y} \\ &\quad \left. - \frac{1}{2} P_s \left\langle \frac{\partial \chi_3}{\partial x} \frac{\partial \psi_3}{\partial y} - \frac{\partial \chi_3}{\partial y} \frac{\partial \psi_3}{\partial x} \right\rangle \frac{\partial H_3}{\partial y} \right] > 0 \end{aligned} \quad (7-23)$$

The inference from (6-3) that the horizontal Maxwell stresses should act to destabilize disturbances which are baroclinically unstable but barotropically stable obviously applies to the two layer case as well (indeed, the disturbances of this type for the nonmagnetic case found by Pedlosky (1964a) were for the two layer model).

As in the continuous case, the necessary conditions for instability involving the potential vorticity and zonal flow (Pedlosky 1964a) no longer hold with the toroidal magnetic field present. Because of the length of the new conditions, and the obscurity of their interpretation, we omit them here.

Again, just as in the continuous case, the bounds on the phase velocities and growth rates of unstable disturbances in the two layer model can easily be shown to be the same as in the nonmagnetic problem. That is, the conditions (6-25), (6-28) and (6-29) for the continuous case apply equally well to the two-layer case. To see this, we first form the potential vorticity equation for levels 1 and 3 by substituting for W_2 from (7-3) into (7-1) and (7-2), from which we obtain

$$\begin{aligned} \left(\frac{\partial}{\partial t} + U_1 \frac{\partial}{\partial x}\right) \nabla^2 \psi_1 + \frac{\partial \varrho_1}{\partial y} \frac{\partial \psi_1}{\partial x} + 4\epsilon \sin^2 \phi_0 \left(\frac{\partial}{\partial t} + U_1 \frac{\partial}{\partial x}\right) (\psi_3 - \psi_1) & \quad (7-24) \\ & = P_s H_1 \nabla^2 \frac{\partial \chi_1}{\partial x} - P_s \frac{\partial^2 H_1}{\partial y^2} \frac{\partial \chi_1}{\partial x} \end{aligned}$$

$$\begin{aligned} \left(\frac{\partial}{\partial t} + U_3 \frac{\partial}{\partial x}\right) \nabla^2 \psi_3 + \frac{\partial \varrho_3}{\partial y} \frac{\partial \psi_3}{\partial x} - 4\epsilon \sin^2 \phi_0 \left(\frac{\partial}{\partial t} + U_3 \frac{\partial}{\partial x}\right) (\psi_3 - \psi_1) & \quad (7-25) \\ & = P_s H_3 \nabla^2 \frac{\partial \chi_3}{\partial x} - P_s \frac{\partial^2 H_3}{\partial y^2} \frac{\partial \chi_3}{\partial x} \end{aligned}$$

where
$$\frac{\partial \varrho_1}{\partial y} = \beta - \frac{\partial^2 U_1}{\partial y^2} - 4\epsilon \sin^2 \phi_0 (U_3 - U_1) \quad (7-26)$$

$$\frac{\partial \varrho_3}{\partial y} = \beta - \frac{\partial^2 U_3}{\partial y^2} + 4\epsilon \sin^2 \phi_0 (U_3 - U_1) \quad (7-27)$$

Then, assuming normal mode disturbances of the form

$$\psi_1, \psi_3, \chi_1, \chi_3 = [\tilde{\psi}_1, \tilde{\psi}_3, \tilde{\chi}_1, \tilde{\chi}_3] e^{ik(x-ct)} \quad (7-28)$$

and substituting them into (7-24) and (7-25) gives

$$\begin{aligned} (U_1 - c) \left(\frac{\partial^2 \tilde{\psi}_1}{\partial y^2} - k^2 \tilde{\psi}_1 \right) + \frac{\partial \varrho_1}{\partial y} \tilde{\psi}_1 + 4\epsilon \sin^2 \phi_0 (U_1 - c) (\tilde{\psi}_3 - \tilde{\psi}_1) \\ = P_5 H_1 \left(\frac{\partial^2 \tilde{\chi}_1}{\partial y^2} - k^2 \tilde{\chi}_1 \right) - P_5 \frac{\partial^2 H_1}{\partial y^2} \tilde{\chi}_1 \end{aligned} \quad (7-29)$$

$$\begin{aligned} (U_3 - c) \left(\frac{\partial^2 \tilde{\psi}_3}{\partial y^2} - k^2 \tilde{\psi}_3 \right) + \frac{\partial \varrho_3}{\partial y} \tilde{\psi}_3 - 4\epsilon \sin^2 \phi_0 (U_3 - c) (\tilde{\psi}_3 - \tilde{\psi}_1) \\ = P_5 H_3 \left(\frac{\partial^2 \tilde{\chi}_3}{\partial y^2} - k^2 \tilde{\chi}_3 \right) - P_5 \frac{\partial^2 H_3}{\partial y^2} \tilde{\chi}_3 \end{aligned} \quad (7-30)$$

Substitution of (7-28) into (7-4) and (7-5) gives

$$(U_1 - c) \tilde{\chi}_1 = H_1 \tilde{\psi}_1 \quad (7-31)$$

$$(U_3 - c) \tilde{\chi}_3 = H_3 \tilde{\psi}_3 \quad (7-32)$$

Analogous to the transformation (6-17), we define

$$\tilde{\psi}_1 = (U_1 - c) \mathcal{G}_1 \quad (7-33)$$

$$\tilde{\psi}_3 = (U_3 - c) \mathcal{G}_3 \quad (7-34)$$

so that

$$\tilde{\chi}_1 = H_1 G_1 \quad (7-35)$$

$$\tilde{\chi}_3 = H_3 G_3 \quad (7-36)$$

The boundary conditions (7-15) become

$$G_1 = G_3 = 0, \quad y = 0, l \quad (7-37)$$

Substituting (7-33)-(7-36) into (7-29) and (7-30), and using the relations

$$(U_n - c) \frac{d^2}{dy^2} [(U_n - c) G_n] - (U_n - c) G_n \frac{d^2 U_n}{dy^2} = \frac{d}{dy} \left[(U_n - c)^2 \frac{d G_n}{dy} \right] \quad (7-38)$$

$$H_n \frac{d^2 \chi_n}{dy^2} - \chi_n \frac{d^2 H_n}{dy^2} = \frac{d}{dy} \left[H_n^2 \frac{d \chi_n}{dy} \right]; \quad n = 1, 3$$

we get

$$\frac{d}{dy} \left[(U_1 - c)^2 \frac{d G_1}{dy} \right] - k^2 (U_1 - c)^2 G_1 + \beta (U_1 - c) G_1 \quad (7-39)$$

$$+ 4E \sin^2 \phi_0 (U_1 - c)(U_3 - c)(G_3 - G_1) = P_5 \frac{d}{dy} \left(H_1^2 \frac{d G_1}{dy} \right) - k^2 P_5 H_1^2 G_1$$

$$\frac{d}{dy} \left[(U_3 - c)^2 \frac{d G_3}{dy} \right] - k^2 (U_3 - c)^2 G_3 + \beta (U_3 - c) G_3$$

$$+ 4E \sin^2 \phi_0 (U_1 - c)(U_3 - c)(G_1 - G_3) = P_5 \frac{d}{dy} \left(H_3^2 \frac{d G_3}{dy} \right) - k^2 P_5 H_3^2 G_3 \quad (7-40)$$

Then if we multiply (7-39) by G_1 , (7-40) by G_3 , integrate over the cross section from $y=0$ to l , rearrange by parts, applying the

boundary conditions (7-37), and add, defining, analogously to the continuous case,

$$\begin{aligned} \left| \frac{dG_n}{dy} \right|^2 + k^2 |G_n|^2 &= Q_n \\ 4\epsilon \sin^2 \phi_0 |G_3 - G_1|^2 &= R \\ |G_n|^2 &= S_n \quad ; \quad n=1,3 \end{aligned} \tag{7-41}$$

we find that

$$\begin{aligned} \int_0^l (U_1 - C)^2 (Q_1 + \frac{R}{2}) dy + \int_0^l (U_3 - C)^2 (Q_3 + \frac{R}{2}) dy &= \\ \beta \int_0^l [(U_1 - C) S_1 + (U_3 - C) S_3] dy + \int_0^l (U_3 - U_1)^2 \frac{R}{2} dy & \\ + P_s \int_0^l (H_1^2 Q_1 + H_3^2 Q_3) dy & \end{aligned} \tag{7-42}$$

The imaginary part of (7-42) gives, since $C_i \neq 0$,

$$\int_0^l (U_1 - C_r) (Q_1 + \frac{R}{2}) dy + \int_0^l (U_3 - C_r) (Q_3 + \frac{R}{2}) dy = \beta \int_0^l (S_1 + S_3) dy \tag{7-43}$$

But (7-43) is obviously just the two level representation of (6-24).

Therefore it follows that in the two level case, we still have

$$U_{\min} - \frac{\beta/2}{k^2 + \pi^2/l^2} \leq C_r \leq U_{\max} \tag{7-44}$$

From the real part of (7-42), we can also easily show that the semicircle theorem

$$0 \geq \left(C_r - \frac{U_{\max} + U_{\min}}{2} \right)^2 + C_i^2 - \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{\max} - U_{\min}}{2} - \left(\frac{U_{\max} - U_{\min}}{2} \right)^2 \quad (7-45)$$

still holds, and that the bounds on the growth rate

$$C_i^2 \leq \left(\frac{U_{\max} - U_{\min}}{2} \right)^2 + \frac{\beta}{k^2 + \pi^2/l^2} \frac{U_{\max} - U_{\min}}{2} \quad (7-46)$$

remain the same as in the continuous case. It goes without saying that if the flow is barotropic, the semicircle radius is again reduced by an amount

$$\left(P_s H^2 \right)_{\min}.$$

As a final integral theorem, we can place limits on the phase velocities of neutral waves in the vortex. Since these waves will in general be some combination of Alfvén and vorticity waves, unlike the unstable waves their bounds will be different from the nonmagnetic case, in which only vorticity waves are allowed. The procedure for finding the bounds in the magnetic case is, however, similar to that of Pedlosky (1964b).

Let us assume that C (which is real, in this case) is outside the range of U_1 and U_3 . Then the transformations (7-33) and (7-34) are permissible. From (7-39) and (7-40), multiplying the former by G_1 and the latter by G_3 , integrating by parts, applying the boundary conditions (7-37), and adding the result, using the definitions (7-41), we get

$$\int_0^l [(U_1 - C)^2 - P_s H_1^2] Q_1 dy + \int_0^l [(U_3 - C)^2 - P_s H_3^2] Q_3 dy + \int_0^l (U_1 - C)(U_3 - C) R dy = \beta \int_0^l [(U_1 - C) S_1 + (U_3 - C) S_3] dy \quad (7-47)$$

Now suppose $C > U_{max}$. Then the right hand side of (7-47) will be negative. Therefore, from the left hand side, C cannot be so large that both

$$\begin{aligned} (U_1 - C)^2 - P_s H_1^2 &> 0 \\ (U_3 - C)^2 - P_s H_3^2 &> 0 \end{aligned}$$

everywhere, so that we must have

$$[(U - C)^2]_{min} < P_s (H^2)_{max} \quad (7-48)$$

But since $C > U_{max}$, it follows that we must have

$$U_{max} - C \geq -\sqrt{P_s} |H|_{max} \quad (7-49)$$

or that

$$C \leq U_{max} + \sqrt{P_s} |H|_{max} \quad (7-50)$$

But $\sqrt{P_s} |H|_{max}$ is simply the maximum Alfvén velocity in the system.

We denote it by U_{Amax} , so that C has an upper bound given by

$$C \leq U_{max} + U_{Amax} \quad (7-51)$$

In words, the phase velocity of a neutral wave cannot exceed the sum of the maximum flow and Alfvén velocities in the system.

To find the lower limit on C , we note that if $C < U_{min}$, then (7-47) implies

$$\left[(U-c)^2 - P_s H^2 \right]_{\min} \int_0^l (Q_1 + Q_3) dy \leq (U_{\max} - c) \beta \int_0^l (S_1 + S_3) dy \quad (7-52)$$

From (7-41)

$$\int_0^l (Q_1 + Q_3) dy = k^2 \int_0^l (S_1 + S_3) dy + \int_0^l \left[\left(\frac{dG_1}{dy} \right)^2 + \left(\frac{dG_3}{dy} \right)^2 \right] dy \quad (7-53)$$

but by a well known theorem of calculus

$$\int_0^l \left[\left(\frac{dG_1}{dy} \right)^2 + \left(\frac{dG_3}{dy} \right)^2 \right] dy \geq \frac{\pi^2}{l^2} \int_0^l (G_1^2 + G_3^2) dy \quad (7-54)$$

In addition

$$\left[(U-c)^2 - P_s H^2 \right]_{\min} = (U_{\min} - c)^2 - U_{A \max}^2 \quad (7-55)$$

From (7-53)-(7-55), then,

$$(U_{\min} - c)^2 - P_s (H^2)_{\max} \leq \frac{(U_{\max} - c) \beta}{k^2 + \pi^2/l^2} \quad (7-56)$$

from which we have

$$c^2 + \left(\frac{\beta}{k^2 + \pi^2/l^2} - 2U_{\min} \right) c - \left(\frac{U_{\max} \beta}{k^2 + \pi^2/l^2} + U_{A \max}^2 - U_{\min}^2 \right) \leq 0 \quad (7-57)$$

Therefore the lower limit on C is

$$c \geq U_{\min} - \frac{\beta/2}{k^2 + \pi^2/l^2} - \left[\frac{\beta^2}{4(k^2 + \pi^2/l^2)} + \frac{\beta}{k^2 + \pi^2/l^2} (U_{\max} - U_{\min}) + U_{A \max}^2 \right]^{1/2} \quad (7-58)$$

Thus the presence of the zonal magnetic field widens the range in which phase velocities of neutral waves can lie. We note as a check that if

$\beta = 0$ and $U = 0$, we obtain from (7-51) and (7-58)

$$U_{Amax} \geq c \geq -U_{Amax} \quad (7-59)$$

for Alfvén waves alone, which is what we would expect.

The limits (7-51) and (7-58) will prove to be useful in isolating marginally stable waves in flows with both horizontal and vertical shear, as we shall do for one such flow (parabolic in the upper layer) in Chapter 9.

8. Stability of two layer baroclinic flow in a uniform
zonal magnetic field

Derivation of the eigenvalue equation

Up to now, study of the stability of the hydromagnetic circumpolar vortex has been limited to rather general theorems. In this chapter and the next we turn to the examination of the stability of particular zonal flow and toroidal magnetic field profiles, and the structure and energetics of the unstable disturbances which result. In general, such stability problems are quite hard to solve unless (although sometimes even when) one limits the available energy sources for perturbations to a single kind. In principle there are three possibilities in our system, namely available potential associated with meridional temperature gradients, available kinetic associated with meridional shears in the zonal flow, and available magnetic associated with meridional shears in the toroidal field. The question is which one to choose. In accordance with Chapter 3, we will be considering available potential energy as the source for perturbation growth. In Chapter 6 we pointed out that it may be impossible for the shears in the toroidal field to provide the primary energy source for perturbations. In the light of this possibility, and for simplicity, we may, at least as a start, take the toroidal field to be uniform. In this chapter, then, we consider the further simplified case in which the flow is purely baroclinic i.e. no horizontal shear in the zonal flow. In Chapter 9, we examine one case where the zonal flow has horizontal as well as vertical shear, in particular, a parabolic profile. In both cases, in accordance with the remarks at the beginning of Chapter 7, we limit ourselves to the two level model.

With these simplifications, the purely baroclinic problem can be reduced to solving a quartic equation for the complex phase velocity C . This equation may be found from equations (7-29)-(7-32). Since the initial state is independent of y , (to satisfy the boundary condition at the side walls) the amplitude functions $\tilde{\Psi}_1, \tilde{\Psi}_3, \tilde{\chi}_1, \tilde{\chi}_3$ may be written in the form

$$\tilde{\Psi}_1, \tilde{\Psi}_3, \tilde{\chi}_1, \tilde{\chi}_3 = [\hat{\Psi}_1, \hat{\Psi}_3, \hat{\chi}_1, \hat{\chi}_3] \sin \frac{\pi y}{l} \quad (8-1)$$

where, since the walls are artificial, we consider only the lowest mode. (The lowest mode should also give the highest growth rates). The amplitudes with carries are then (complex) constants. Then substituting (8-1) into (7-29)-(7-32), setting $H_1 = H_3 = H = \text{constant}$ we get

$$\begin{aligned} -(U_1 - C) \left(k^2 + \frac{\pi^2}{l^2} \right) \hat{\Psi}_1 + \frac{\partial \rho_1}{\partial y} \hat{\Psi}_1 + 4 \epsilon \sin^2 \phi_0 (U_1 - C) (\hat{\Psi}_3 - \hat{\Psi}_1) \\ = -P_s H \left(k^2 + \frac{\pi^2}{l^2} \right) \hat{\chi}_1 \end{aligned} \quad (8-2)$$

$$\begin{aligned} -(U_3 - C) \left(k^2 + \frac{\pi^2}{l^2} \right) \hat{\Psi}_3 + \frac{\partial \rho_3}{\partial y} \hat{\Psi}_3 - 4 \epsilon \sin^2 \phi_0 (U_3 - C) (\hat{\Psi}_3 - \hat{\Psi}_1) \\ = -P_s H \left(k^2 + \frac{\pi^2}{l^2} \right) \hat{\chi}_3 \end{aligned} \quad (8-3)$$

$$(U_1 - C) \hat{\chi}_1 = H \hat{\Psi}_1 \quad (8-4)$$

$$(U_3 - C) \hat{\chi}_3 = H \hat{\Psi}_3 \quad (8-5)$$

Then, substituting from (8-4) and (8-5) into (8-2) and (8-3) for $\hat{\chi}_1$ and $\hat{\chi}_3$,

we get

$$\left\{ (P_5 H^2 - (U_1 - C)^2) \left(k^2 + \frac{\pi^2}{l^2} \right) + (U_1 - C) \frac{\partial q_1}{\partial y} - 4E \sin^2 \phi_0 (U_1 - C)^2 \right\} \hat{\Psi}_1 + 4E \sin^2 \phi_0 (U_1 - C)^2 \hat{\Psi}_3 = 0 \quad (8-6)$$

$$\left\{ (P_5 H^2 - (U_3 - C)^2) \left(k^2 + \frac{\pi^2}{l^2} \right) + (U_3 - C) \frac{\partial q_3}{\partial y} - 4E \sin^2 \phi_0 (U_3 - C)^2 \right\} \hat{\Psi}_3 + 4E \sin^2 \phi_0 (U_3 - C)^2 \hat{\Psi}_1 = 0 \quad (8-7)$$

The system (8-6)-(8-7) has a solution if and only if the determinant of the coefficients of $\hat{\Psi}_1$ and $\hat{\Psi}_3$ vanishes. Let us define

$$\left. \begin{aligned} A &= \frac{8E \sin^2 \phi_0}{k^2 + \pi^2/l^2} \quad ; \quad G = \frac{\beta}{8E \sin^2 \phi_0} \\ U_0 &= U_3 - U_1 \quad ; \quad P_5 H^2 = P \\ C' &= C - \frac{U_1 + U_3}{2} \end{aligned} \right\} \quad (8-8)$$

so that from (8-6)-(8-8) after rearranging, we obtain the eigenvalue equation for C' .

$$\begin{aligned} (A+1)C'^4 + AG(A+2)C'^3 - \left(\frac{U_0^2}{2} - A^2 G^2 + (A+2)P \right) C'^2 \\ - AG \left(\frac{U_0^2}{4} (A+2) + 2P \right) C' + \left(P - \frac{U_0^2}{4} \right) \left(P - \frac{U_0^2}{4} (1-A) \right) - \frac{A^2 G^2 U_0^2}{4} = 0 \end{aligned} \quad (8-9)$$

Stability criteria

To find the stability criteria, we must therefore solve a quartic equation for C' . This can be done by using standard procedures in, for example, mathematical tables by Burington (1953). But we can outline the regions of instability rather well by considering (8-9) for several limiting cases. The results which follow are presented also in Fig. 3. As the simplest of limiting cases, suppose we set $G = 0$ (no β effect) and $P = 0$ (no magnetic field). Then (8-9) reduces to

$$\left(C'^2 - \frac{U_0^2}{4}\right) \left(C'^2 - \frac{U_0^2}{4} + A \left(C'^2 + \frac{U_0^2}{4}\right)\right) = 0 \quad (8-10)$$

The first factor was introduced by the magnetic field, and need not concern us. The second factor is readily solved for C'^2 , giving

$$C'^2 = \frac{U_0^2}{4} \frac{1-A}{1+A} \quad (8-11)$$

Thus one wave will grow exponentially if

$$A > 1 \quad (8-12)$$

This result is represented graphically in Fig. 3a.

The onset of instability in this case is thus independent of the vertical shear. In terms of the wave number, (8-12) may be written as

$$k^2 + \frac{\pi^2}{l^2} < 8\epsilon \sin^2 \phi_0 \quad (8-13)$$

This is the two level analog of the result for the continuous case with linear vertical shear, the so-called Eady problem (Eady, 1949). In the continuous case, the criterion is approximately $k^2 + \pi^2/l^2 < 5.76 \epsilon \sin^2 \phi_0$, so there is good correspondence. Expression (8-11) says that for a given static stability ($\epsilon = \text{const}$) and latitude, if the channel is not too narrow, all wavelengths longer than some critical value will be unstable. However, if the channel is narrow enough (l small enough) then all wavelengths will be neutral. Physically, this is, in some sense, because potential energy releasing trajectories can no longer fit into the channel.

Since from (8-11), $C_r' = 0$ for the unstable waves, they all move with the average velocity of the zonal flow, i.e. $C_r = (U_1 + U_3)/2$. The growth rate for unstable waves is given by $k C_i'$. As a function of A , it will be a maximum where

$$\frac{d}{dA} (k C_i')^2 = 0 \quad (8-14)$$

From (8-11) and the defining relation for A in (8-8), we may write

$$(k C_i')^2 = U_0^2 \left(\frac{K_1}{A} + K_2 \right) \frac{A-1}{A+1} \quad (8-15)$$

where $K_1 = 2\epsilon \sin^2 \phi_0$; $K_2 = -\pi^2/4l^2$

Then substitution of (8-15) into (8-14), noting that

$$K_2/K_1 = -\frac{1}{A} \left(\frac{\lambda}{2l} \right)^2, \text{ we get}$$

$$A^2 - 2 \left[1 - \frac{1}{4} \left(\frac{\lambda}{2l} \right)^2 \right] A - 1 = 0 \quad (8-16)$$

or

$$A \text{ (max growth rate)} = 1 - \frac{1}{4} \left(\frac{\lambda}{l} \right)^2 + \left(1 + \left[1 - \frac{1}{4} \left(\frac{\lambda}{l} \right)^2 \right]^2 \right)^{1/2} \quad (8-17)$$

We can see that for $\lambda/l < 1$, the location of the maximum is not far from that where $l = \infty$, i.e.

$$A \text{ (max growth rate, } l = \infty) = 1 + \sqrt{2} \quad (8-18)$$

From the magnetograms of Howard and Bumba (1965b) the most appropriate ratio λ/l probably lies in the range $1/2 < \lambda/l < 1$.

What does the presence of a toroidal magnetic field do to the very simple stability criterion (8-10)? From (8-9), again setting $G = 0$, but retaining P , we obtain the equation

$$(A+1)C'^4 - \left(\frac{U_0^2}{2} + (A+2)P \right) C'^2 + \left(P - \frac{U_0^2}{4} \right) \left(P - \frac{U_0^2}{4} (1-A) \right) = 0 \quad (8-19)$$

Since (8-17) is biquadratic (the odd powers of C' are missing), it can be treated as a quadratic for C'^2 , so that the solution is

$$C'^2 = \frac{\frac{U_0^2}{2} + (A+2)P}{2(A+1)} \pm \left[\left(\frac{\frac{U_0^2}{2} + (A+2)P}{2(A+1)} \right)^2 - \frac{\left(P - \frac{U_0^2}{4} \right) \left(P - \frac{U_0^2}{4} (1-A) \right)}{A+1} \right]^{1/2} \quad (8-20)$$

Instability can occur only if the minus sign is chosen. We can show that the argument of the square root is always positive.

Thus the right hand side of (8-18) will be negative, and therefore instability will occur, if and only if

$$\left(P - \frac{U_0^2}{4}\right) \left(P - \frac{U_0^2}{4}(1-A)\right) < 0 \quad (8-21)$$

Therefore the unstable region is defined by

$$P < \frac{U_0^2}{4} \quad , \quad P > \frac{U_0^2}{4}(1-A) \quad (8-22)$$

or

$$|U_0| > 2\sqrt{P} \quad , \quad |U_0| < \frac{2\sqrt{P}}{\sqrt{1-A}} \quad (8-23)$$

The second condition in (8-22) contains the limiting case (8-10) for $P=0$.

The unstable region defined by (8-23) is sketched in Fig. 3b. We can see, then, that the principal effects of the toroidal magnetic field are to render the flow neutral to wavelike disturbances for small shears at all wavelengths, and to destabilize the flow for moderate shears at short wavelengths. The stabilizing effect on small shears is seen to be independent of wavelength.

It is interesting to compare this latter result with the condition (6-31), which holds equally as well in the two level as in the continuous case. For $\beta = 0$ and a uniform magnetic field, (6-31) reduces to the statement that if

$$P \geq \left(\frac{U_{max} - U_{min}}{2}\right)^2 \quad (8-24)$$

all disturbances are rendered stable (neutral). Now expression (8-24) is

SCHEMATIC STABILITY DIAGRAMS

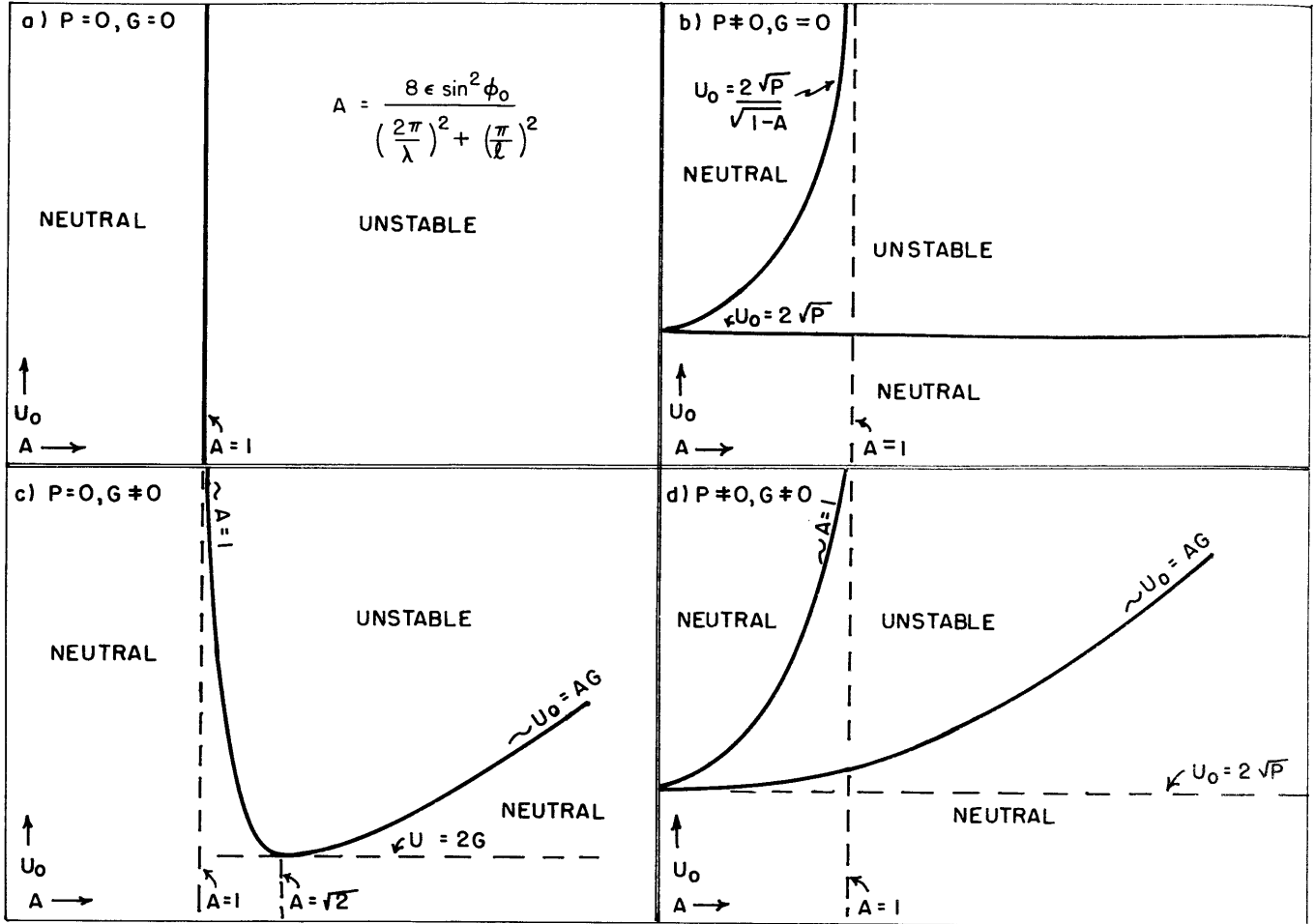


Fig. 3. Schematic stability diagram showing qualitative effects of a uniform zonal magnetic field (P) and/or β effect (G) on the stability criteria for baroclinic disturbances in the two layer model. Solid lines define boundaries of unstable regions; dashed lines denote asymptotes or minimum points of the stability boundaries. Ordinate is the vertical shear of the zonal flow, U_0 ; abscissa $A = 8\epsilon \sin^2 \phi_0 / (k^2 + \pi^2/l^2)$, where $k = 2\pi/\lambda$, λ the zonal wavelength, l the nondimensional channel width. ϵ (defined in Table 1 and text) is inversely proportional to the static stability.

a condition on barotropic flow, but if we identify $U_{max} - U_{min}$ with U_0 , it is identical with the condition on baroclinic flow given by the first inequality in (8-22), namely that for

$$P > \frac{U_0^2}{4} \quad (8-25)$$

the magnetic field renders all baroclinic disturbances neutral. Thus at least for the two level model, the minimum shear necessary for instability is the same for purely baroclinic and purely barotropic flows.

From (8-20), we can see that, just as in the nonmagnetic case, the phase velocity of the unstable wave is just the mean velocity $(U_1 + U_3)/2$ of the shear flow. Thus the two unstable modes (one growing, one decaying) are in no sense Alfvén waves, but rather baroclinic waves whose stability characteristics are somewhat altered by the presence of the magnetic field. The two neutral modes in (8-20), arising when the positive sign is chosen, do represent Alfvén waves, but whose phase velocities are altered by the zonal flow present, and which are made dispersive by the stable stratification. These phase velocities must fall within the bounds found for neutral waves in Chapter 7, i.e. (7-51) and (7-58).

Let us now see how including the β effect ($G \neq 0$) affects the stability curves. As a reference point, let us again consider the non-magnetic case. With $P=0$, $G \neq 0$ (8-9) reduces to (neglecting the common factor $(c'^2 - \frac{U_0^2}{4})$)

$$(A+1)c'^2 + AG(A+2)c' + A^2G^2 + \frac{A-1}{4}U_0^2 = 0 \quad (8-26)$$

Solving by the quadratic formula again, we get

$$C' = \frac{-AG(A+2)}{2(A+1)} \pm \frac{1}{2(A+1)} \left[A^2 G^2 (A+2)^2 - 4(A+1) \left(A^2 G^2 + \frac{A-1}{4} U_0^2 \right) \right]^{1/2} \quad (8-27)$$

There will be instability if the argument of the square root becomes negative.

This reduces to the requirement

$$U_0^2 > \frac{A^4 G^2}{A^2 - 1} \quad (8-28)$$

for instability. Thus, the short wavelength (A small) boundary of the unstable region is asymptotic as $U_0 \rightarrow \infty$ to the neutral curve for $G=0$ (see Fig. 3a). For long wavelengths (A large) the neutral curve is asymptotic to

$$|U_0| = AG \quad (8-29)$$

There is a minimum shear required for instability, which we can find by setting the derivative of (8-28) with respect to A equal to zero. This gives

$$U_0(\min) = 2G, \quad \text{at } A = \sqrt{2} \quad (8-30)$$

All of these results are shown graphically in Fig. 3c.

Since $G \neq 0$ ($\beta \neq 0$) the phase velocities of the unstable waves are no longer equal to $(U_1 + U_3)/2$, but rather lag, by an amount

$$\frac{-AG(A+2)}{2(A+1)} \quad (8-31)$$

as we can see from (8-27). This is the well known property of Rossby waves.

The neutral curve for this case (see Fig. 3c) has, as we would expect, a shape quite similar to that found by Phillips 1954.

When neither P nor G vanish, we would expect a neutral curve which in some way combines the characteristics of Figs. 3b and 3c to result. We can argue plausibly that it should take the schematic form given in Fig. 3d. The left hand edge of the neutral curve will still be asymptotic to $A=1$, as $U_0 \rightarrow \infty$. That is, if in (8-9), we divide by U_0^4 and assume that c'/U_0 remains finite, we obtain

$$(A+1)\left(\frac{c'}{U_0}\right)^4 - \frac{1}{2}\left(\frac{c'}{U_0}\right)^2 + \frac{1}{16}(1-A) = 0 \quad (8-32)$$

which has as solutions,

$$c'^2 = \frac{U_0^2}{4}, \quad \frac{U_0^2}{4} \frac{1-A}{1+A} \quad (8-33)$$

the same as for the case $P=0, G=0$.

The cusp at $U_0 = 2\sqrt{P}$ also remains, as we can see by letting $A \rightarrow 0$ in (8-9). This gives

$$c'^4 - 2\left(P + \frac{U_0^2}{4}\right)c'^2 + \left(P - \frac{U_0^2}{4}\right)^2 = 0 \quad (8-34)$$

for which the solution is

$$c'^2 = P + \frac{U_0^2}{4} \pm U_0 \sqrt{P} \quad (8-35)$$

With $G=0$, we saw that $C_r' = 0$ everywhere in the unstable

region. Even when $G \neq 0$, this will still be the case along the line $A=0$. If the cusp remains, there should be only one point along $A=0$ where $C'_r = 0$ ($C_i = 0$ there also, since the point will be part of the neutral curve). From (8-35), this will be only at the point satisfying the equation $P + U_0^2/4 - U_0\sqrt{P} = 0$, which is given by $U_0 = 2\sqrt{P}$.

At the cusp, the slope of the right hand part of the neutral curve will still be horizontal. For A large, this segment of the neutral curve will be asymptotic to the curve for $P=0$, i.e. $\sim U_0 = AG$. We can see this by letting U_0 and A become large in (8-9) in such a way that U_0/AG remains of order unity (and again assuming C'/U_0 remains finite). We get, again dropping the common factor $(C'^2 - U_0^2/4)$

$$C'^2 + AGC' + \frac{U_0^2}{4} = 0 \quad (8-36)$$

with solution

$$C' = -\frac{AG}{2} \pm [A^2G^2 - U_0^2]^{1/2} \quad (8-37)$$

so that, as before, the unstable region lies above the curve $U_0 = AG$. This curve will be approached asymptotically from higher (lower) values of U_0 if the ratio P/G^2 is greater than (less than) one.

In general, then, the minimum shear required for instability no longer depends on the β effect, or on the static stability (ϵ), but

rather is determined only by the magnetic field strength. This is in accordance with the general theorems of chapters 6 and 7, where it was shown that the vanishing of the potential vorticity gradient (in one of the two layers, for the two layer case) was no longer a necessary condition for instability. The minimum shear occurs for the limiting case of $\lambda^2 \rightarrow 0$. This confirms our conclusion from expression (6-16) that at the shortest wavelengths, the magnetic field should dominate in determining the stability criteria.

Finally, it is notable that the requirement for instability that $P/U_0^2 < 1/4$ justifies a posteriori our scaling assumption on P , made in Chapter 4, namely $P \lesssim 1$.

Figure 3 then gives a schematic picture of how the various physical parameters determine the stability curves. To facilitate comparison of the resulting composite curve (Fig. 3d) with results in the nonmagnetic case (e.g. Phillips 1954, Pedlosky 1964b) we have plotted it for several values of P with a fixed G , in Fig. 4. These numerical results were found using the standard algebraic technique found in Mathematical Tables and Formulas by Burington, and an IBM 7094 computer. The abscissa, A , again involves the width of the channel as well as the wavelength of the disturbance. In calculating the curves, we have taken $\epsilon \sin^2 \phi_0 = 1/2$, $\beta = 0.707$, so that $G = 0.177$. (The values corresponding to Pedlosky's 1964b calculations would be $\epsilon \sin^2 \phi_0 = 3/4$; $G = 0.250$).

We can see from Fig. 4, then, that, for a given shear (and a channel wide enough so that $(\lambda/l)^2 < 1$), increasing the magnetic field can render

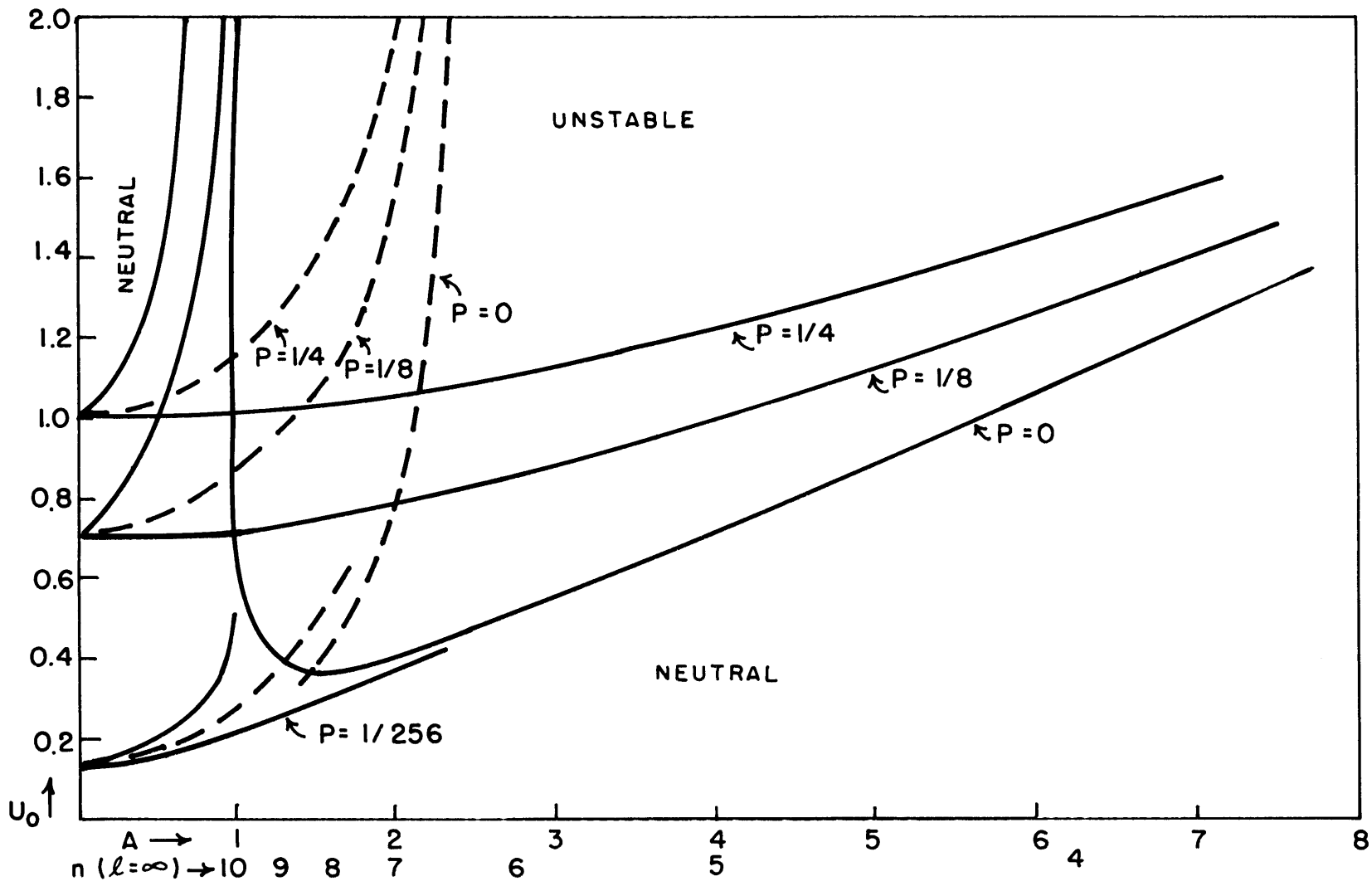


Fig. 4. Numerical evaluation of stability diagram for $G = 0.177$,
 $(\beta = 0.707)$, $E \sin^2 \phi = 0.5$, for various values of P .
 Ordinate and abscissa same as in Fig. 3. Number of waves n
 around latitude circle 45° plotted as second (lower) abscissa for
 $l = \infty$. For $\lambda/l < 1$, $A \approx \lambda^2/10$, where λ is in units
 of 10^5 km. Solid lines denote stability boundaries; dashed lines
 denote most unstable disturbances.

a wide range of long waves neutral, while destabilizing short ones. The wave length of maximum instability is also shifted toward shorter wavelengths, as indicated by the dashed lines. We can also see that the neutral curve takes on a rather peculiar shape for very small magnetic field strengths. That is, the neutral curve for $P = 1/256$ adds only a small tail to the curve in the nonmagnetic case. Presumably as $P \rightarrow 0$, this tail shrinks to a line.

If the channel has infinite width, or at least is wide enough so that $(\lambda/l)^2 < 1$, then the abscissa in Fig. 4 is approximately $\lambda^2/10$. In channels with finite widths, A will be bounded as $\lambda^2 \rightarrow \infty$, so that in Fig. 4, only the region to the left of the limiting value of A will apply. For example, if $l = 2$, $A(\text{max}) = 1.6$. For relatively narrow channels, then, Fig. 4 does not give a very good representation of the region of instability. Therefore, since in Chapter 9, we will have to consider only narrow channels, we shall plot the stability curve as a function of λ^2 , for different channel widths. Also, since we are primarily interested in the effect of the magnetic field, we shall use as the ordinate the parameter P/U_0^2 , which we may call a magnetic Richardson number. The growth rates are then in units of $k c_i / U_0$, the phase velocities (measured relative to the flow in the lower layer i.e., we set $U_1 = 0$), as C_r / U_0 . Three channel widths are chosen: $l = \infty, 2, 1$. The values $E \sin^2 \phi_0 = 1/2$, $G = 0.177$ are used again. The results are plotted in Figs. 5 and 6. The values $l = 2, 1$ were chosen for ease of comparison with the results of Chapter 9, where the horizontal shear is included.

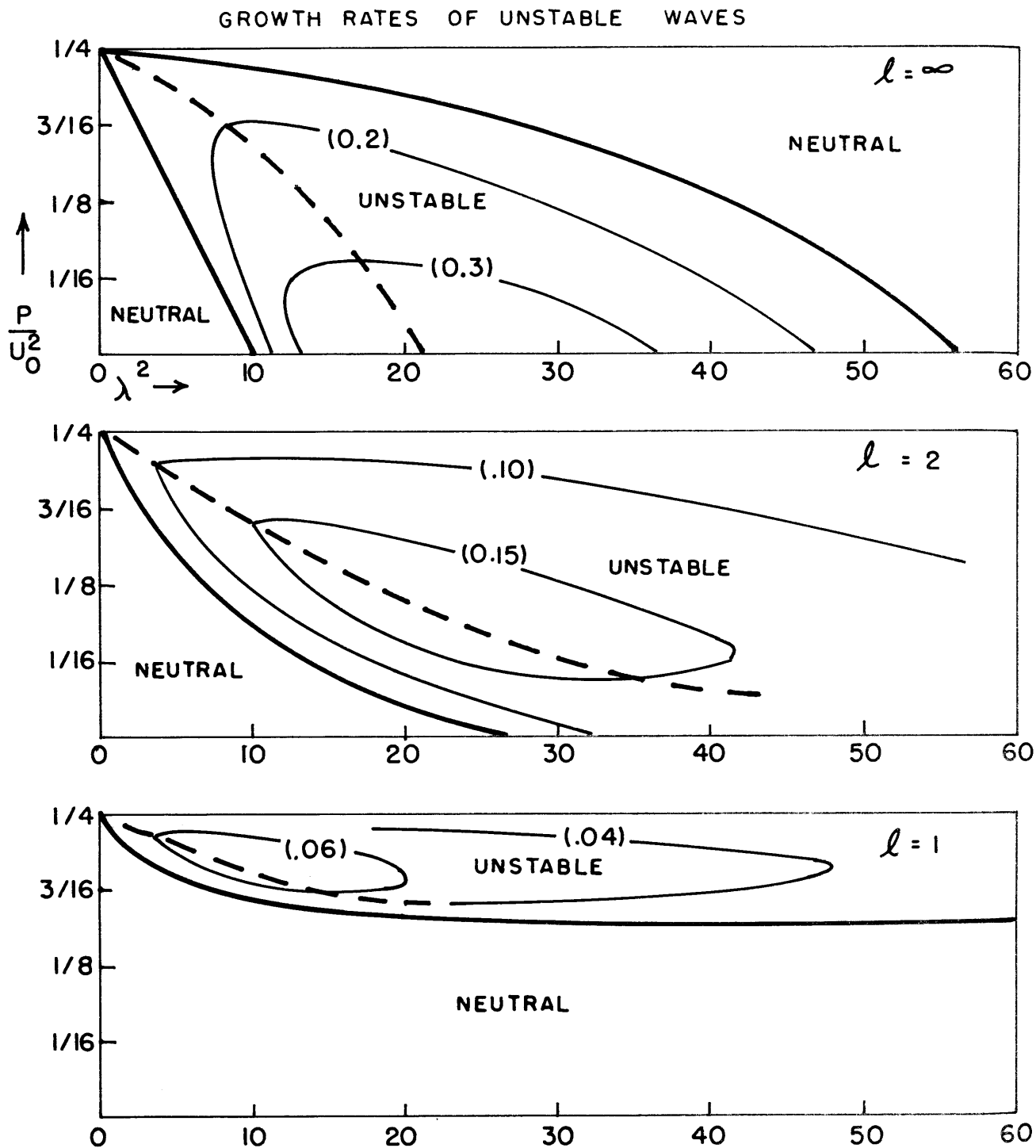


Fig. 5. Growth rates Kc_i/U_0 for channels of width $l = \infty$, 2, 1, with $G = 0.177$. Ordinate is P/U_0^2 , a magnetic Richardson number, abscissa is square of wavelength (λ in units of 10^5 km). Dashed lines denote most unstable waves. Reciprocal of growth rate gives "e folding" time approximately in solar rotations.

PHASE VELOCITIES OF UNSTABLE WAVES

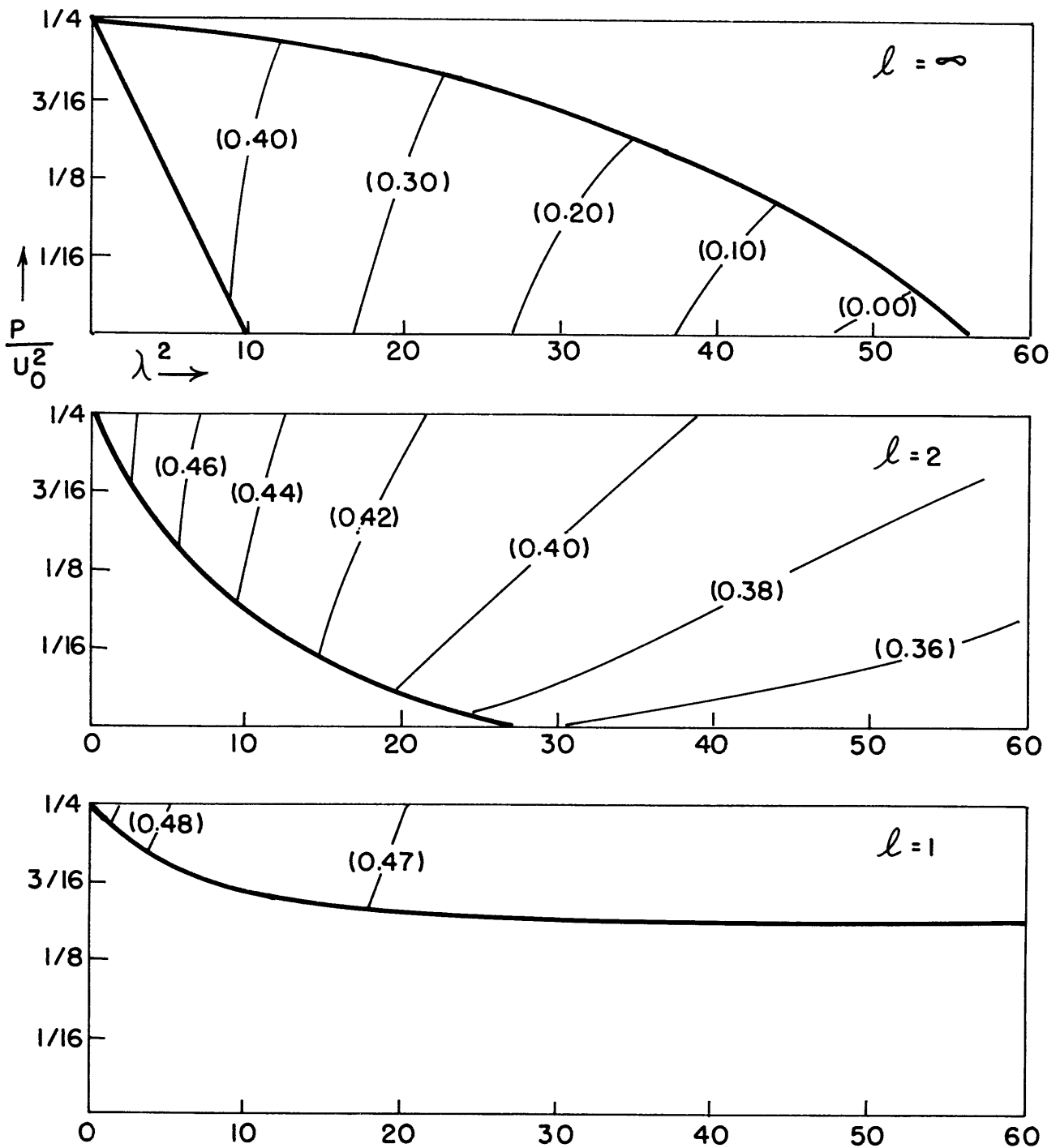


Fig. 6. Phase velocities C_r/U_0 of unstable waves whose growth rates are given in Fig. 5. C_r measured relative to zonal flow at level 1.

Figures 5 and 6 demonstrate clearly how much the stability of the waves and their phase velocities depend upon the channel width and the toroidal field strength. Maximum growth rates decrease strongly with decreasing channel width, ranging from 0.3/solar rotation for $l = \infty$, to 0.06/solar rotation or so for $l = 1$.

Even more striking are the changes in the neutral curves which bound the unstable region. For the narrow channels, the stabilizing effect of β for long waves is much reduced. In contrast, short wavelengths, particularly for relatively weak magnetic fields, become stable for the narrow channels. For example, with $l = 2$, for small magnetic fields ($P/U_0^2 \ll 1/4$) the shortest unstable wavelength is almost doubled. For $l = 1$, all wavelengths are neutral for $P/U_0^2 < \sim 0.15$.

On the other hand, contrary to the Eady case, a narrowed channel does not render all wavelengths neutral for all magnetic field strengths. However narrow the channel, there is a finite range of magnetic fields of which disturbances longer than a certain wavelength are unstable. For very wide channels, increasing the field strength while holding the shear constant results in a reduced growth rate for long waves ($\lambda^2 \gg \sim 15$) but can lead to increased growth rates at small and moderate field strengths at short wavelengths ($\lambda^2 < 15$). For the narrow channels, at almost all wavelengths the maximum growth rate occurs at some intermediate value of P/U_0^2

Finally, consistent with the reduced β effect on the neutral curves, the narrow channels also have much smaller ranges of phase velocities (Fig. 6). For a given shear and wavelength the phase velocities increase with the magnetic field strength.

Structure of disturbances and energy conversions

In addition to defining the regions of instability, it is of interest to examine the structure of the unstable disturbances, and the energy conversions this structure implies.

The structure will be described completely when the relative phase angles and amplitudes of all the dependent variables are found. Let us measure these relative to $\hat{\psi}_1$, the stream function at the level 1. That is, we set $\hat{\psi}_1 \equiv 1$.

The amplitude and phase of $\hat{\psi}_3$ can then be found from either (8-6) or (8-7). Using the definitions (8-8), we may write, from (8-7)

$$\hat{\psi}_3 = \frac{-\frac{A}{2} \left(\frac{U_0}{2} - c' \right)^2}{P - \left(\frac{U_0}{2} - c' \right)^2 + \left(\frac{U_0}{2} - c' \right) \left(A G + A \frac{U_0}{2} \right) - \frac{A}{2} \left(\frac{U_0}{2} - c' \right)^2} \quad (8-38)$$

From (8-4), (8-5) and (8-8), we may write

$$\hat{\chi}_1 = \frac{-H}{c' + \frac{U_0}{2}} \quad (8-39)$$

$$\hat{\chi}_3 = \frac{H \hat{\psi}_3}{\frac{U_0}{2} - c'} \quad (8-40)$$

To find the amplitudes and phases of the vertical motion and vertical magnetic field, we first define

$$w_2, h_2 = \left[\hat{w}_2, \hat{h}_2 \right] \sin \frac{\pi y}{l} e^{ik(x-ct)} \quad (8-41)$$

Then substitution for Ψ_1 , Ψ_3 , from (8-1) and (7-28), and for w_2 from (8-41), into (7-3), using the definitions (8-8), gives

$$\hat{w}_2 = 2ik\epsilon \sin\phi \left[\hat{\Psi}_3 \left(\frac{U_0}{2} + c' \right) + \left(\frac{U_0}{2} - c' \right) \right] \quad (8-42)$$

Finally, substituting for w_2 , h_2 from (8-41) into (7-8), we get

$$\hat{h}_2 = -\frac{H \hat{w}_2}{c'} \quad (8-43)$$

To simplify matters, let us in calculating the phases neglect the β effect and take the channel to have infinite width. It is then easy to relate the cases of positive and negative zonal flow. Since $\beta = 0$, $c_r' = 0$. Then substitution for c'^2 from (8-20) into (8-38), after considerable algebraic manipulation, yields

$$|\hat{\Psi}_3| = 1 \quad (8-44)$$

Thus the stream function amplitudes are the same in the two layers. From (8-39) and (8-40), then, it follows that

$$|\hat{\chi}_1|, |\hat{\chi}_3| = \frac{|H|}{\left(c_i^2 + \frac{U_0^2}{4}\right)^{1/2}} \quad (8-45)$$

Therefore the magnetic flux functions have the same amplitudes in the two layers.

Let us define the phase angles of the dependent variables in the following manner:

$$\left. \begin{aligned}
 \hat{\psi}_1 &= |\hat{\psi}_1| \\
 \hat{\psi}_3 &= |\hat{\psi}_3| e^{i\alpha_3} \\
 \hat{\chi}_1 &= |\hat{\chi}_1| e^{i\gamma_1} \\
 \hat{\chi}_3 &= |\hat{\chi}_3| e^{i\gamma_3} \\
 \hat{w}_2 &= |\hat{w}_2| e^{i\alpha_w} \\
 \hat{h}_2 &= |\hat{h}_2| e^{i\gamma_h}
 \end{aligned} \right\} \quad (8-46)$$

Then, from (8-39) and (8-40)

$$\gamma_3 = \pi + \alpha_3 - \gamma_1 \quad (8-47)$$

The phase angles for positive and negative but equal magnitude shears can be easily related. Using a bar to distinguish between the two cases, the phases are related according to

$$\begin{aligned}
 \bar{\alpha}_3 &= -\alpha_3 \\
 \bar{\alpha}_w &= -\alpha_w \\
 \bar{\gamma}_1 &= \pi - \gamma_1 \\
 \bar{\gamma}_3 &= \pi - \gamma_3 \\
 \bar{\gamma}_h &= \pi - \gamma_h
 \end{aligned} \quad (8-48)$$

We have calculated the phase angles of the various dependent variables for the case $G=0$, $l=\infty$, and again $\epsilon \sin^2 \phi_0 = 1/2$, along the curve of maximum growth rate. These are presented in Figs. 7b, c.

UNSTABLE DISTURBANCE STRUCTURE

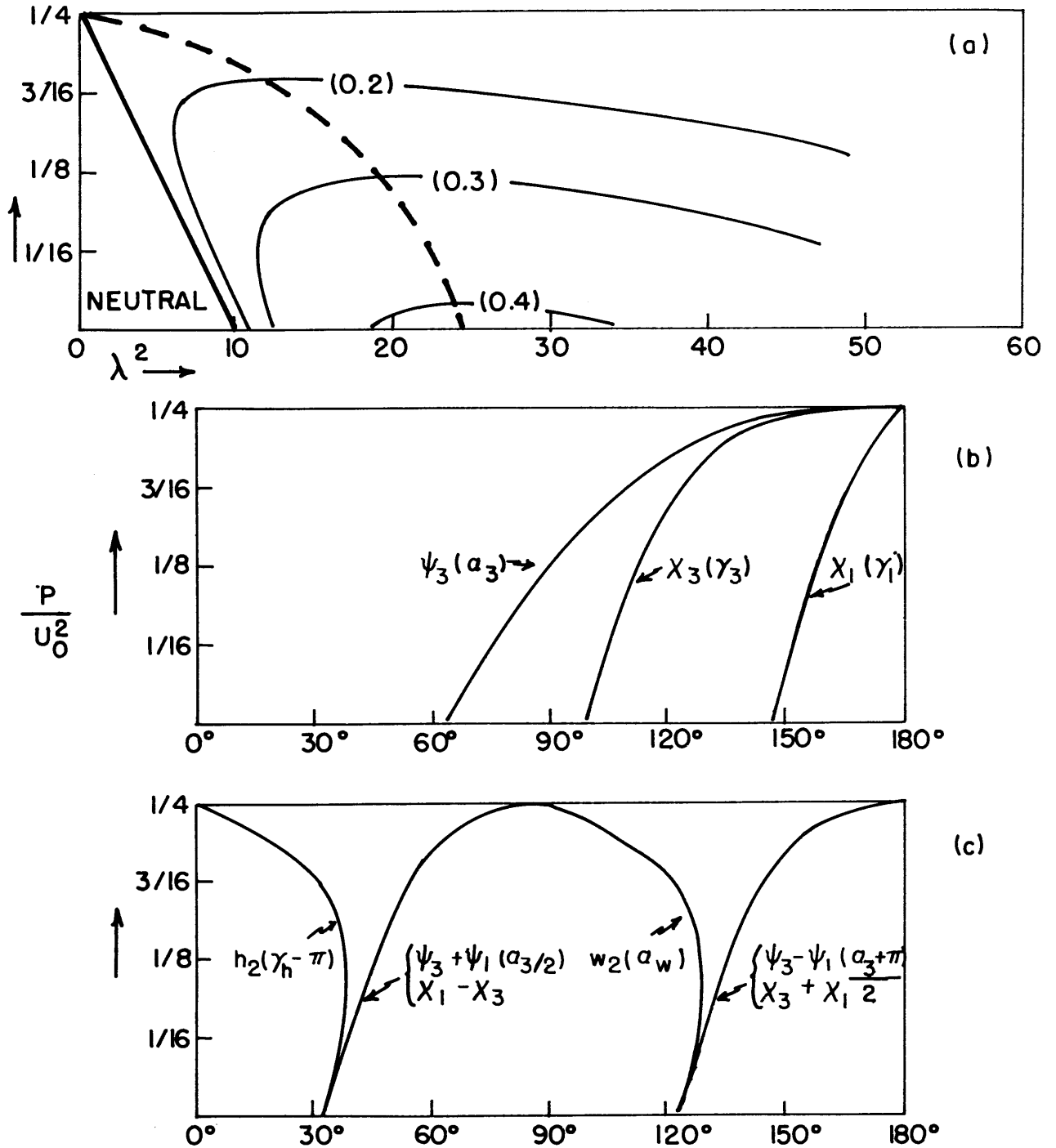


Fig. 7. a) Growth rates ki/U_0 of unstable waves for $l = \infty$, $G = 0$ (no β effect). Ordinate and abscissa same as in Fig. 5. Dashed curve denotes most unstable waves.
 b, c) Phase angles of disturbance variables for most unstable waves of Fig. 7a. Phases measured relative to ψ_1 (see definitions (8-46) in text).

Figure 7a gives the unstable region and the curve of maximum growth rate (dashed). These are for positive shear U_0 (cold pole, warm equator) and positive toroidal field.

It is easiest to interpret these phase relationships in terms of the energy conversions they imply. Since the zonal flow and toroidal magnetic field are initially independent of y , the only non zero conversions of energy initially will be $(\langle A \rangle, A')$, (A', K') , (K', M') . In the absence of a magnetic field, the first two conversions are positive for baroclinically unstable waves. That is, there is a conversion from zonal available potential to eddy available potential, some of which in turn is converted into eddy kinetic energy. With the toroidal field present, we would expect these two conversions to go on qualitatively as before, though altered somewhat by the toroidal field, and to have an additional process, represented by (K', M') , in which part of the eddy kinetic energy is further converted into eddy magnetic energy. From Figs. 7a, b we can see that this is indeed the case for the most unstable waves. The generally stabilizing effect of the toroidal field strength as it is increased is also quite evident.

In order for the first link in this chain of energy conversions to be in the desired sense, the mean temperature field $\psi_3 - \psi_1$, must lag the mean stream field $(\psi_3 + \psi_1)/2$, so that there will be an eddy heat flux from warm to cold latitudes. From Fig. 7c, it is evident that the amount of this lag is 90° , no matter what the magnetic field strength. As this field strength is increased, this conversion nonetheless

decreases toward zero. This is effected through the tending toward zero of the mean stream field $(\Psi_3 + \Psi_1)/2$. That is, although $|\hat{\Psi}_1| = |\hat{\Psi}_3| = 1$, the angle $\alpha_3 \rightarrow 180^\circ$ (see Fig. 7b), so that $\hat{\Psi}_3 \rightarrow -\hat{\Psi}_1$.

The second conversion, (A', K') , will be positive only if there is net rising of warm and sinking of cold gas. In terms of the phase angles, this will be true if $0 < \alpha_w - \alpha_3/2 < \pi$. From Fig. 7c, it is evident that this is always the case. The stabilizing effect on this conversion of increasing the magnetic field strength results in $\alpha_w - \alpha_3/2 \rightarrow 0$; i.e. the phase difference between the vertical motion and temperature perturbations approaches 90° , shutting off the conversion. In addition, the magnitude of the vertical motion approaches zero, since $\hat{\Psi}_3 \rightarrow -1$, and $C' \rightarrow 0$, as we can see from (8-42). (If the β effect were included, this would no longer be the case).

Finally, the conversion (K', M') will be positive if $0 < \delta_1 < \pi$ (from which it follows that $0 < \delta_3 - \alpha_3 < \pi$, by 8-47). From Fig. 7b this is true for the most unstable waves. Since $\sin \delta_1 = C_i / (C_i^2 + \frac{U_0^2}{4})^{1/2}$, it will be true for all waves for which $C_i > 0$, i.e., for all unstable waves. As the field strength is increased, $\delta_1 \rightarrow 180^\circ$, since $C_i \rightarrow 0$, thus shutting off this last conversion.

Although the above discussion was applied directly to the case of $U_0 > 0, H > 0$, using the relations (8-48), we may easily carry

it through for negative shears and/or negative toroidal fields. Inclusion of the β effect will make the phase relations somewhat more complicated, but should not alter the fundamental result. In short, in the unstable baroclinic hydromagnetic disturbances we have found, zonal available potential energy is converted into eddy available potential energy, some of which is further converted into eddy kinetic energy, some of which in turn is converted to eddy magnetic energy.

Changes in the initial zonally symmetric state

The phase relations presented in Fig. 7, and discussed in the previous section, also imply changes in the initial zonally symmetric state. These changes are defined by the tendency equations (7-9)-(7-14). In the absence of horizontal shear in either the zonal flow or the toroidal field, the relative phases of the various variables are, of course, independent of latitude. This means that the unstable waves are not tilted, so that the Reynolds and Maxwell stresses vanish. From equations (7-9), (7-10) and (7-11), then, we can see that we may solve for the time tendency in the mean stream function as if the magnetic field were not present. Its effect is, of course, still felt, since the phase relations are a function of the magnetic field strength.

The solution of the closed system (7-9)-(7-11) for $\langle \psi_1 \rangle$ and $\langle \psi_3 \rangle$ thus is formally precisely the same as that found by Phillips (1954) for the nonmagnetic case. We may solve for $\partial H_{1,3} / \partial t$ separately at levels one and three, from (7-12) and (7-13). Since $H_1 = H_3 = H$ they will give the same answer. Finally, as we saw from (8-43), the unstable

disturbances produce an exponentially growing perturbation magnetic field

h_2 . From equation (7-14) we can also find the symmetric meridional (poloidal) magnetic field $\langle h_2 \rangle$ produced by the disturbances.

In the absence of Reynolds and Maxwell stresses, and noting that

$$\theta_2 = 2 \sin \phi_0 \langle \psi_3 - \psi_1 \rangle, \quad \text{the tendency equations (7-9)-(7-11) for}$$

the flow and temperature fields may be written in terms of the mean and perturbation stream functions as

$$\frac{\partial \langle \psi_3 \rangle}{\partial t} - \frac{\partial \langle \psi_1 \rangle}{\partial t} = - \frac{\langle w_2 \rangle}{2E \sin \phi_0} - \frac{1}{2} \frac{\partial}{\partial y} \left\langle \left(\psi_3' - \psi_1' \right) \frac{\partial}{\partial x} (\psi_3' + \psi_1') \right\rangle \quad (8-49)$$

$$\frac{\partial^2}{\partial y^2} \frac{\partial \langle \psi_1 \rangle}{\partial t} = 2 \sin \phi_0 \langle w_2 \rangle \quad (8-50)$$

$$\frac{\partial^2}{\partial y^2} \frac{\partial \langle \psi_3 \rangle}{\partial t} = -2 \sin \phi_0 \langle w_2 \rangle \quad (8-51)$$

From equations (8-50) and (8-51), and the boundary condition (7-15);

i.e. $\frac{\partial^2 \langle \psi_{1,3} \rangle}{\partial y^2 \partial t} = 0$ at $y = 0, l$ it is evident that

$$\frac{\partial \langle \psi_1 \rangle}{\partial t} = - \frac{\partial \langle \psi_3 \rangle}{\partial t} \quad (8-52)$$

Then substituting (8-52) and (8-50) into (8-49) for $\frac{\partial \langle \psi_3 \rangle}{\partial t}$ and $\langle w_2 \rangle$

respectively, and evaluating the eddy heat flux term, we obtain a single equation for $\frac{\partial \langle \psi_1 \rangle}{\partial t}$ namely

$$\frac{\partial^2}{\partial y^2} \frac{\partial \langle \psi_1 \rangle}{\partial t} - 4m \frac{\partial \langle \psi_1 \rangle}{\partial t} = -E \sin \frac{2\pi y}{l} \quad (8-53)$$

where we have defined

$$m = 2\epsilon \sin^2 \phi_0 \quad (8-54)$$

$$E = 2m \frac{k\pi}{l} \sin \alpha_3 e^{2kct} \quad (8-55)$$

Since $\partial^2 \langle \psi_1 \rangle / \partial y \partial t = 0$ at $y = 0, l$, the complete solution of (8-53) is of the form

$$\begin{aligned} \frac{\partial \langle \psi_1 \rangle}{\partial t} = & -\frac{\pi D}{l\sqrt{m}} \left[\sinh 2\sqrt{m} y + \frac{1 - \cosh 2l\sqrt{m}}{\sinh 2l\sqrt{m}} \cosh 2\sqrt{m} y \right] \\ & + D \sin \frac{2\pi y}{l} \end{aligned} \quad (8-56)$$

where $D = E / 4(m + \pi^2/l^2) \quad (8-57)$

The sign of $\partial \langle \psi_1 \rangle / \partial t$ is thus determined by the sign of D , which, from (8-55), is determined by $\sin \alpha_3$. Together with (8-52), (8-56) then implies that, for $U_0 > 0$ (cold pole, warm equator)

$$\begin{aligned} \frac{\partial}{\partial t} \langle \psi_3 - \psi_1 \rangle &< 0 & 0 < y < l/2 \\ \frac{\partial}{\partial t} \langle \psi_3 - \psi_1 \rangle &> 0 & l/2 < y < l \end{aligned} \quad (8-58)$$

For $U_0 < 0$ (warm pole, cold equator), the inequalities in (8-58) are reversed. They are unaffected by a change in sign of the toroidal field. Thus from (8-58) the meridional eddy flux of heat is raising the temperature

in the cold latitudes, and lowering it in warm latitudes, just as in the nonmagnetic case.

Substituting (8-56) into (8-50), we may calculate the symmetric vertical and meridional circulation which results, given by

$$\langle W_2 \rangle = -\frac{2D\pi\sqrt{m}}{l\sin\phi_0} \left[\sinh 2\sqrt{m}y + \frac{1 - \cosh 2l\sqrt{m}}{\sinh 2l\sqrt{m}} \cosh 2\sqrt{m}y + \frac{\pi}{l\sqrt{m}} \sin \frac{2\pi y}{l} \right] \quad (8-58)$$

$$\langle V_3 \rangle = -\frac{2D\pi}{l\sin\phi_0} \left[\cosh 2\sqrt{m}y + \frac{1 - \cosh 2l\sqrt{m}}{\sinh 2l\sqrt{m}} \sinh 2\sqrt{m}y - \cos \frac{2\pi y}{l} \right] \quad (8-59)$$

Equation (8-59) is found from integrating the symmetric continuity equation

$$\frac{\partial \langle V_3 \rangle}{\partial y} = 2 \langle W_2 \rangle = -\frac{\partial \langle V_1 \rangle}{\partial y} \quad (8-60)$$

applying the boundary condition $\langle V_{1,3} \rangle = 0$ at $y = 0, l$.

From (8-58) and (8-59), then, we can see that just as in Phillips (1954), a three celled symmetric meridional circulation develops, with thermally direct (warm gas rising, cold gas sinking) cells in low and high latitudes, and a thermally indirect one in middle latitudes.

Finally, by equations (8-50) and (8-51), we see that the mean vertical circulation alters the zonal flow in each layer. For positive shear U_0 , the flow in the upper layer (level 3) is decreased in middle latitudes, and increased in low and high latitudes. The opposite is true at level 1. (The flow in each layer remains unchanged at $y = 0, l$, since $\partial \langle V_{1,3} \rangle / \partial t = 0$ there).

As stated, the changes in the toroidal field can be found from (7-12), (7-13). Substituting the real parts of the separated solutions for $\psi_1, \psi_3, \chi_1, \chi_3$ from (8-1) and (7-28), we get, from both (7-12), (7-13)

$$\frac{\partial H}{\partial t} = H \left(\frac{\pi}{l} \right)^2 \frac{\sin \delta_1}{(c_i^2 + \frac{v_0^2}{4})^{1/2}} \cos \frac{2\pi y}{l} e^{2k c_i t} \quad (8-61)$$

Thus only the unstable disturbances, for which $\delta_1 \neq 0, \pi$, can change the toroidal field. Since, from (8-48), the angles for positive and negative vertical shears are related according to $\bar{\delta}_1 = \pi - \delta_1$, $\sin \delta_1 = \sin \bar{\delta}_1$, so that the toroidal field tendency is independent of the sign of the shear in the zonal flow.

From (8-61), then, we see that the unstable disturbances, through the "mixed stress", act to decrease the magnitude of the toroidal field in middle latitudes, and increase it near the poles and equator. Since the total toroidal flux must remain constant (easily seen by integrating (7-12), (7-13) from wall to wall and applying the boundary conditions (7-15)), the disturbances must carry flux from middle into low and high latitudes.

Lastly, from (7-14), we may calculate the symmetric poloidal field $\langle h_2 \rangle$ produced by the disturbances, which we may write as

$$\frac{\partial \langle h_2 \rangle}{\partial t} = - \frac{H k \pi}{4 l} \left[|\hat{w}_2| / |\hat{\chi}_1 + \hat{\chi}_3| \cos(\alpha_w - \alpha_3/2) + |\hat{h}_2| / |\hat{\psi}_1 + \hat{\psi}_3| \sin(\delta_n - \alpha_3/2) \right] e^{2k c_i t} \sin \frac{2\pi y}{l} \quad (8-62)$$

For the most unstable disturbances, from Fig. 7c, we can see that both terms in the brackets of (8-62) are positive (for both positive and negative vertical shear in the zonal flow). Therefore if $H > 0$

$$\frac{\partial \langle h_z \rangle}{\partial t} < 0 \quad 0 < y < l/2$$

$$\frac{\partial \langle h_z \rangle}{\partial t} > 0 \quad l/2 < y < l$$
(8-63)

For $H < 0$, the inequalities are reversed.

Maximum amplitude occurs at $y = l/4, 3l/4$. The meridional component associated with (8-62) is easily found from

$$\frac{\partial \langle H_{3y} \rangle}{\partial y} = - \frac{\partial \langle H_{1y} \rangle}{\partial y} = 2 \langle h_z \rangle$$
(8-64)

Using the boundary condition $\langle H_{1y} \rangle, \langle H_{3y} \rangle = 0$ at $y = 0, l$. The horizontal fields are of opposite signs in the two layers, and all of one sign within each layer, of the form

$$\frac{\partial \langle H_{1y} \rangle}{\partial t} = - \frac{\partial \langle H_{3y} \rangle}{\partial t} = - \frac{H h K}{4} e^{2kt} \left[1 - \cos \frac{2\pi y}{l} \right]$$
(8-65)

where K stands for the factor in square brackets in (8-62).

Thus the disturbances produce a single celled poloidal field, such that the vertical field is of the same sign in high latitudes as the original toroidal field, and of the opposite sign in low latitudes.

Application to the sun

Because this is a perturbation study, and because of the many simplifying and restricting assumptions made, application of the results to the solar observations cannot be very precise. We should instead limit ourselves to comparison of qualitative effects and orders of magnitude.

First of all, we can see that unless U_0 is only slightly larger than $2\sqrt{P}$, the most unstable disturbances have wavelengths comparable to those of the field patterns in Bumba and Howard's (1965b) magnetograms (see also the Appendix). For example, in Fig. 4, if we assume $\lambda/l < 1$, $A \sim \lambda^2/10$, where λ is in units of 10^5 km, so that for $P = 1/8$, $U_0 = 1.2$, the most unstable wavelength is about 4×10^5 km, corresponding to a zonal wave number (lower abscissa scale in Fig. 4) of about 8, as compared to a crude estimate of 6 or 7 from the magnetograms. For an initial state in which U_0 is only slightly above $2\sqrt{P}$, the most unstable wavelength becomes significantly shorter. These short waves are, however, of more limited significance, since for them some of the scaling approximations break down (the Rossby number becomes too large).

The condition $U_0 > 2\sqrt{P}$ required for instability can be looked upon as a magnetic constraint on unstable waves. For a given magnetic field then, a certain minimum amount of available potential energy will be needed to overcome this constraint. Within the scaling assumptions of our model, we can estimate how large a meridional temperature gradient would be needed to give unstable disturbances in the presence of a reasonable solar toroidal field.

Suppose we take $M = 100$ gauss. Since we are applying our model to a layer deep in the convection layer, we take $\rho \sim 10^{-4}$ gm/cm³. Then $\bar{P} = 3/16$ corresponds to a scale velocity $U \sim 65$ m/sec. In setting down the scaling requirements in Chapter 4, we assumed the vertical scale D in the flow is of the same order or less than the scale height S' . We may take $S' \sim 10^3$ km at these depths, so that if $D \sim S'$, the shear $\partial u^*/\partial z^* \sim 6.5 \times 10^{-5}$ sec⁻¹. This is about 10^{-2} of a typical terrestrial atmospheric value. The dimensional form of the thermal wind relating the shear to the meridional temperature gradient may be written approximately as

$$\frac{\partial u^*}{\partial z^*} = - \frac{g}{2\Omega \sin\phi_0} \frac{\partial \ln T^*}{\partial y^*} \quad (8-66)$$

For the sun $g = 2.7 \times 10^4$ cm sec⁻², $2\Omega \sin\phi_0 = 3 \times 10^{-6}$ sec⁻¹, so that the fractional temperature difference $\Delta T/T$ between equator and pole would be about

$$\Delta T/T \sim 7 \times 10^{-4} \quad (8-67)$$

Thus for an average temperature $T \sim 50,000^\circ\text{K}$, the equator to pole temperature difference need be only $\Delta T \sim 35^\circ\text{K}$. This value is well within the spread of values summarized by Beckers (1960) for the photosphere.

It does not seem too implausible to expect the small scale convection, in the manner outlined in Chapter 3, to produce a ΔT of this size, since the ratio $\Delta T/T$ is so small (only about 1/300 of its value for the sun - imposed gradient in the earth's atmosphere).

The interpretation of the requirement $U_0 > 2\sqrt{P}$ for instability can, of course, be turned around. If we can place limits on the equator to pole temperature differences, then we can place an upper limit on the uniform toroidal field strengths attainable. If we use the maximum photosphere temperature difference reported ($\sim 100^\circ\text{K}$) as an upper bound on the value in the deeper layer we are considering, the maximum uniform toroidal field strength is about 400 gauss.

The perturbation vertical magnetic fields h'_z produced by the unstable disturbances, though of second order, are important because they may be compared directly to the magnetograms. We already know they will have the proper horizontal scale. In magnitude, if $\delta \sim 10^{-2}$, $R_0 \sim 10^{-1}$, $M \sim 100$ gauss, then the perturbation vertical fields will be of order $\delta R_0 M \sim 10^{-1}$ gauss, about one order of magnitude less than observed. This difference is probably not significant, since taking $\delta \sim 2 \times 10^{-2}$, $R_0 \sim 2 \times 10^{-1}$, $M \sim 200$ gauss gives $\delta R_0 M \sim 1$ gauss. Thus the vertical fields produced by our model are consistent with the observed values both in magnitude and horizontal scale.

By our scaling arguments, the symmetric poloidal field $\langle h_z \rangle$ produced should be an order of magnitude smaller than h'_z , i.e. 10^{-2} to 10^{-1} gauss. However, the observed value for the sun is often as large as 1 gauss. This discrepancy is not unreasonable if the observed value is the cumulative result of several successive disturbances.

On the sun, the unipolar field is seen only near the poles, while our model has it just as strong near the equator. However, if the poloidal

field produced in the model is still single celled when spherical geometry is introduced, the vertical fields would be much stronger near the poles, due to the convergence of meridians.

There is another property of these poloidal fields that is of interest. In the sun, the lower horizontal branch of the poloidal field is presumably much more concentrated than the upper branch, which extends into the corona and even beyond. It is this lower branch which Babcock (1961) has postulated is drawn out into the zonal direction by the differential rotation. This process was supposed to produce the new toroidal field from which a new sunspot cycle emerged.

In our model, we can see from (8-65) that the lower branch of the poloidal field $\langle H_{1y} \rangle$ is such that, if it were stretched into the zonal direction in the manner put forth by Babcock, it would create a toroidal field of the opposite sign from the original field. Thus, if in accordance with our remarks at the end of Chapter 5, we relax the scaling restriction $L/a \ll 1$, Babcock's process will be taking place in our system. Since we know that the conversion $(K', \langle M \rangle)$ defined in (5-25), which involves the "mixed stress", cannot over long time periods maintain the energy in the toroidal field by itself (c.f. the end of Chapter 4 and the end of Chapter 5), the energy conversion $(\langle K \rangle, \langle M \rangle)$ implied by Babcock's process may indeed give the necessary boost.

However, the reappearance of sunspots and the progression of the favored latitude for their formation toward the equator would seem to be more strongly determined by the "mixed stress" in (5-7) and (7-12), (7-13)

than by Babcock's process. That is, we showed in equation (8-61) that the effect of the large scale baroclinic disturbances is to decrease the magnitude of the toroidal field in middle latitudes, and increase it in low and high latitudes. If sunspots preferentially form where the subsurface toroidal field is strongest, as is often assumed, then this effect could give rise to the equatorward progression of the zone of sunspot formation. Admittedly, this process also allows for sunspots forming closer to the poles as the cycle progresses, but this may be suppressed by other factors, such as the spherical geometry, when they are included.

In closing, then, we see that even in this simple case, the structure of the unstable baroclinic hydromagnetic disturbances have many properties of interest for the solar problem. In the next chapter, in which we include horizontal as well as vertical shear in the zonal flow, several other features of interest will come to light. Most importantly, we shall see how the horizontal Reynolds stresses could maintain the differential rotation, though opposed by the Maxwell stresses, and how the magnetic fields in the disturbances take on a tilted structure not unlike that observed on the magnetograms.

9. Stability of two layer baroclinic flow with horizontal shear
in a uniform zonal magnetic field.

Derivation of the eigenvalue equation

Because we know the sun has a differential rotation, it is desirable to generalize the stability study in Chapter 8 to include horizontal as well as vertical shear. In particular, given the form of the differential rotation defined by its average state over several sunspot cycles (Fig. A-1), we would like to examine the stability properties of a flow which is parabolic at least in the upper layer.

In general, the mathematical approach to be employed in a stability study of flow of continuously varying shear depends critically on the possible existence and number of singularities in the equations which fall within the region of interest. If there are no such singularities, then the solution, while it may be computationally tedious, is straightforward: the dependent variables may be expanded in a power series in the meridional coordinate (y) , and the coefficients calculated from the resulting recursion relation, after applying the proper boundary conditions.

In essence, this is what Pedlosky (1964b) did for the nonmagnetic case. He chose to confine the horizontal shear to one, the upper, layer, taking the flow in the lower layer to be uniform. In this nonmagnetic case, singularities will become a problem only if, on the neutral curves which bound an unstable region (often called curves of marginal stability), the phase velocity lies in the range of the zonal flow in the upper layer. Pedlosky showed that for profiles in the upper layer which had no extremum

of potential vorticity, the phase velocity of marginally stable waves must be less than the minimum velocity in the upper layer. Thus this type of flow profile would not give rise to singularities in the equations, and the power series technique could be used to completely define the unstable region.

Pedlosky then analyzed numerically one profile of this type, which was of parabolic shape in the upper layer, and uniform (for definiteness, zero) in the lower layer. In our notation, this flow is of the form

$$U_3 = U_0(1 - by^2); U_1 = 0 \quad 0 \leq y \leq 1 \quad (9-1)$$

where $0 \leq b \leq 1$.

He found that the same minimum value of vertical shear was required for instability as when no horizontal shear was present (just enough vertical shear to make the potential vorticity in the lower layer negative). However, the wavelength at which this minimum vertical shear gave unstable disturbances was shorter, as was, in general, the most unstable wavelength and range of unstable wavelengths for any given shear. In general, for a given shear, the phase velocities were smaller, reflecting an increased β effect due to the positive mean zonal vorticity in the flow. The imaginary part C_i was also smaller, but due to the shift of the unstable region toward shorter wavelengths the growth rate $k C_i$ was not necessarily smaller.

Examination of the structure of the unstable disturbances indicated that horizontal Reynolds stresses were acting to increase the horizontal shear, i.e. they were transporting momentum against the gradient. In order to do this, these disturbances took on a characteristic tilted structure,

upstream away from the maximum of mean zonal flow. At the same time, just as in the case of no horizontal shear, the disturbances released zonal available potential energy by transporting heat from warm to cold latitudes, and converted the resulting eddy available potential energy into eddy kinetic through rising motion at warm and sinking at cold longitudes. Thus these disturbances were in a sense baroclinically unstable and barotropically stable, and as such were performing the energy conversions in basically the same manner as do the cyclones and anticyclones in the actual atmosphere.

We would like to extend Pedlosky's analysis to the magnetic case, to see whether disturbances still exist with this chain of energy conversions, but which in addition show the action of horizontal Maxwell stresses in braking the mean zonal flow as suggested in Starr and Gilman (1965b), and which produce vertical fields whose gross horizontal structure is similar to the solar magnetograms of Bumba and Howard (1965b).

However, with the addition of a zonal magnetic field, even a uniform one, we are no longer able to classify the flow profiles according to the marginally stable waves they allow, as Pedlosky did. Thus, for flows without potential vorticity extrema, such as the parabolic flow we wish to consider, we cannot be sure there are no marginally stable disturbances with phase velocities which lie within the range of the flow, giving us singularities in the equations. This difficulty is closely related to the fact that in the magnetic case, the potential vorticity is no longer conserved.

In addition, inclusion of the magnetic field introduces new singularities into the equations, which we will discuss in detail when the equations are derived.

Nevertheless, since we may increase the magnetic field strength continuously from zero, it seems reasonable to expect that we can find one class of unstable waves whose curves of marginal stability remain within Pedlosky's classification at least for moderate or small magnetic fields. This proves to be the case, although the singularities introduced by the magnetic field serve to reduce considerably the range of P/U_0^2 over which the power series expansion is valid, by an amount which increases with the horizontal shear parameter b . The series solutions we are able to find do give us the Reynolds and Maxwell stresses acting in the desired manner, and the vertical fields tilted in a way qualitatively similar to the magnetograms.

With a flow profile defined by (9-1) let us now reduce the problem to one equation, which most conveniently is written in terms of $\tilde{\chi}_1$. Let us define

$$m = 2\epsilon \sin^2 \phi_0 \quad (9-2)$$

$$I = U_0(1 - by^2) - c \quad (9-3)$$

Then, using (9-1) for the profile and remembering that $H_1 = H_3 = H$ we may write (7-31) and (7-32) as

$$\tilde{\psi}_1 = -\frac{c}{H} \tilde{\chi}_1 \quad (9-4)$$

$$\tilde{\psi}_3 = \frac{I}{H} \tilde{\chi}_3 \quad (9-5)$$

Then, substitution of (9-2)-(9-5) into (7-29) and (7-30), using (7-26), (7-27), and remembering that $P_s H^2 = P$, gives

$$(C^2 - P) \left(\frac{\partial^2 \tilde{\chi}_1}{\partial y^2} - k^2 \tilde{\chi}_1 \right) - \beta c \tilde{\chi}_1 - 2mcI(\tilde{\chi}_3 - \tilde{\chi}_1) = 0 \quad (9-6)$$

$$(I^2 - P) \left(\frac{\partial^2 \tilde{\chi}_3}{\partial y^2} - k^2 \tilde{\chi}_3 \right) + \frac{\partial}{\partial y} (I^2) \frac{\partial \tilde{\chi}_3}{\partial y} + 3I \tilde{\chi}_3 + 2mcI(\tilde{\chi}_3 - \tilde{\chi}_1) = 0 \quad (9-7)$$

We may solve (9-6) for $\tilde{\chi}_3$ in terms of $\tilde{\chi}_1$ and its derivatives and substitute into (9-7). If we, for convenience, further define

$$\begin{aligned} B &= C^2 - P \\ E &= k^2 B + 3C \end{aligned} \quad (9-8)$$

the final result may be written as

$$\begin{aligned} & BI^2(I^2 - P) \frac{\partial^4 \tilde{\chi}_1}{\partial y^4} + PB \frac{\partial}{\partial y} (I^2) \frac{\partial^3 \tilde{\chi}_1}{\partial y^3} \\ & + \left[(I^2 - P) (2cmI^3 - EI^2 + B \left[2 \left(\frac{\partial I}{\partial y} \right)^2 - I \frac{\partial^2 I}{\partial y^2} - I^2 k^2 \right]) \right. \\ & \quad \left. + B \left(I^3 [2mc + \beta] - \frac{1}{2} \left[\frac{\partial}{\partial y} (I^2) \right]^2 \right) \right] \frac{\partial^2 \tilde{\chi}_1}{\partial y^2} \\ & + \frac{\partial}{\partial y} (I^2) [2mcI^3 - PE] \frac{\partial \tilde{\chi}_1}{\partial y} \\ & + \left[(I^2 - P) \left(E \left[I \frac{\partial^2 I}{\partial y^2} - 2 \left(\frac{\partial I}{\partial y} \right)^2 + k^2 I^2 \right] - 2mck^2 I^3 \right) \right. \\ & \quad \left. + \frac{E}{2} \left(\frac{\partial}{\partial y} (I^2) \right)^2 - EI^3 (2mc + \beta) + 2mc\beta I^4 \right] \tilde{\chi}_1 = 0 \end{aligned} \quad (9-9)$$

To obtain the series solution, the coefficients in (9-9) will have to be expanded in powers of y , but first, let us look at the possible singularities. They will occur whenever any of the factors multiplying

$\partial^4 \tilde{\chi}_1 / \partial y^4$ vanish, i.e. where

$$[U_0(1-by^2) - c]^2 = 0 \quad (9-10)$$

$$[U_0(1-by^2) - c]^2 = P \quad (9-11)$$

$$c^2 = P \quad (9-12)$$

The first of these equations gives the possible singularities in the non-magnetic case. Pedlosky (1964b) showed that no neutral wave could have a phase velocity greater than the maximum flow velocity in the system, so that in order for this singularity to be avoided, we must have C_r less than the minimum velocity in the upper layer, i.e.

$$C_r < U_0(1-b) \quad (9-13)$$

In the magnetic case, we have proved in Chapter 7 that the phase velocity of a neutral wave cannot exceed the sum of the maximum flow velocity and the maximum Alfvén velocity in the system (see (7-50)). For parabolic flow in a uniform field, then, this requires $C_r < U_0 + \sqrt{P}$. Therefore, to avoid the singularities defined by (9-11), we must require (for $U_0 > 0$)

$$C_r < U_0(1-b) - \sqrt{P} \quad (9-14)$$

Thus if (9-14) is satisfied, (9-13) will be also, and both singularities will be avoided.

Finally, from (9-12), we must also have

$$C_r \neq \sqrt{P} \quad (9-15)$$

With these restrictions in mind, let us now write equation (9-9) in a form more suitable for expansion in a power series. For convenience, we define the following symbols

$$\left. \begin{aligned} F_1 &= C^2 - P \\ F_2 &= U_0 - C \\ F_3 &= -U_0 b \\ F_4 &= m C \\ F_5 &= \beta C + k^2(C^2 - P) \\ F_6 &= \beta C + 2k^2(C^2 - P) \\ F_7 &= 2mC + \beta + 2U_0 b \\ F_8 &= (C^2 - P)(2mC + \beta + 2U_0 b) - 2mC P \\ F_9 &= 2mC P k^2 - (2mC + \beta + 2U_0 b)(\beta C + k^2(C^2 - P)) \\ F_{10} &= 2mC \beta + k^2(\beta C + k^2(C^2 - P)) \end{aligned} \right\} (9-16)$$

Then the coefficients in equation (9-9) may be expanded out in the form

$$\begin{aligned} & \sum_{n=0}^4 Q_{2n} y^{2n} \frac{\partial^4 \tilde{\chi}_1}{\partial y^4} + \sum_{n=0}^1 R_{2n+1} y^{2n+1} \frac{\partial^3 \tilde{\chi}_1}{\partial y^3} \\ & + \sum_{n=0}^5 S_{2n} y^{2n} \frac{\partial^2 \tilde{\chi}_1}{\partial y^2} + \sum_{n=0}^4 T_{2n+1} y^{2n+1} \frac{\partial \tilde{\chi}_1}{\partial y} \\ & + \sum_{n=0}^5 W_{2n} y^{2n} \tilde{\chi}_1 = 0 \end{aligned} \quad (9-17)$$

Where we have defined

$$Q_0 = F_1 F_2^4 - P F_1 F_2^2$$

$$Q_2 = 4 F_1 F_2^3 F_3 - 2 P F_1 F_2 F_3$$

$$Q_4 = 6 F_1 F_2^2 F_3^2 - P F_1 F_3^2$$

$$Q_6 = 4 F_1 F_2 F_3^3$$

$$Q_8 = F_1 F_3^4$$

$$R_1 = 4 P F_1 F_2 F_3$$

$$R_3 = 4 P F_1 F_3^2$$

$$S_0 = 2 F_4 F_2^5 - F_6 F_2^4 + F_8 F_2^3 + P F_6 F_2^2 + 2 P F_1 F_2 F_3$$

$$S_2 = 10 F_4 F_2^4 F_3 - 4 F_6 F_2^3 F_3 + 3 F_8 F_2^2 F_3 + 2 P F_6 F_2 F_3 - 6 P F_1 F_3^2$$

$$S_4 = 20 F_4 F_2^3 F_3^2 - 6 F_6 F_2^2 F_3^2 + 3 F_8 F_2 F_3^2 + P F_6 F_3^2$$

$$S_6 = 20 F_4 F_2^2 F_3^3 - 4 F_6 F_2 F_3^3 + F_8 F_3^3$$

$$S_8 = 10 F_4 F_2 F_3^4 - F_6 F_3^4$$

$$S_{10} = 2 F_4 F_3^5$$

$$T_1 = 8 F_4 F_2^4 F_3 - 4 P F_5 F_2 F_3$$

$$T_3 = 32 F_4 F_2^3 F_3^2 - 4 P F_5 F_3^2$$

$$T_5 = 48 F_4 F_2^2 F_3^3$$

$$T_7 = 32 F_4 F_2 F_3^4$$

$$T_9 = 8 F_4 F_3^5$$

$$W_0 = -2 k^2 F_4 F_2^5 + F_{10} F_2^4 + F_9 F_2^3 - P k^2 F_5 F_2^2 - 2 P F_2 F_3 F_5$$

$$W_2 = -10 k^2 F_4 F_2^4 F_3 + 4 F_{10} F_2^3 F_3 + 3 F_9 F_2^2 F_3 - 2 P k^2 F_5 F_2 F_3 + 6 P F_3^2 F_5$$

$$W_4 = -20 k^2 F_4 F_2^3 F_3^2 + 6 F_{10} F_2^2 F_3^2 + F_9 F_2 F_3^2 - P k^2 F_5 F_3^2$$

$$W_6 = -20 k^2 F_4 F_2^2 F_3^3 + 4 F_{10} F_2 F_3^3 + F_9 F_3^3$$

$$W_8 = -10 k^2 F_4 F_2 F_3^4 + F_{10} F_3^4$$

$$W_{10} = -2 k^2 F_4 F_3^5$$

We then assume a solution to (9-17) of the form

$$\tilde{\chi}_1 = \sum_{s=0}^{\infty} B_s y^s \quad (9-23)$$

Substitution of (9-23) into (9-17) yields the following lengthy recursion relation for coefficients B_s :

$$\begin{aligned} & s(s-1)(s-2)(s-3)Q_0 B_s + \\ & + \left[\begin{array}{l} (s-2)(s-3)(s-4)(s-5)Q_2 \\ + (s-2)(s-3)(s-4)R_1 \\ + (s-2)(s-3)S_0 \end{array} \right] B_{s-2} + \left[\begin{array}{l} (s-4)(s-5)(s-6)(s-7)Q_4 \\ + (s-4)(s-5)(s-6)R_3 \\ + (s-4)(s-5)S_2 \\ + (s-4)T_1 \\ + W_0 \end{array} \right] B_{s-4} \\ & + \left[\begin{array}{l} (s-6)(s-7)(s-8)(s-9)Q_6 \\ + (s-6)(s-7)S_4 \\ + (s-6)T_3 \\ + W_2 \end{array} \right] B_{s-6} + \left[\begin{array}{l} (s-8)(s-9)(s-10)(s-11)Q_8 \\ + (s-8)(s-9)S_6 \\ + (s-8)T_5 \\ + W_4 \end{array} \right] B_{s-8} \\ & + \left[\begin{array}{l} (s-10)(s-11)S_8 \\ + (s-10)T_7 \\ + W_6 \end{array} \right] B_{s-10} + \left[\begin{array}{l} (s-12)(s-13)S_{10} \\ + (s-12)T_9 \\ + W_8 \end{array} \right] B_{s-12} \\ & + W_{10} B_{s-14} = 0 \end{aligned}$$

Because this recursion relation contains only even subscripts, the solutions will be either symmetric or antisymmetric about the point $y = 0$. Which solution we choose depends upon the boundary conditions. In either case, we require $\tilde{\chi}_1, \tilde{\chi}_3, \tilde{\psi}_1, \tilde{\psi}_3$ all to vanish at the wall at $y = +1$. For (9-17) for $\tilde{\chi}_1$ alone, this requires

$$\tilde{\chi}_1(1), \frac{\partial^2 \tilde{\chi}_1(1)}{\partial y^2} = 0 \quad (9-25)$$

If we wish to put the other wall at $y = 0$, we require

$$\tilde{\chi}_1(0), \frac{\partial^2 \tilde{\chi}_1(0)}{\partial y^2} = 0 \quad (9-26)$$

Conditions (9-25) and (9-26) will result in the solution containing only odd powers of y , i.e. the antisymmetric solution.

We could instead require

$$\frac{\partial \tilde{\chi}_1(0)}{\partial y}, \frac{\partial^3 \tilde{\chi}_1(0)}{\partial y^3} = 0 \quad (9-27)$$

which would give us the symmetric solution. This boundary condition is equivalent to taking the symmetric solution when we require

$$\tilde{\chi}_1(-1), \frac{\partial^2 \tilde{\chi}_1(-1)}{\partial y^2} = 0 \quad (9-28)$$

giving us a channel between walls at $y = \pm 1$. The symmetric solution, from conditions (9-25) and (9-27) was the one calculated in the nonmagnetic case by Pedlosky (1964b). We shall consider both the symmetric and antisymmetric solutions. While the antisymmetric solution is probably physically

more relevant to the solar problem, as we shall discuss at the end of this chapter, the symmetric case is the more unstable, and can be found for larger horizontal shears, so that we have carried the symmetric case further.

The antisymmetric solution should in the absence of horizontal shear be simply the series for $\sin n\pi y$, $n=1,2,\dots$ in the range $0 \leq y \leq 1$

The symmetric solution with no horizontal shear will represent $\cos \frac{n\pi y}{2}$

$n=1,3,5,\dots$ in the range $-1 \leq y \leq +1$, or equivalently, $\sin \frac{n\pi y}{2}$

in the range $0 \leq y \leq +2$ (to conform with Chapter 8). We plotted the stability diagrams showing the marginally stable curves, growth rates and phase velocities for the lowest modes of both of these cases in Figs. 5 and 6. We expect, then, that the stability diagrams for the lowest mode when horizontal shear is present, will evolve continuously from Figs. 5 and 6 as the shear parameter b is increased. The higher modes can also be calculated, although they are of less interest, because they will, in general, grow more slowly, and because the walls are artificial. In the actual computations, some care must be exercised to keep the modes separate.

We can see in Figs. 5 and 6 that only the short wavelength edge of the unstable region is discernible for these narrow channels. This will remain the case when shear is added. Therefore there is only one marginally stable curve to be considered. Examination of this curve in Fig. 6 shows that for $l=2$, and $l=1$, everywhere along it $C_r > \sqrt{P}$. This result allows us to define more closely the bounds on C_r required to avoid singularities. For if $C_r > \sqrt{P}$ when $b=0$, it must remain so when $b > 0$ -- otherwise at some value of b we would have to have

$C_r = \sqrt{P}$, in violation of the restriction (9-15). Therefore, from this deduction and (9-14), the phase velocities of marginally stable waves in the absence of singularities must be bounded according to

$$\sqrt{P} < C_r < U_0(1-b) - \sqrt{P} \quad (9-29)$$

One can see that these bounds can be highly restricting. For example, if $b = 0.5$, the largest value of P/U_0^2 for which a marginally stable wave satisfying (9-29) can exist is $P/U_0^2 = 1/16$ (for which we must have $C_r/U_0 = 0.25$). In the case without shear, the highest value would be $P/U_0^2 = 1/4$, so that for $b = 0.5$, at least three-fourths of the range of P/U_0^2 would contain singularities. Despite this great reduction in range, we can still obtain interesting results within what is left.

Our first job, then, is to try to find marginally stable curves as functions of P/U_0^2 and λ^2 for various values of b whose phase velocities fall within the range defined by (9-29). The manner in which this is done is essentially the same as that used by Pedlosky (1964b).

In the antisymmetric case, one solution is obtained from (9-24) by setting

$$B_1 = 1 \quad , \quad B_3 = 0 \quad ; \quad \text{the other results from taking } B_1 = 0 \quad ; \\ B_3 = 1 \quad . \quad \text{Let us denote these two solutions, respectively, by}$$

$$D_1 = \sum_{s=0}^{\infty} B_{2s+1}^{(1)} y^{2s+1} \quad (9-30)$$

$$D_3 = \sum_{s=0}^{\infty} B_{2s+1}^{(3)} y^{2s+1} \quad (9-31)$$

Similarly, the symmetric solutions are obtained by setting $B_0 = 1$, $B_2 = 0$; $B_0 = 0$, $B_2 = 1$. Analogously, we denote them by

$$D_0 = \sum_{s=0}^{\infty} B_{2s}^{(0)} y^{2s} \quad (9-32)$$

$$D_2 = \sum_{s=0}^{\infty} B_{2s}^{(2)} y^{2s} \quad (9-33)$$

Then satisfaction of the boundary condition (9-25) requires

$$F \equiv \begin{vmatrix} D_{0,1}(1) & D_{2,3}(1) \\ \frac{\partial^2 D_{0,1}(1)}{\partial y^2} & \frac{\partial^2 D_{2,3}(1)}{\partial y^2} \end{vmatrix} = 0 \quad (9-34)$$

Thus $F(m, \beta, U_0, b, P, \lambda^2; c) = 0$ is the eigenvalue equation for the problem, with the complex phase velocity c the eigenvalue.

Stability criteria, phase velocities, growth rates

Just as in the nonmagnetic case, the marginally stable curves will be characterized by the coalescence of two roots of (9-34). For one of these roots, as one enters the unstable region from the marginal curve, the imaginary part will be positive, denoting an exponentially growing mode. The other will give its complex conjugate, an exponentially decaying mode. To find the neutral curves, we first fix values for some of the parameters, setting $m = 1$, $\beta = 0.707$, the same values we used in the no shear case. We then set $C_i = 0$, and for different shears b , vary C_r/U_0 in $P/U_0^2, \lambda^2$ space to look for zeros of F . The locus of points where

two such roots coalesce for a given $P/U_0^2, \lambda^2$ defines the curve of marginal stability. In doing this, we are guided by some of Pedlosky's (1964b) results. For example, as we have stated, he found the region of instability shifted toward shorter wavelengths when horizontal shear was included. Analogously, we should expect our unstable regions do to the same, so that we should look to the left of the marginally stable curves in Figs. 5 and 6. Pedlosky also found the phase velocity C_r to be less positive when shear was added, due to the increased " β effect" of the positive vorticity in the mean flow. We should expect the same in the magnetic case. Using these guidelines, and the restrictions imposed by (9-29), we have evaluated F numerically on the IBM 7094 at the MIT Computation Center and have found new marginally stable curves for both the symmetric and antisymmetric cases, for several shears, as plotted in Fig. 8.

We can see from Fig. 8 that for both the symmetric and antisymmetric cases the region of instability is shifted toward shorter wavelengths. In the symmetric case, for a given shear, the percentage reduction in wavelength is virtually independent of the magnetic field strength. In the antisymmetric case, this reduction is strongly dependent on the field strength in the narrow range of parameters for which it could be calculated, decreasing for increasing field strength.

For each shear, there was a short wavelength limit beyond which the curve of marginal stability could not be computed by using power series, because the singularity defined by (9-11) was within the channel. The light solid lines in Fig. 8 denote the locus of the limit of the marginally stable

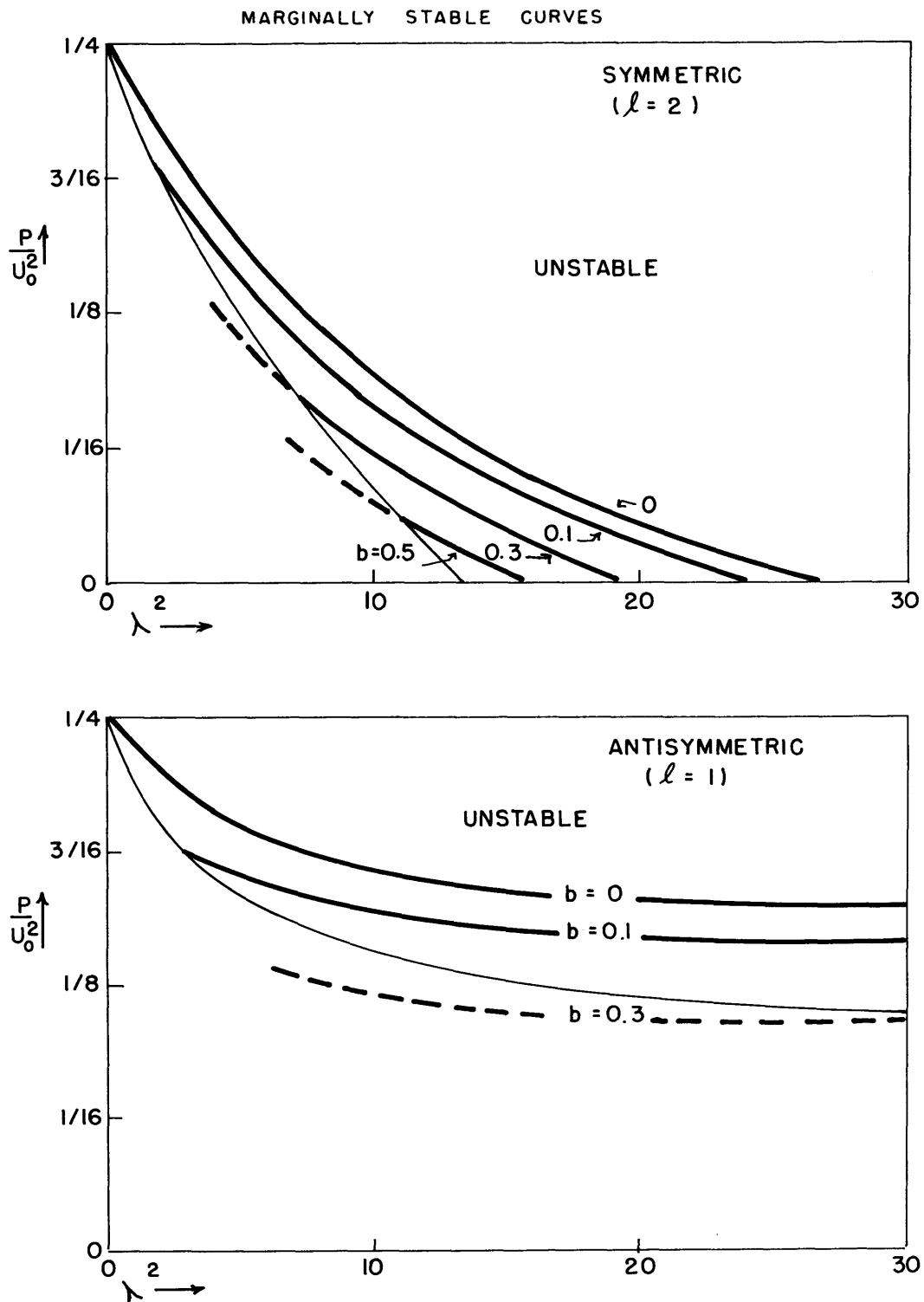


Fig. 8. Marginally stable curves (heavy solid lines) for the parabolic flow profile $U_1 = 0$; $U_3 = U_0(1 - by^2)$, for several values of b , for disturbances symmetric (upper figure) and antisymmetric (lower figure) in the range $-1 \leq y \leq +1$. Light solid lines denote short wave limit as a function of the shear parameter b to the left of which singularities of the equations will be present in the range of y . Heavy dashed lines represent computer extension of the marginally stable curves beyond this limit which exist because power series is truncated.

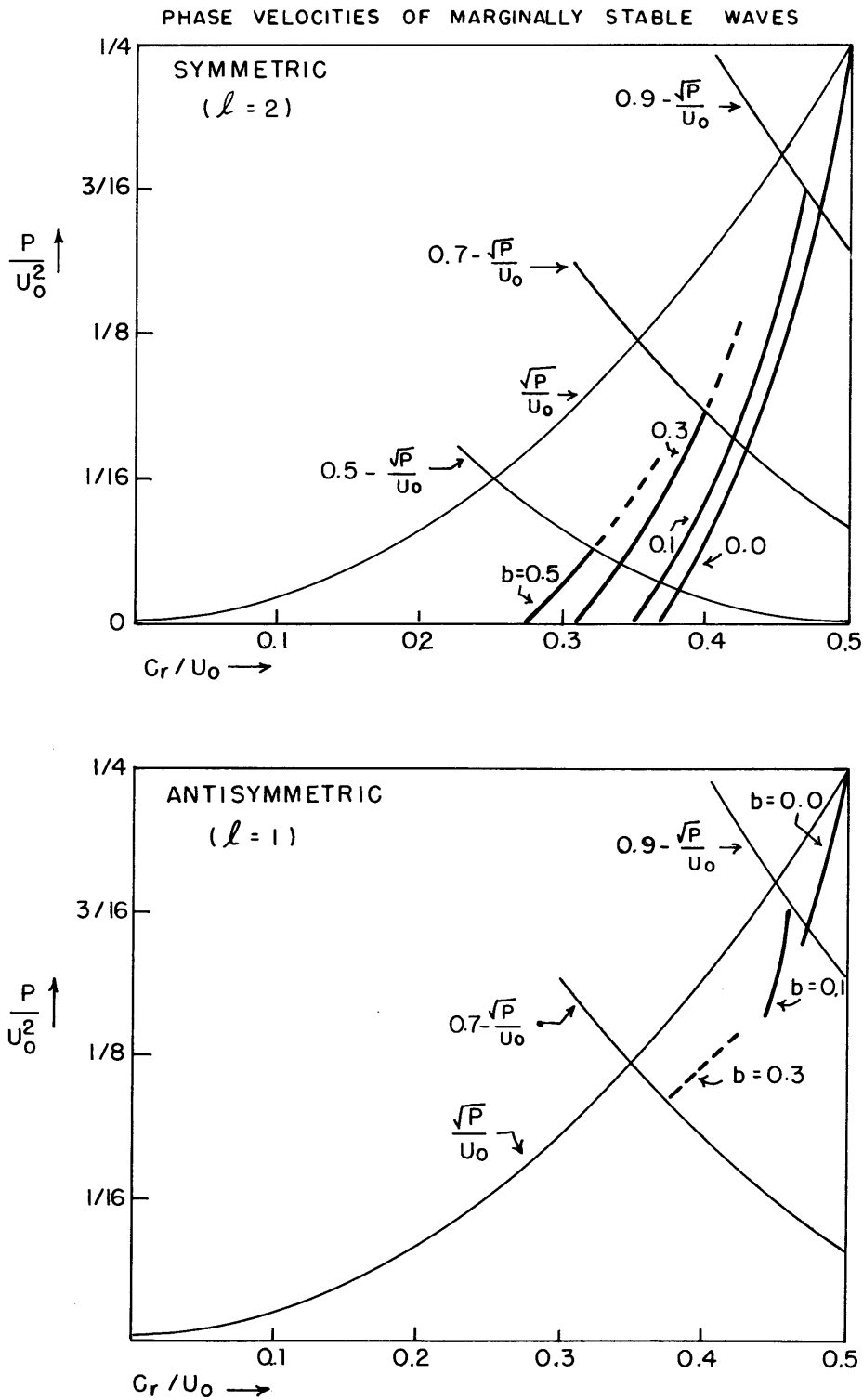


Fig. 9. Phase velocities C_r/U_0 of marginally stable waves (heavy solid lines) for various shears in parabolic profile (same as Fig. 8). Light solid lines denote singularity boundaries as defined by limits in (9-29). Those light solid lines sloping upward to the left are defined by $C_r/U_0 = (1-b) - \sqrt{P}/U_0$; the one sloping upward to the right by $C_r/U_0 = \sqrt{P}/U_0$. For a given shear b , points above this curve give singularities in the domain $-1 \leq \eta \leq +1$. Heavy dashed lines denote extensions of phase velocities of marginally stable waves into singular region, found for truncated power series.

curves for different shears. On the short wavelength side of this limiting curve, we do not know whether the series solution converges. However, since the machine must always work with a truncated series, it may give roots on the short wavelength side. The heavy dashed lines are extensions of the curves of marginal stability found from such calculations. It is not clear whether they represent convergent solutions.

We can see more clearly the approach of the curves of marginal stability to the singularities by plotting the phase velocity C_r against P/U_0^2 , as we have done in Fig. 9. From Fig. 9, we see that indeed just as in the nonmagnetic case, the phase velocities at the edge of the unstable regions decrease with increasing horizontal shear. For a given shear, they also increase with increasing magnetic field, just as they do when no shear is present.

The light solid lines in Fig. 9 represent the limits on C_r to avoid singularities, set by (9-29). Thus the marginal curves for a given shear must remain below the tent shaped limiting curve for that shear. One can see, then, that the singularity encountered is that defined by (9-11) in all cases. The heavy dashed lines again show the extensions of the marginal curves with truncated series beyond the boundary of the singular points.

One can also see from Figs. 8 and 9, as we have already pointed out, that for increasing shear, the range of P/U_0^2 over which we can find solutions by the power series technique can be greatly reduced. This is most severe in the antisymmetric case.

In addition to calculating the marginal curves for different horizontal shears, we have also computed the growth rates and phase velocities of unstable waves for several cases. To do this, we evaluate F for a grid of values of C_r and C_i . The root values C_r, C_i occur where the real and imaginary parts of F go to zero simultaneously. These results are presented in Fig. 10. In Fig. 10a, b, we present the phase velocities and growth rates of symmetric and antisymmetric disturbances for equal $P/U_0^2 (= 5/32)$ and $b (= 0.1)$. The symmetric disturbances are seen to be much more unstable, and unstable for shorter wavelengths. The phase velocities for the symmetric case also drop off much faster with increasing wavelength. These results are qualitatively equivalent to those for $b=0$, from Figs. 5 and 6, as we would expect, since we are not able to put in very much shear. We would expect the antisymmetric mode to be the less unstable whether or not there is horizontal shear, because the cross-channel scale is smaller than for the symmetric case.

In Figs. 10c, d, we show the effect of increasing the magnetic field strength for a given shear ($b = 0.3$) in the symmetric case. We can see that, just as with no shear, increasing the magnetic field increases the phase velocity, and shifts the most unstable wavelength toward shorter waves. The most unstable wave is more sharply defined, but the magnitude of the growth rate remains nearly the same (increasing slightly), the decrease in C_i being slightly overbalanced by the increase in k .

Finally, in Fig. 10e, f, we compare the symmetric no-shear case with a shear case ($b = 0.3$) for equal magnetic field strengths. Here, just as in the nonmagnetic problem, increasing the shear decreased the phase velocity and made the most unstable wave shorter, but about as unstable as without shear.

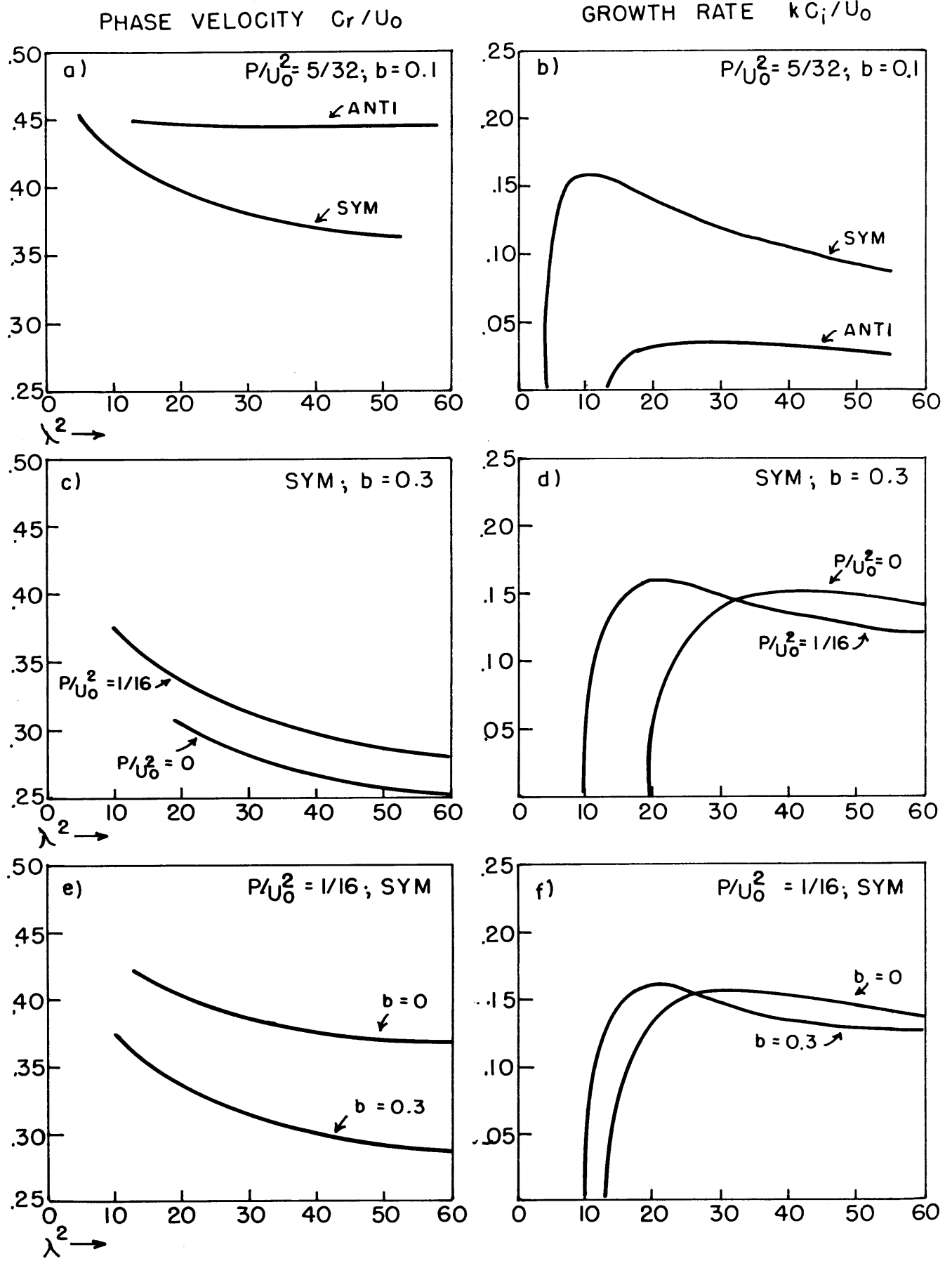


Fig. 10. Phase velocities and growth rates of unstable waves for several cases.

We can see that the effects on the growth rate of increasing the magnetic field and increasing the horizontal shear are about the same. Their effects on the phase velocities, however, are opposite. The similarity of effects on the growth rates is reasonable, since both increasing the field strength and increasing the horizontal shear effectively decrease the baroclinicity of the flow relative to the constraint of the magnetic field.

Structure of disturbances; stresses, energy conversions, and changes in initial state

It is interesting also to look at the structure of the unstable disturbance for which we have just calculated the curves of marginal stability, phase velocities, and growth rates. Since we have the two solutions D_0 , D_2 , or D_1 , D_3 , which together satisfy respectively the symmetric or antisymmetric boundary conditions, we may write the whole solution for $\tilde{\chi}_1$ in the form

$$\tilde{\chi}_1 = D_{0,1} + \zeta_1 D_{2,3} \tag{9-34}$$

where

$$\zeta_1 = - \frac{D_{0,1}(1)}{D_{2,3}(1)} \tag{9-35}$$

Once we have the complete series for $\tilde{\chi}_1$, it is a simple matter to calculate term by term the series for $\tilde{\chi}_3$, $\tilde{\psi}_1$, $\tilde{\psi}_3$, \tilde{w}_2 and \tilde{u}_2 , to give us the complete structure of the disturbance. In Figs. 11 and 12 we have presented the results for the symmetric solution for the point $P/U_0^2 = 1/16$,

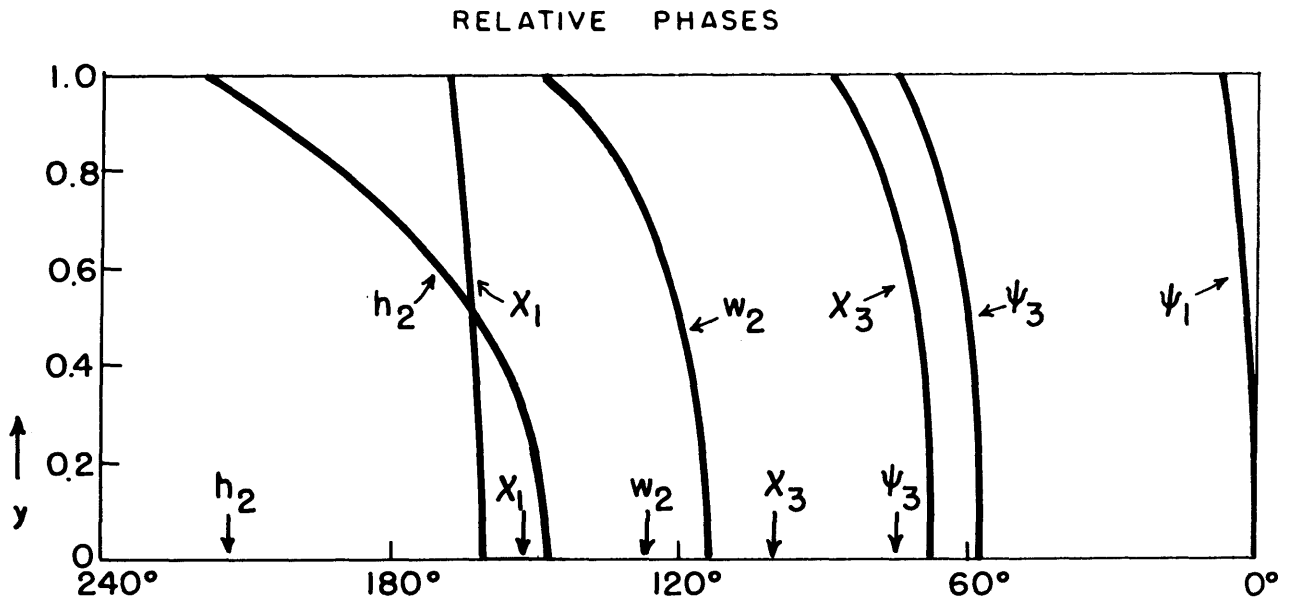


Fig. 11. Phases of dependent variables as a function of y for the case $P/U_0^2 = 1/16$, $b = 0.3$, $G = 0.177$ ($\beta = 0.707$), $m = 1$ ($e \sin^2 \phi_0 = 0.5$), $\lambda^2 = 20$ for which $C_r/U_0 = 0.3376$, $C_i/U_0 = 0.1124$. Labeled arrows on abscissa indicate phases of most unstable wave for same P/U_0^2 , β , m , but with $b = 0$ (for which $\lambda^2 \approx 21$) taken from Fig. 7. Phases plotted increasing to left so that phase curves give crests of a wave advancing to the right (in the same sense as the sun's rotation).

$\beta = 0.707$, $m = 1$, $b = 0.3$, $\lambda^2 = 20$. At this point, $C_r/U_0 = 0.3376$, $C_i/U_0 = 0.1124$ come closest to satisfying the boundary conditions. The series were calculated to the 48th power in y , and converged at a satisfactory rate.

Figure 11 gives the phases of the disturbances as a function of y . For convenience of comparison with the no-shear case, we have plotted the phases relative to $\psi_1(0)$. They are defined in the same way as the no-shear case, i.e. by equations (8-46). We have furthermore plotted along the axis the phases of a disturbance with the same magnetic field strength, but with no horizontal shear in the flow, and no β effect (from Fig. 7). The phases are plotted increasing toward the left (the opposite of Fig. 7) so that the constant phase lines represent the wave crests of a wave advancing toward the right (in the same sense as the sun's rotation).

We can see from Fig. 11 that, with the exception of \tilde{h}_z , the relative phases, and particularly the phase differences, do not change a great deal when the β effect and horizontal shear are added. Thus we would expect all the energy conversions dependent upon these phase differences, i.e. $(\langle A \rangle, A')$, (A', K') , (K', M') to have the same signs as in the no-shear case, as indeed we shall see that they do. Also, the changes in the initial state should proceed in an at least qualitatively similar manner to the no shear case. In addition, the increase in the phases with y implies the existence of Reynolds and Maxwell stresses, and consequently non-zero values for the energy conversions $(K', \langle K \rangle)$, $(\langle K \rangle, M')$. The amount of this "tilt" in the crests of the variables with y is about 3 times as large in ψ_3 and χ_3 as in ψ_1 and χ_1 . This is understandable since the horizontal

shear is confined to the upper layer. The phase shift with y is still more pronounced in \tilde{w}_z , and is most pronounced in \tilde{h}_z , being around 70° , or 1/5 of a wavelength.

The amplitudes of the various variables are plotted in Fig. 12. We can see that the horizontal shear does not drastically change the amplitude as a function of y over the no-shear case. That is, the curve for $|\tilde{\psi}_3(0)| \cos \frac{\pi y}{2l}$, the amplitude function in the no-shear case, fits quite closely to $\tilde{\psi}_3$ in the shear case. $|\tilde{\chi}_3(0)| \cos \frac{\pi y}{2}$ departs somewhat more strongly from $\tilde{\chi}_3$, but still not radically.

The poleward eddy flux of heat still renders $(\langle A \rangle, A')$ positive, while a vertical eddy flux of heat makes (A', K') positive. These fluxes are respectively represented in the form

$$\langle (\psi_3' - \psi_1') \frac{\partial}{\partial x} (\psi_3 + \psi_1) \rangle = \frac{k}{2} \left[(\tilde{\psi}_{3i} - \tilde{\psi}_{1i}) (\tilde{\psi}_{3r} + \tilde{\psi}_{1r}) - (\tilde{\psi}_{3r} - \tilde{\psi}_{1r}) (\tilde{\psi}_{3i} + \tilde{\psi}_{1i}) \right] \quad (9-36)$$

$$\langle 2 w_z' (\psi_3' - \psi_1') \rangle = 2 \tilde{w}_{2r} (\tilde{\psi}_{3r} - \tilde{\psi}_{1r}) + 2 \tilde{w}_i (\tilde{\psi}_{3i} - \tilde{\psi}_{1i}) \quad (9-37)$$

where, as usual, the subscripts r and i denote the real and imaginary parts. Both (9-36) and (9-37) are positive everywhere in the channel (indicating poleward and upward heat transports), as we can see from Fig. 13. Both transports are a maximum at $y = 0$. In the antisymmetric case, both would be zero there.

The conversion (K', M') is still positive, indicating the perturbation magnetic fields are growing, because ψ and χ are not exactly in phase or exactly 180° out of phase with each other in either layer.

RELATIVE AMPLITUDES

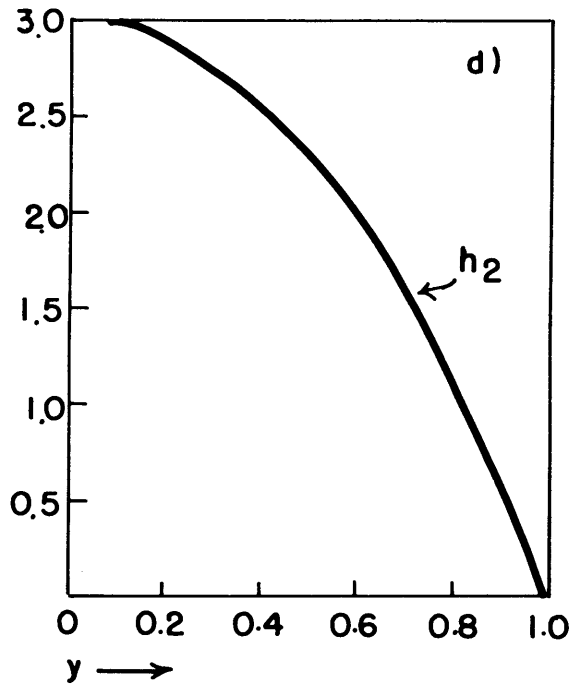
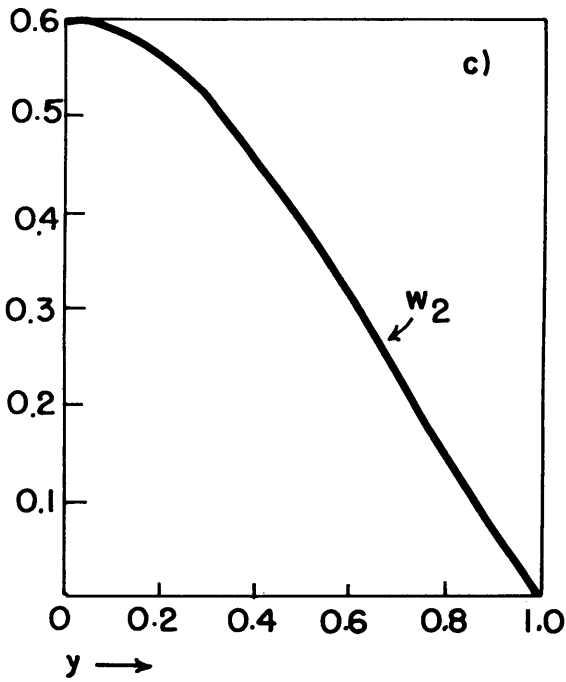
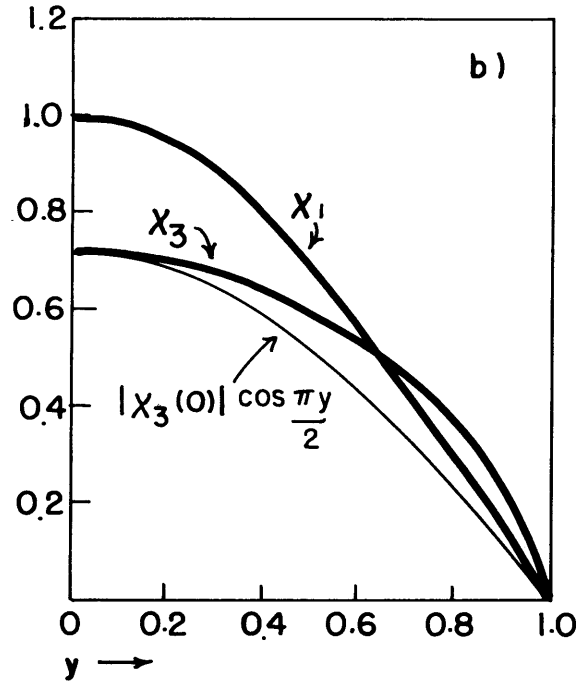
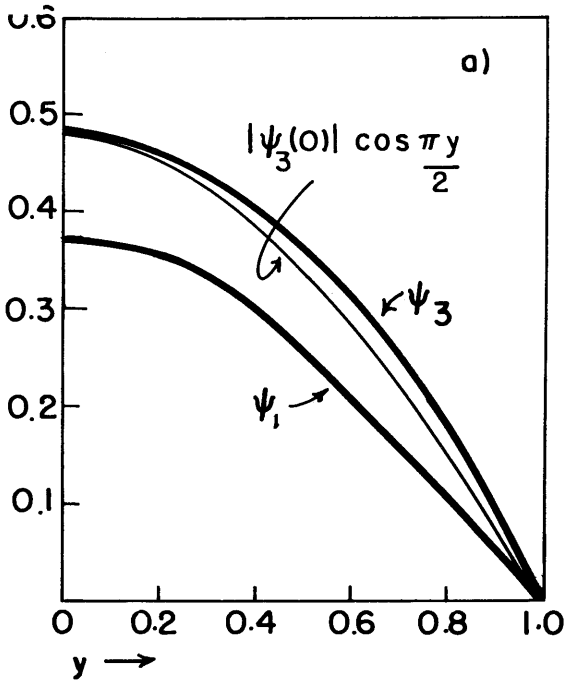


Fig. 12. Amplitudes of various variables (heavy solid lines) as function of y for same case as in Fig. 11, measured relative to $\tilde{\chi}_1(0)$. For comparison, amplitude functions for $b=0$ are plotted with equal amplitude at $y=0$ as case with shear.

EDDY TRANSPORTS OF HEAT

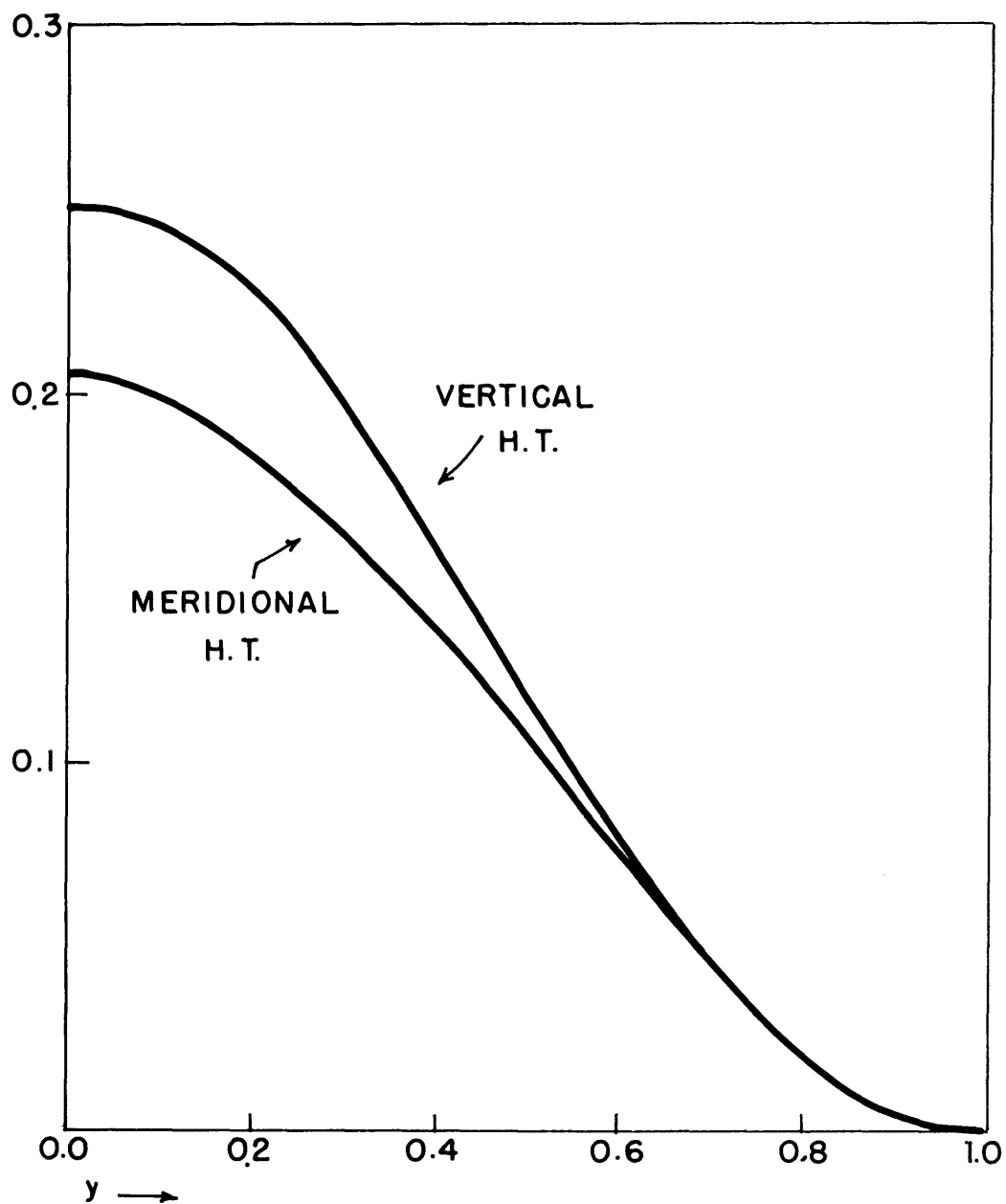


Fig. 13. Meridional and vertical heat transports by disturbances for same case as Fig. 11. Amplitude relative to $\tilde{\chi}_1(0)$.

The Reynolds stresses for the upper and lower layers may be evaluated according to

$$-\left\langle \frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial y} \right\rangle = \frac{k}{2} \left(\tilde{\psi}_i \frac{\partial \tilde{\psi}_r}{\partial y} - \tilde{\psi}_r \frac{\partial \tilde{\psi}_i}{\partial y} \right) \quad (9-38)$$

by substituting in, respectively, $\tilde{\psi}_3$ and $\tilde{\psi}_1$. Similarly, the Maxwell stresses for each of the two layers may be found from

$$P_s \left\langle \frac{\partial \chi}{\partial x} \frac{\partial \chi}{\partial y} \right\rangle = -\frac{k P_s}{2} \left(\tilde{\chi}_i \frac{\partial \tilde{\chi}_r}{\partial y} - \tilde{\chi}_r \frac{\partial \tilde{\chi}_i}{\partial y} \right) \quad (9-39)$$

These stresses are plotted in Fig. 14. With the coordinate system used, a positive value for the stress indicates momentum transport toward higher values of y (higher latitude). The Reynolds stresses are thus in each layer transporting momentum toward $y=0$, i.e. toward the latitude of maximum zonal velocity. Since the Reynolds stress is antisymmetric about $y=0$, the transport in the range $-1 \leq y \leq 0$ will also be toward the maximum of the zonal flow. This upgradient transport of momentum would thus tend to increase the horizontal shear in the flow. In terms of energy, eddy kinetic energy would therefore be converted into mean zonal kinetic energy. This conversion $(K', \langle K \rangle)$, as we can see from its definition (7-19), would be taking place initially only in the upper layer (level 3) since that is the only layer with shear. This energy conversion is the same as what Pedlosky (1964b) showed to be happening in the non-magnetic case. But in the magnetic case, we can see from Fig.14 that there is an additional conversion process which opposes this. That is, the Maxwell stresses are acting to transport momentum away from the maximum of zonal

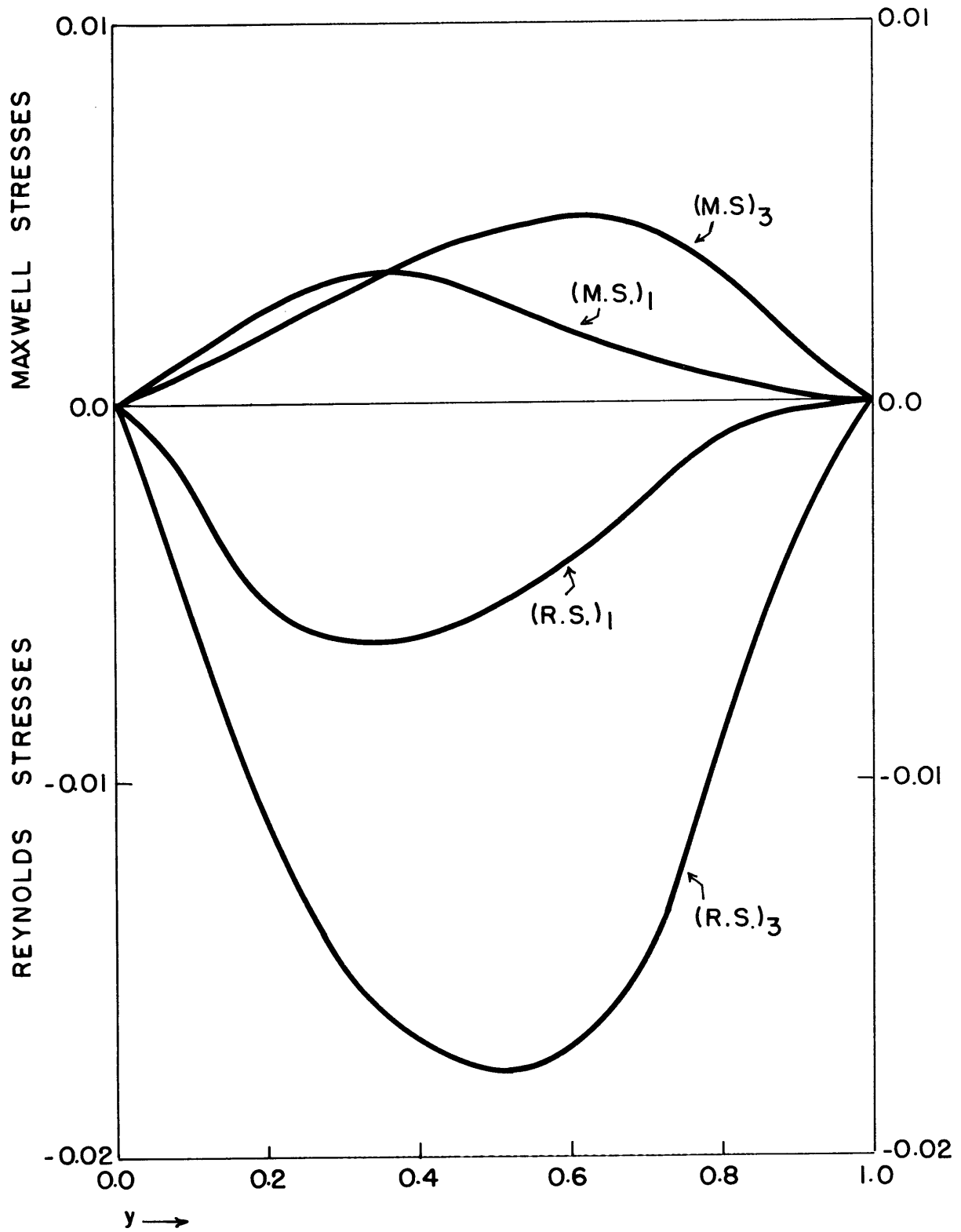


Fig. 14. Reynolds and Maxwell stresses for same case as Fig. 11.
 Amplitudes relative to $\tilde{\chi}_1(0)$.

flow, though at a smaller rate than the Reynolds stresses are adding it. As the Maxwell stresses are also antisymmetric, this is true in the range $-1 \leq y \leq 0$ as well. Thus the energy conversion $(\langle K \rangle, M')$ defined by (7-20), is positive. This means that the perturbation magnetic fields are growing not only directly at the expense of eddy kinetic energy (i.e. from (K', M')) but also indirectly through conversion from mean zonal kinetic energy, which in turn, as we just showed, is supplied from eddy kinetic energy by the Reynolds stresses.

For this particular case, the magnitude of the Maxwell stress transports are about one-third as large as the Reynolds stress fluxes. Although we have not yet had time to look extensively, there may be other values of the parameters for which the Maxwell stresses are still larger relative to the Reynolds stresses.

Let us finally turn to the changes in the initial state produced by these disturbances. In principle one can generalize the analysis of Chapter 8 where $\partial/\partial t \langle \psi_{1,3} \rangle$ and $\partial \langle w_z \rangle / \partial t$ were found, but higher derivatives of the Reynolds and Maxwell stresses and the meridional heat flux would have to be taken, which would be computationally difficult. Rather than try to do this, it seems reasonable to argue as Pedlosky (1964b) did that the resulting mean meridional circulation and change in zonal flow would not be markedly different from the no-shear case (which we showed in Chapter 8 was very similar to the nonmagnetic case, i.e. Phillips, 1954).

The changes in the magnetic structure are much easier to compute. The original toroidal field will be changed by the negative gradient of

what we have called the mixed stress (see (7-12), (7-13)). This mixed stress may be evaluated for each level in the form

$$\left\langle \frac{\partial \chi'}{\partial x} \frac{\partial \psi'}{\partial y} - \frac{\partial \chi'}{\partial y} \frac{\partial \psi'}{\partial x} \right\rangle = \frac{k}{2} \left[\tilde{\psi}_i \frac{\partial \tilde{\chi}_r}{\partial y} + \tilde{\chi}_r \frac{\partial \tilde{\psi}_i}{\partial y} - \tilde{\psi}_r \frac{\partial \tilde{\chi}_i}{\partial y} - \tilde{\chi}_i \frac{\partial \tilde{\psi}_r}{\partial y} \right] \quad (9-40)$$

The results are presented graphically in Fig. 15a. It is interesting to see that the mixed stress is smaller in the upper layer than in the lower, in contrast to the Reynolds and Maxwell stresses.

The tendency equations (7-12), (7-13) for the toroidal field can be written in short form as

$$\frac{\partial}{\partial t} H_{1,3} = -\frac{\partial}{\partial y} (\text{Mixed } S)_{1,3} \quad (9-41)$$

From (9-41) and Fig. 15a, then, we can see that in each layer, the amplitude of H will be decreased at low values of y (low latitudes) and increased at high values of y (high latitudes). If one pictures the complete channel $-1 \leq y \leq +1$, the magnitude of H is therefore decreasing in the middle of the channel and increasing near the edge, because the mixed stress is antisymmetric about $y=0$. This is the same as for the no-shear case. The primary difference from the no-shear case is that in the upper layer, the positive tendency is confined to a narrower band next to the walls.

Finally, we may calculate the production of a mean poloidal field from (7-14). There are two effects which contribute to the tendency

$\partial \langle h_z \rangle / \partial t$. Both have been plotted in Fig. 15b. We can see that

EDDY TRANSPORTS OF MAGNETIC FIELDS

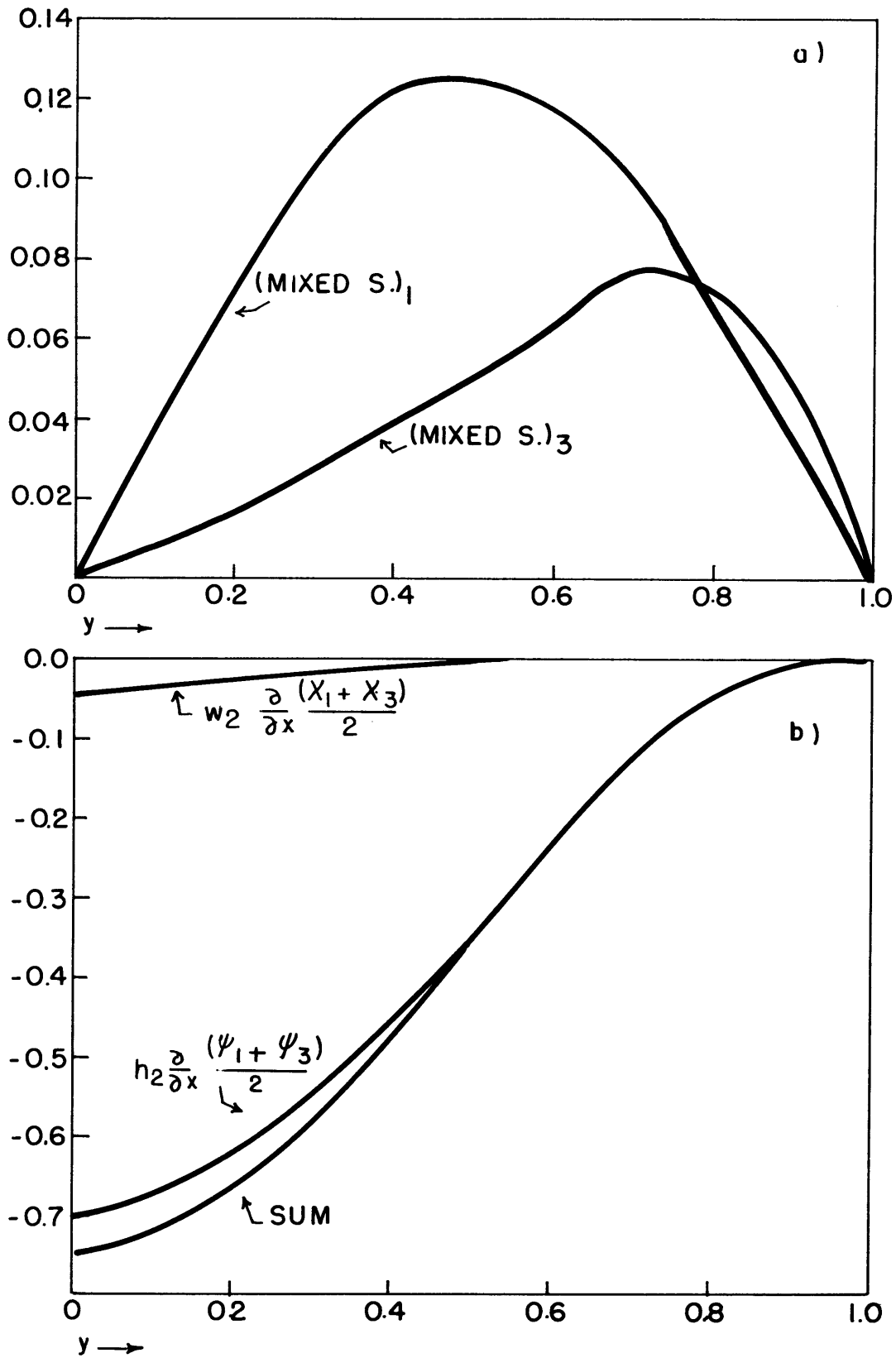


Fig. 15. a) Mixed stresses for same case as in Fig. 11. Amplitudes relative to $\bar{Z}_1(0)$.
 b) Eddy meridional transport of eddy vertical magnetic fields (middle curve); twisting of horizontal fields into vertical by vertical motion (upper curve); and their sum. These processes together determine the production of the mean poloidal field (see equations (7-14) and (9-42)).

by far the major contributor to this production is the eddy meridional transport of eddy vertical magnetic fields. Thus in this case the time tendency is approximately

$$\begin{aligned} \frac{\partial \langle h_z \rangle}{\partial t} &\approx \frac{1}{2} \frac{\partial}{\partial y} \left[-h'_z \frac{\partial}{\partial x} (\psi'_1 + \psi'_3) \right] \\ &= \frac{k}{4} \frac{\partial}{\partial y} \left[h_{2r} (\psi_{3i} + \psi_{1i}) - h_{2i} (\psi_{3r} + \psi_{1r}) \right] \end{aligned} \quad (9-42)$$

From (9-42) and Fig. 15b, then, we can see that in the range $0 \leq y \leq 1$ the poloidal field tendency is everywhere positive. That is, the disturbances are producing an upward directed poloidal field. It is also evident that the tendency is antisymmetric about $y=0$, so that a downward directed poloidal field would be produced in the range $-1 \leq y \leq 0$. This is again just what we found in the no-shear case. For a channel of nondimensional width 2, this field was proportional to $\sin \pi y$ for a range $-1 \leq y \leq +1$. From Fig. 15b, it is obvious that the form of $\partial \langle h_z \rangle / \partial t$ does not change much when horizontal shear is added.

Application to the sun

To apply our results to the sun, it would probably have been more realistic to have examined in detail the antisymmetric rather than the symmetric solution. That is, the antisymmetric solution would be for a channel of nondimensional width $l=1$, for which the maximum of zonal flow would occur at the equatorward boundary of the channel. This would simulate in a crude way the actual mean zonal flow profile for the sun deduced from the various spot displacement statistics. (The narrower

channel, would, on the other hand, be less realistic). The difficulty was that with the antisymmetric solution, we were not able to put in very much horizontal shear ($b \approx 0.1$) before the singularities would enter the domain. Therefore we chose to examine the symmetric solution instead, to demonstrate the effect of a more reasonable amount of horizontal shear (even in this case, we could not take b much greater than 0.3).

The symmetric case corresponds to a flow profile in which the maximum zonal flow is in the middle of the channel. To fit this configuration to the sun, we would have to assume the channel to straddle the equator. This would violate some of our scaling assumptions, for example that we could take the sine of the latitude as a constant to lowest order when solving for the stream function. In addition, the solution applied in this way would give the maximum of meridional heat transport at the equator, as if one solar hemisphere were colder than the other. But there are no grounds, observational or theoretical, for expecting such an asymmetric temperature distribution to exist on the sun.

The symmetric solution would correspond more realistically to the situation when the maximum of zonal flow was in middle latitudes, as it is in the earth's atmosphere. But even so, we can reasonably expect that the Reynolds, Maxwell, and mixed stresses, which are antisymmetric about $y=0$ for both the symmetric and antisymmetric solutions, will be qualitatively similar in the two cases. The production of mean poloidal fields would be different for the antisymmetric solution, as the eddy meridional flux of vertical fields would vanish at $y=0$, rather than be a maximum there, as it is in the symmetric case. This would just

result in a single celled poloidal field being produced between $y = 0$ and $y = +1$ (i.e. one complete cell within each hemisphere) instead of between $y = -1$ and $y = +1$ (one cell for the two hemispheres).

Given these reservations, the solutions we have found basically include all of the desirable structural properties contained in the solutions of Chapter 8 for the case of no horizontal shear. But in addition, they possess additional important properties, properties which meet the characteristics of large scale hydromagnetic disturbances that Starr and Gilman (1965b) (c.f. Chapter 3) sought to explain the differential rotation. That is, these disturbances produce horizontal Reynolds stresses which transport momentum up the gradient, but which are partially offset by horizontal Maxwell stresses transporting momentum down the gradient. Starr and Gilman then invoked unpublished magnetograms of Howard (see also Chapter 1 and Appendix) as evidence that large scale disturbances with the tilted structure required to give these momentum fluxes might be present on the sun. Their argument was that the tilted pattern seen in the essentially vertical magnetic fields represented by the magnetograms suggested that a similarly tilted structure would be found in the horizontal streamlines and magnetic flux lines, if they could be measured. In the structure of the disturbances we have found, as we have already seen in Fig. 11, the streamlines and magnetic flux lines are indeed tilted upstream away from the maximum of zonal flow in the same way as are the vertical fields that the disturbances produce, though the tilt in the vertical fields is more pronounced. Thus these baroclinically unstable

hydromagnetic disturbances appear to be quite capable of producing the desired qualitative correspondence with the observations and of maintaining the differential rotation in the manner put forth by Starr and Gilman (1965b). Whether the disturbances on the sun really are of this type is, of course, another question.

10. Criticisms and possibilities for further work

There are obviously many grounds on which this work can be criticised. Hopefully many of these criticisms may be met by improvements and generalizations of the models and results we have put forth in the previous chapters.

One obvious fault is the use of a two-level model. This has yielded us many interesting results, but we need to know how well these carry over to the physically more realistic continuous case. To get very far at all, we would probably have to limit ourselves to the purely baroclinic problem, i.e. to zonal flows with no horizontal shear. Suppose, then, that the mean zonal flow $U = U(z)$, and that the magnetic field parameter $P = P(z)$ (The variation in P with z could come about by either a variation in the magnetic field strength, or the density, or both). Then it is fairly easy to show that for wavelike disturbances of the form

$$\psi = \tilde{\Psi}(z) \sin \frac{\pi y}{l} e^{ik(x-ct)} \quad (10-1)$$

between walls at $y=0, l$ the equation for the amplitude function $\tilde{\Psi}(z)$ is given by

$$\frac{\partial^2 \tilde{\Psi}(z)}{\partial z^2} + \frac{1}{\epsilon \sin^2 \phi_0} \left(-\alpha^2 + \frac{\beta}{U-c} + \frac{P \alpha^2}{(U-c)^2} \right) \tilde{\Psi}(z) = 0 \quad (10-2)$$

where $\alpha^2 = k^2 + \pi^2/l^2$. As a particular example, if P is constant and U is linear in z (direct hydromagnetic generalization of the Eady case) equation (10-2) is a confluent hypergeometric equation. If $\beta = 0$,

it reduces to a Bessels' equation. However, the order of the Bessel functions will be dependent upon the magnetic field strength, becoming imaginary for large enough field strength (or small enough wavelength). Such functions have not been studied very much, but they could be evaluated on the computer from series representations. The job would again be to solve for C for appropriate boundary conditions to find the regions of instability.

As another possibility, we could look more at the continuous barotropic problem, to see, for example, how the criterion for barotropic instability might be altered for certain flows by the presence of the magnetic field.

In addition to looking for unstable modes in the continuous case, it would be interesting also to pursue in greater detail the neutral modes that are possible, both in the continuous and two level cases. In general, these would be some combination of internal Rossby and Alfvén waves. As we have already pointed out, some work has been done along these lines by Hide (1964).

Finally, as another extension of the perturbation approach, it would seem worthwhile to try to find out, for the two level case, the significance of the singularities introduced by the magnetic field, in particular, the singularity defined by (9-11), which we encountered in the numerical calculations of Chapter 9. As one speculation, could it be that flows with this singularity within their domain are barotropically unstable, even if the potential vorticity has no extremum within the domain? As another possibility, could the Maxwell stresses over-balance the Reynolds stresses in

some flows of this type, thus depleting both available potential and available kinetic energy of the zonal flow? Unfortunately, due to the complexity of the equations, it may be quite difficult to pursue analytically the effect of these singularities very far or in any very general way.

As we discussed at the end of Chapter 9, the antisymmetric solution for the interval between $y=0$ and $y=1$ would in several respects be more applicable than the symmetric case to the solar problem. However, because of the presence of the singularities, we could not calculate this solution for a case which had very much horizontal shear (we chose $b = 0.1$) To get around this difficulty, it would seem worthwhile to solve numerically on a horizontal grid the initial value problem for the same set of scaled equations (both the linear and nonlinear cases could be done), instead of assuming a wavelike or exponentially growing disturbance at the outset. This could be done for a much wider channel than was feasible for the power series approach, and could be done for a number of initial flow profiles, including the one which probably most closely simulates the solar differential rotation, namely the half parabola with its maximum flow velocity at $y=0$, i.e. (9-1). In this way, we could study in more detail the structure of a growing disturbance for a wider range of initial conditions. Finite amplitude nonlinear effects could also be examined.

Several variations of this approach could also be made. For example, we could restrict the disturbances to a single or restricted number of zonal wavelengths, retaining grid points in the meridional (cross channel) coordinate. If we also restricted the wave numbers in the meridional

coordinate, we could examine in a simple way the nonlinear interactions between waves of different scales, analogous to the nonmagnetic work of Lorenz (1960a, b, 1962, 1963).

Another unsatisfactory feature of the present system of equations, one which we have not discussed before, is the difficulty of applying them near to the equator where the Coriolis parameter $2\Omega\sin\phi$ approaches zero. Because we have expanded in a Rossby number defined by $Ro = U/2\Omega L$ rather than by $Ro = U/2\Omega L\sin\phi_0$, we have left the problem of the vanishing of $\sin\phi$ in the heliostrophic balance itself, i.e. equation (4-49), which we repeat below

$$\sin\phi_0 \underline{k} \times \underline{V}_\psi = -\nabla p \quad (4-49)$$

Professor E.N. Lorenz of this department has suggested to me one relatively simple way of getting around this difficulty. He points out that if we retain the variability of $\sin\phi$ in (4-49), we can still arrive at system of equations which possesses an energy invariant in the absence of heating and dissipation.

The difference will be, of course, that the stream function will no longer be linearly proportional to the perturbation pressure, expressed by (4-51). Consequently the vertical derivative of the stream function will no longer be proportional to the potential temperature through (4-57), but through a more complicated formula. While it can be shown that the new system of equations would not be entirely consistent from the point of view of scale analysis (i.e. some terms would be neglected which are as large or larger than some terms retained), this system would have the advantage of

allowing us, at least in a mathematically reasonable way, to deal with flows which straddle the equator.

The physical description implied by the new equations thus may not be very accurate within a few degrees of the equator. However, this might not be of too great importance, since at least the mean-solar equatorial jet is so broad in latitude. That is, these equations could still give reasonable results for the middle and low latitudes taken as a whole.

With the variability of the Coriolis parameter explicitly included in the heliostrophic balance, it would probably be quite difficult to find wavelike solutions analytically, but again the initial value problem could be tried. Since the modified equations if heating and dissipation are suppressed retains an energy invariant, any increase in the total energy for a numerical time integration would have to be due to computational sources (for a general discussion of the importance of energy invariance in hydrodynamical modeling, see Lorenz (1960b)).

Once we have included the equator into the system, it is a relatively simple matter to eliminate another undesirable feature of our earlier systems, namely the lateral boundaries. This can be done by writing the equations with the same approximations as before, but in spherical rather than cartesian coordinates. The resulting system could then be applied to a complete spherical shell. Professor Lorenz is now developing such a system for the non-magnetic problem, to be solved numerically on the LGP 30 computer. In its present form, it has two layers in the vertical, with zonal wave numbers

assigned, but grid points for the meridional coordinate. It should not be difficult to generalize this model to include the effect of horizontal magnetic fields.

Unfortunately, none of the above generalizations will give a system which allows for the feedback of the vertical magnetic fields, which, as we have already discussed, is needed before we can have any hope of simulating a solar cycle. To get these feedbacks, it appears that we will have to at least go to some hydromagnetic generalization of the so-called "balance" equation models (c.f. Charney, 1962), or even of the so-called "primitive" equations (also Charney, 1962), in which the heliostrophic assumption is dropped, but the hydrostatic assumption retained. There are now fairly good stable finite difference schemes for numerical solution of such equations (Lilly, 1965). However, as evidenced by the analogous numerical general circulation models for the terrestrial atmosphere (e.g. Smagorinsky (1963), Mintz (1965)), this would be a very considerable undertaking. Before embarking on such a course, it would appear to be prudent to first study in greater detail some simpler models.

In the above discussion of criticisms and possibilities for further work, we directed our attention to specific defects and possible generalizations of the systems of equations developed in earlier chapters. Even if these changes were made, the resulting systems would remain within the more general theoretical framework we postulated for the convective layer of the sun in Chapter 3. That is, the essential link between the available potential energy of the system and the kinetic energy of the large scale

motions would be some form of baroclinic instability. However, in the final analysis, there may be other mechanisms which are also reasonable in the light of the solar observations taken as a whole. Probably the most likely alternative possibility is that of nonlinear interactions of the large scale motions inferred from the spot statistics with the smaller scale modes of motion known to be present, particularly the supergranulation. That is, there may be a reverse cascade of kinetic energy flow from the supergranulation scale eddies to the suprasunspot scale eddies. If this were so, then the reverse cascade would proceed all the way from the supergranulation scale to the differential rotation itself. This possibility was briefly suggested also in Chapter 3, and in Starr and Gilman (1965a). If this were in fact the mechanism, then the problem of producing the necessary meridional temperature gradients, and the gravitationally stable (in the mean) vertical stratification would be circumvented. Leighton's theory of the dynamics of the large scale magnetic regions would fit naturally with such a scheme. For these reasons and others, it would be well worth while to look into this possibility much more deeply.

In closing this thesis, it can be remarked that although we have directed our modeling attempts toward the solar problem, the same approach may be applicable in some form to the dynamics of other cosmical magnetic fields, such as those of the planets Earth and Jupiter, and of magnetic stars.

Bibliography

- Abetti, G., 1957: The Sun. New York, MacMillan, 336 pp.
- Alfven, H., 1945a: Magneto-hydrodynamic waves and sunspots I.
Mon. Not. Roy. Ast. Soc. 105, 3.
- , 1945b: Magneto-hydrodynamic waves and sunspots II.
Mon. Not. Roy. Ast. Soc. 105, 382.
- , 1950: Discussion of the origin of the terrestrial and
solar magnetic fields. Tellus 2, 74.
- , 1961: On the origin of cosmic magnetic fields. Ap. J. 133,
1049.
- Babcock, H.W. and Babcock, H.D., 1955: The sun's magnetic field 1952-
1954. Ap. J. 121, 349.
- Babcock, H.D., 1959: The sun's polar magnetic field. Ap. J. 130, 364.
- Babcock, H.W., 1961: The topology of the sun's magnetic field
and the 22 year cycle. Ap. J. 133, 572.
- Backus, G.E. and Chandrasekhar, S. 1956: On Cowling's theorem on the
impossibility of self maintained axisymmetric homogeneous dynamos.
Proc. Nat. Ac. Sci. 42, 105.
- Beckers, J.M., 1960: Temperature variation on the sun with heliographic
latitude. Bull. Ast. Inst. Neth. 16, 133.
- Bjerknes, V., 1926: Solar Hydrodynamics, Ap. J. 64, 93.
- , 1937: Application of line integral theorems to the hydro-
dynamics of terrestrial and cosmic vortices. Astro. Norv. 2, 263.
- Bullard, E. and Gellman, H., 1954: Homogeneous dynamos and terrestrial
magnetism. Phil. Trans. Roy. Soc. 247, 213.
- Burington, R.S. 1953: Handbook of Mathematical tables and formulas.
Handbook Publishers, Sandusky, Ohio, 296 pp.
- Bumba, V. and Howard, R., 1965a: A study in the development of active
regions on the sun. Ap. J. 141, 1492.
- Bumba, V. and Howard, R., 1965b: Large scale distribution of solar
magnetic fields. Ap. J. 141, 1502.

- Chandrasekhar, S., 1953: The instability of a layer of fluid heated below and subject to Coriolis forces. Proc. Roy. Soc. (Ser. A.) 217, 306.
- Charney, J.G., 1962: Integration of the primitive and balance equations. Proceedings of the International Symposium on Numerical Weather Prediction in Tokyo. Tokyo, Meteorological Society of Japan 656 pp, p. 131.
- Charney, J.G. and Stern, M.E., 1962: On the stability of internal baroclinic jets in a rotating atmosphere. J. At. Sci. 19, 159.
- Chistyakov, V.F. 1960: The circulatory nature of the eleven year cycle of solar activity. Sov. Ast. 4, 405.
- Clark, A., 1965: Some exact solutions in magnetohydrodynamics with astrophysical applications. Phys. Fluids 8, 644.
- Cowling, T.G., 1953: Solar Electrodynamics, Chapter 8 in The Sun, Kuiper, G.P. Ed. Chicago, University of Chicago Press, 745 pp.
- _____, 1934: The magnetic field of sunspots. Mon. Not. Roy. Ast. Soc. 94, 39.
- Dungey, J.W. and Loughhead, R.E., 1954: Twisted magnetic fields in conducting fluids. Aust. J. Phys. 7, 5.
- Eady, E.T., 1949: Long waves and cyclone waves, Tellus 1, 33.
- Eddington, A.S. 1925: Circulating currents in rotating stars. Observatory 48, 285.
- Elsasser, W.M., 1956: Hydromagnetic dynamo theory. Rev. Mod. Phys., 28, 135.
- Evans, J.W., Main, P., Michard, R., and Servajean, R., 1962: Correlations in the time variations of macroscopic inhomogeneities in the solar atmosphere. Ap. J. 136, 682.
- Evans, J.W., and Michard, R., 1962a: Observational study of macroscopic inhomogeneities in the solar atmosphere I. Ap. J. 135, 812.
- _____, 1962b: Observational study of macroscopic inhomogeneities in the solar atmosphere II. Ap. J. 136, 487.
- _____, 1962c: Observational study of macroscopic inhomogeneities in the solar atmosphere III. Ap. J. 136, 493.
- Ferraro, V.C.A., 1937: The non-uniform rotation of the sun and its magnetic field. Mon. Not. Roy. Soc. 97, 485.
- Fultz, D., 1961: Developments in controlled experiments on large scale geophysical problems. Adv. Geoph. 7.

Goldberg, Leo, 1953: Chapter 1 (introduction) of The Sun, Kuiper, G.P., Ed., Chicago, Univ. of Chicago Press, 745 pp.

Hart, A.B., 1954: Motions in the sun at the photospheric level IV. Mon. Not. Roy. Soc. 114, 17.

—————, 1956: Motions in the sun at the photospheric level VI. Mon. Not. Roy. Soc. 116, 38.

Hide, R., 1958: An experimental study of thermal convection in a rotating fluid. Phil. Trans. Roy. Soc. (ser. A.) 250, 441.

—————, 1964: Hydromagnetic oscillations of a rotating spherical shell of an incompressible fluid and the theory of the geomagnetic secular variation. Scientific Report HRF/SR12, Hydrodynamics of Rotating Fluids Project, Dept. of Geology and Geophysics, M.I.T.

Hide, R. and Roberts, P.H., 1962: Some elementary problems in magneto-hydrodynamics. Article in Advances in applied mechanics 1. New York, Academic Press p. 215.

Howard, L.N., 1961: Note on a paper of John W. Miles. J. Fl. Mech. 10, 509.

Howard, R., 1959: Observations of solar magnetic fields. Ap. J. 130, 193.

de Jager, C., 1959: Structure and dynamics of the solar atmosphere. Handb. d. Physik. 52, 80.

—————, 1964: Energy transport and "turbulence" in a sunspot. Bull. Ast. Inst. Neth. 17, 253.

Kiepenheuer, K.O., 1953: Solar activity, Chapter 6 in The Sun, Kuiper G.P. Ed. Chicago, University of Chicago Press, 745 pp.

Kippenhahn, R., 1963: Differential rotation in stars with convective envelopes. Ap. J. 137, 664.

Klyakotko, M.A., 1958: On the structure of the velocity field in motions in latitude of non-recurring sunspots. Sov. Ast. 2, 689.

Krogdahl, W., 1944: Stellar rotation and large scale currents. Ap. J. 99, 191.

Leighton, R.B., 1959: Observations of solar magnetic fields in plage regions. Ap. J. 130, 366.

—————, 1964: Transport of magnetic fields on the sun. Ap. J. 140, 1547.

- Leighton, R.B. (unpublished). Possible explanation of the correlation of latitude and longitude changes of sunspot groups.
- Leighton, R.B., Noyes, R.W., Simon, G.W., 1962: Velocity fields in the solar photosphere I. Ap. J. 135, 474.
- Lilly, D.C., 1965: On the computational stability of numerical solutions of the time-dependent non-linear geophysical fluid dynamics problems. Mon. Wea. Rev. 93, 11.
- Lorenz, E.N., 1955: Available potential energy and the maintenance of the general circulation. Tellus 7, 157.
- , 1960a: Maximum simplification of the dynamic equations. Tellus 12, 243.
- , 1960b: Energy and numerical weather prediction. Tellus 12, 364.
- , 1962: Simplified dynamic equations applied to the rotating-basin experiments. J. At. Sci. 19, 39.
- , 1963: The mechanics of vascillation J. At. Sci. 20, 448.
- Minnaert, M., 1953: The photosphere, Chapter 3 in The Sun, Kuiper, G.P., Ed. Chicago, University of Chicago Press 745 p.
- Mintz, Y. 1965: Very long-term global integration of the primitive equations of atmospheric motion. WMO Technical Note No. 66, "WMO-IUGG Symposium on Research and development aspects of long range forecasting, Boulder Colorado 1964," Geneva, p. 141.
- Mulders, G.F.W., and Slaughter, C.D., 1965: Absence of temperature difference between the sun's equatorial and polar limb near solar minimum. Pub. Ast. Soc. Pac. 77, 295.
- Ness, N.F. and Wilcox, J.M., 1964: Solar origin of the interplanetary magnetic field. Phys. Rev. Letters 13, 461.
- , 1965: Sector structure of the quiet interplanetary magnetic field. Science 148, 1592.
- Newton, H.W. and Nunn, M.L., 1951: The sun's rotation derived from sunspots 1934-44 and additional results. Mon. Not. Roy. Ast. Soc. 111, 413.
- Opik, E.J., 1951: Rotational currents. Mon. Not. Roy. Ast. Soc. 111, 278.
- Pagel, B.E.J., 1961: The distribution of temperature in neighborhood of the solar limb. Ap. J. 133, 924.

- Parker, E.N., 1955a: Hydromagnetic dynamo models. Ap. J. 122, 293.
- , 1955b: The formation of sunspots from the solar toroidal field. Ap. J. 121, 491.
- , 1963a. A kinematical theory of turbulent magnetic fields Ap. J. 138, 226.
- , 1963b: Kinematical hydromagnetic theory audits application to the low solar photosphere. Ap. J. 138, 552.
- Pedlosky, J., 1964a: The stability of currents in the atmosphere and the ocean: Part I. J. At. Sci. 21, 201.
- , 1964b: The stability of currents in the atmosphere and the ocean: Part II. J. At. Sci. 21, 342.
- Phillips, N.A., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two level, quasi-geostrophic model. Tellus 6, 274.
- Phillips, N.A., 1956: The general circulation of the atmosphere: a numerical experiment. Q.J. Roy. Met. Soc. 82, 123.
- Randers, G. 1938: Convection currents in rotating stars. Astr. Norv. 3, 97.
- , 1942: On the rotation of stars with convective cores. Ap. J. 95, 454.
- Roxburgh, I.W., 1964: On stellar rotation, I; The rotation of the upper main sequence stars. Mon. Not. Roy. Ast. Soc. 128, 157.
- Schwarzschild, M., 1942: On stellar rotation. Ap. J. 95, 441.
- , 1947: On stellar rotation II. Ap. J. 106, 427.
- , 1958: Structure and evolution of the stars, Princeton, Princeton Univ. Press, 296 pp.
- , 1959: Photographs of the solar granulation taken from the stratosphere. Ap. J. 130, 345.
- , 1961: Convection in stars. Ap. J. 134, 1.
- Schwarzschild, M., and Bahug, J., 1961: The temperature fluctuations in the solar granulation. Ap. J. 134, 337.

- Severny, A., 1964: Solar magnetic fields. Space Science Reviews 3, 451.
- Simon, G.W. and Leighton, R.B., 1964: Velocity fields in the solar photosphere III. Large Scale motions. The chromospheric network, and magnetic fields. Ap. J. 140, 1120.
- Starr, V.P., 1948: An essay on the general circulation of the earth's atmosphere. Jou. of Met. 5, 39.
- _____ and Gilman, P.A., 1965a: Energetics of the solar rotation. Ap. J. 141, 1119.
- _____, 1965b: On the structure and energetics of large scale hydromagnetic disturbances in the solar photosphere. Tellus, 17, 334.
- Stern, M.E., 1963: Joint instability of hydromagnetic fields which are separately stable. Phys. Fluids 6, 636.
- Thompson, P.D., 1961: Numerical weather analysis and prediction. New York, MacMillan, 170 pp.
- Tuominen, J., 1955: On the mean latitudinal movements of sunspots. Vistas in Astronomy, 1, p. 631.
- Van de Hulst, H.C., 1953: The chromosphere and corona. Chapter 5 in The Sun, Kuiper, G.P., Ed. Chicago, University of Chicago Press, 745 pp.
- Veronis, G.: Large-Amplitude Benard convection. To be published in the Journal of Fluid Mechanics.
- Ward, F., 1964: The general circulation of the solar atmosphere from observational evidence. Pageoph 58, 157.
- _____, 1965a: The general circulation of the solar atmosphere and the maintenance of the equatorial acceleration. Ap. J., 141, 534.
- _____, 1965b: The effect of some systematic errors in the determination of the general circulation of the solar atmosphere. Pageoph 60, 126.
- _____ (unpublished) reply to Leighton, R.B., unpublished.
- Whitney, C.A., 1963: Thermal response of the solar atmosphere. Ap. J. 138, 537.
- von Zeipel, H. 1924a: The radiative equilibrium of a rotating system of gaseous masses. Mon. Not. Roy. Ast. Soc. 84 (Supp.), 665.
- _____, 1924b: The radiative equilibrium of a slightly oblate rotating star. Mon. Not. Roy. Ast. Soc. 84 (Supp.), 684.

Appendix

In this appendix we present, in graphical form, data on sunspot displacements (Figs. (A-1)-(A-5)) and the large scale line of sight magnetic fields (Figs. (A-6)-(A-7)). The sunspot displacement statistics we compiled by Dr. F. Ward from Greenwich Observatory data, for five solar cycles, including the years 1905-1954 (the averages for 5 cycles are unpublished). Each plotted point in Figs. (A-1)-(A-5) represents an average value for the northern and southern hemispheres, for latitude belts of 5° centered about the latitude of the plotted point. Figures (A-6) and (A-7) are two of many synoptic charts of solar magnetic fields, published by R. Howard and V. Bumba in the *Astrophysical Journal* of May 15, 1965. These charts were compiled by them from their solar magnetograms. The magnetograms are records of the line of sight magnetic fields seen on the sun, as deduced from the Zeeman splitting of selected spectral lines. To compile the charts, magnetograms are taken each day, atmospheric conditions permitting, primarily along the central meridian of the sun. The fields are generally strongest at the center of the solar disk, falling off toward each pole and toward the east and west limbs, strongly suggesting that these fields are primarily radial. For other details, see the figure legends, and Chapter 1 of the text.

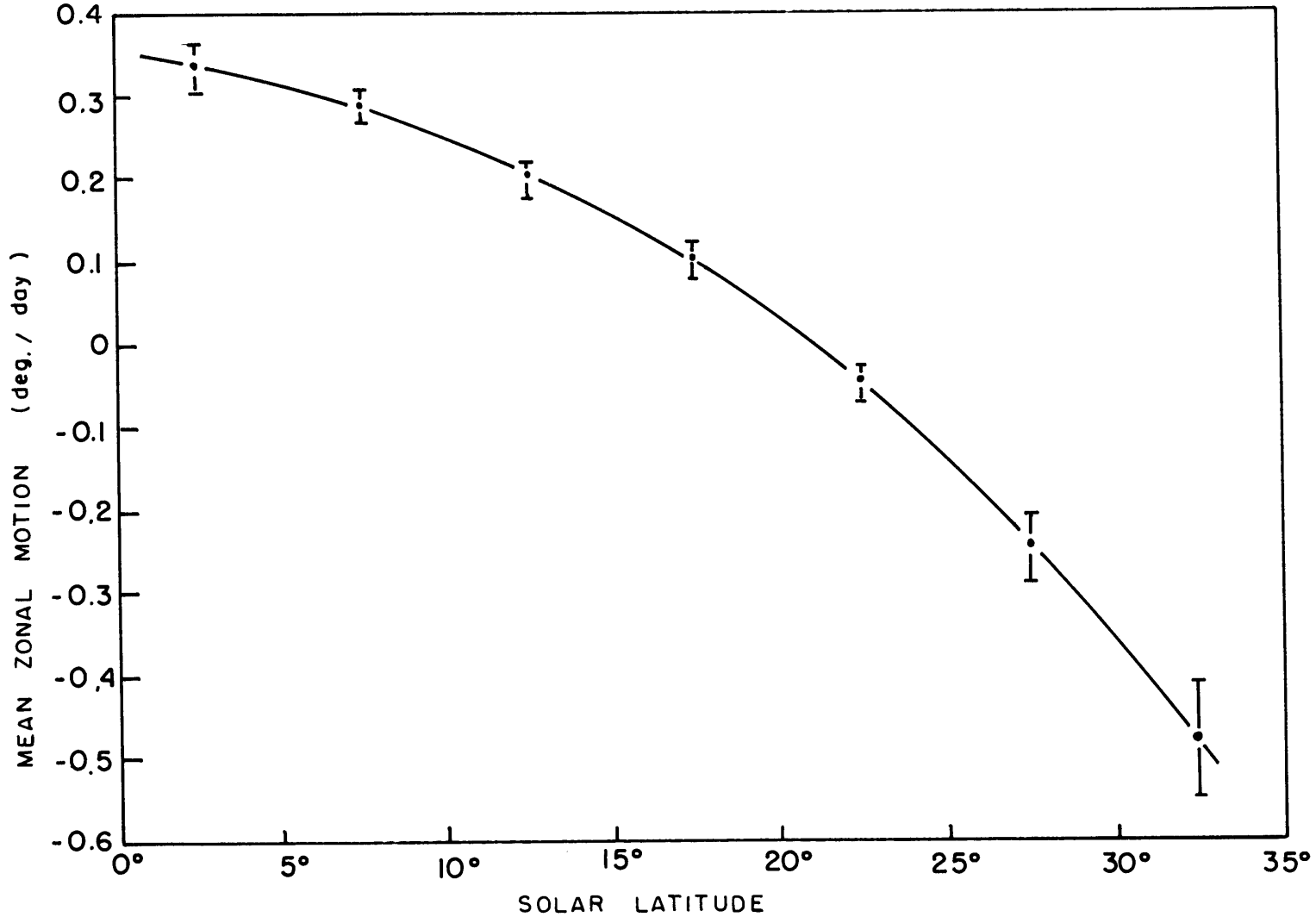


Fig. A-1. Mean zonal motion (differential rotation) of sunspots in degrees longitude per day. Zonal motion is measured relative to the Greenwich convention for the solar rotation, 14.186 deg/day. 95% confidence limits for the data are plotted for each point.

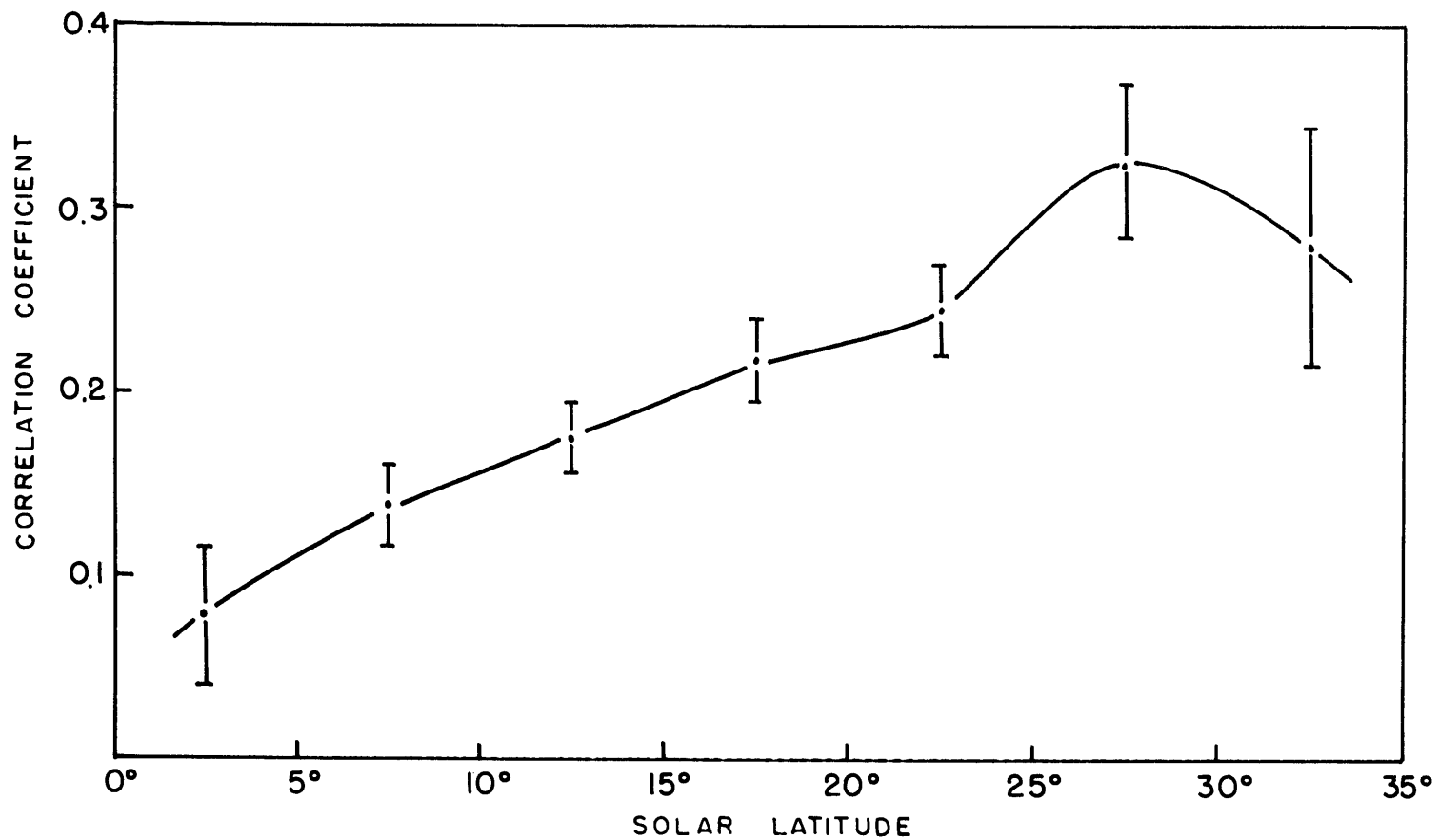


Fig. A-2. Correlation coefficient of the longitudinal and latitudinal spot motions as a function of solar latitude. 95% confidence limits are plotted for each point. A positive correlation indicates equatorward motion is correlated with motion toward the west limb (eastward motion, in meteorological coordinates).

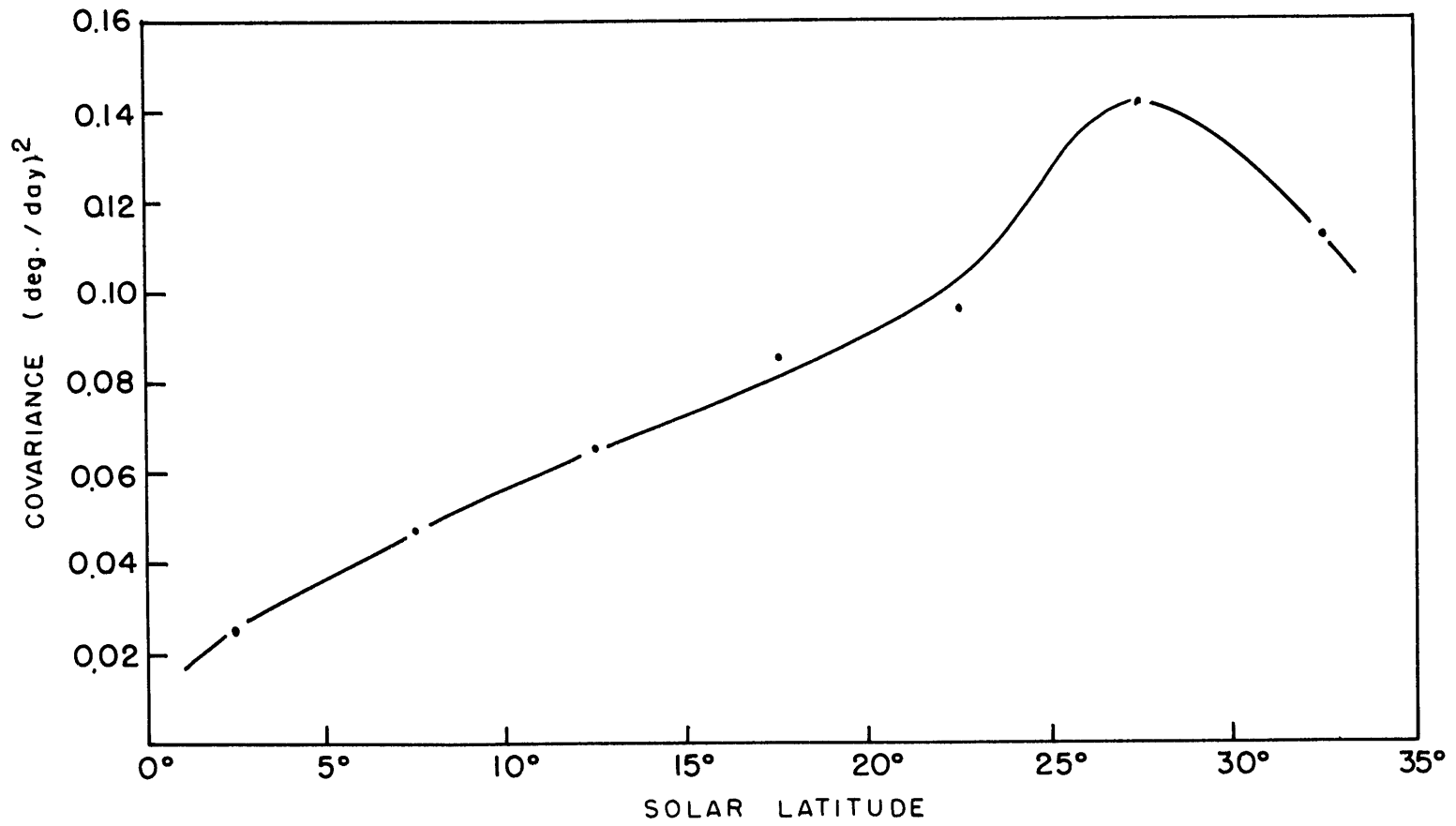


Fig. A-3. Covariance of latitudinal and longitudinal motions in $(\text{deg./day})^2$ as a function of solar latitude.

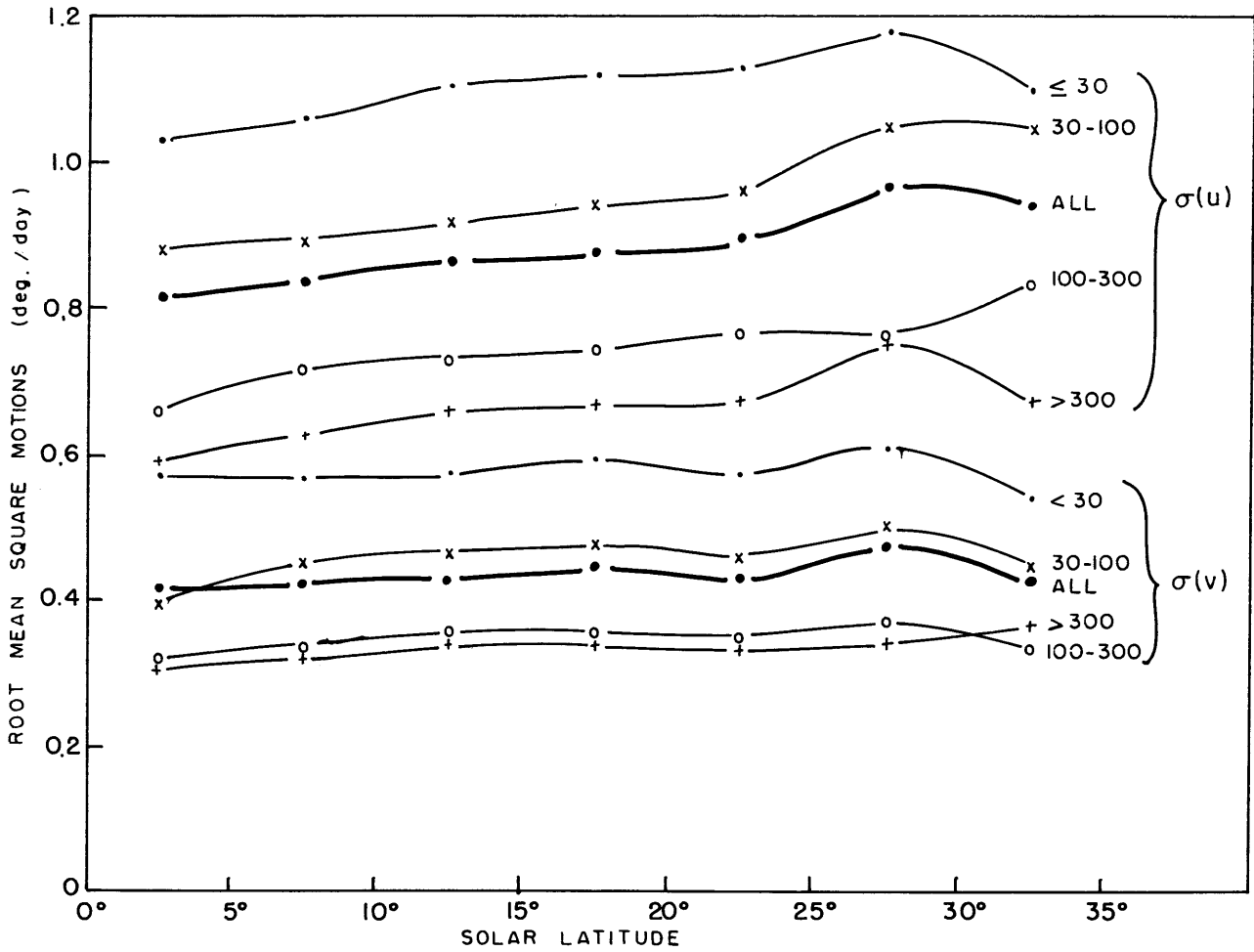


Fig. A-4. Root mean square longitudinal ($\sigma(u)$) and latitudinal ($\sigma(v)$) motions of spots separated according to spot areas (measured in 10^{-6} of the solar disk). Note that $\sigma(u) \sim 2\sigma(v)$.

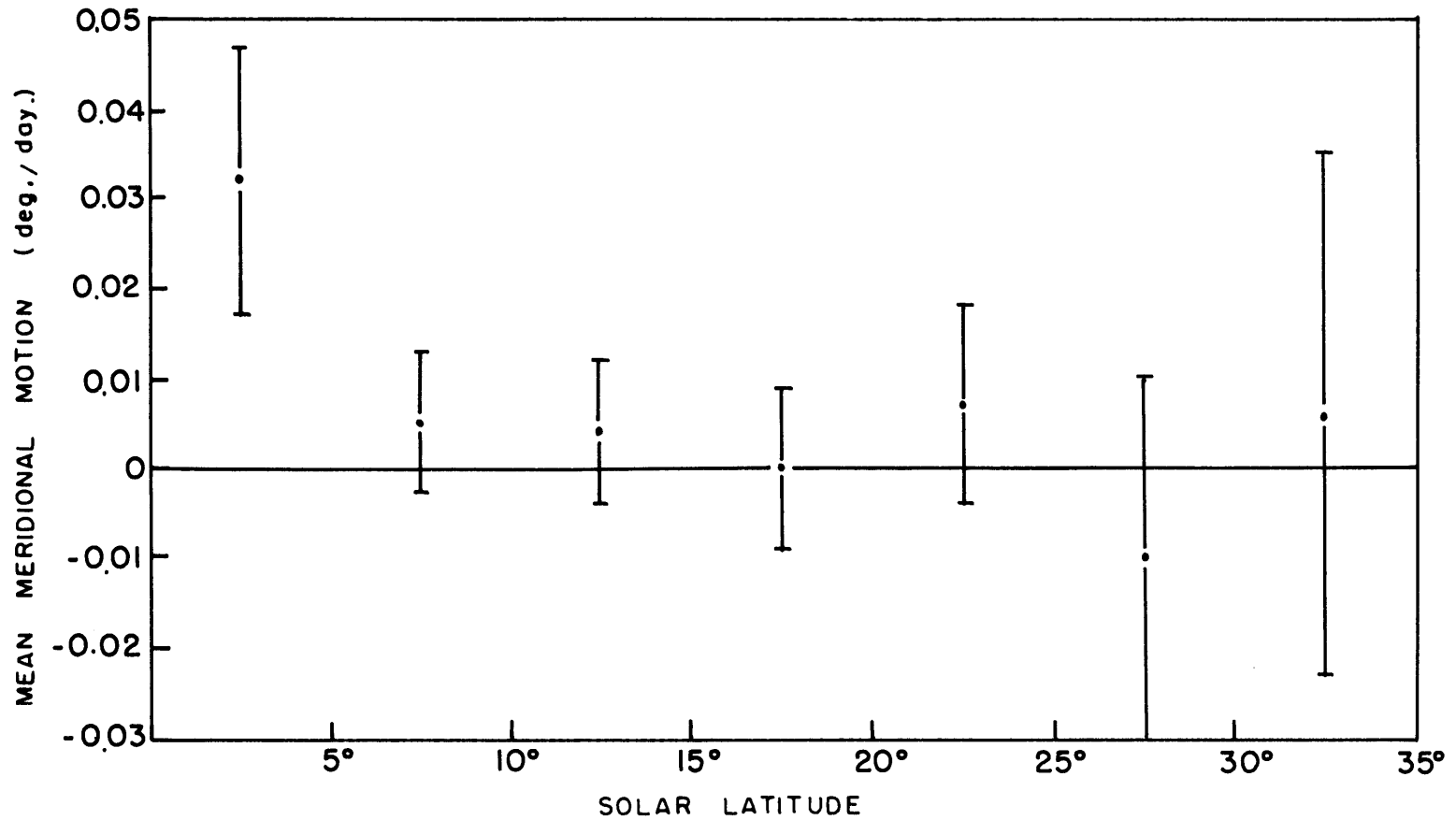


Fig. A-5. Mean meridional motion in degrees latitude per day as a function of solar latitude. 95% confidence limits are plotted for each point. Note that only in the 0-5° latitude belt is the mean meridional motion significant at the 5% level.

SOLAR MAGNETOGRAM AT ACTIVITY MAXIMUM
(Bumba and Howard, 1965)

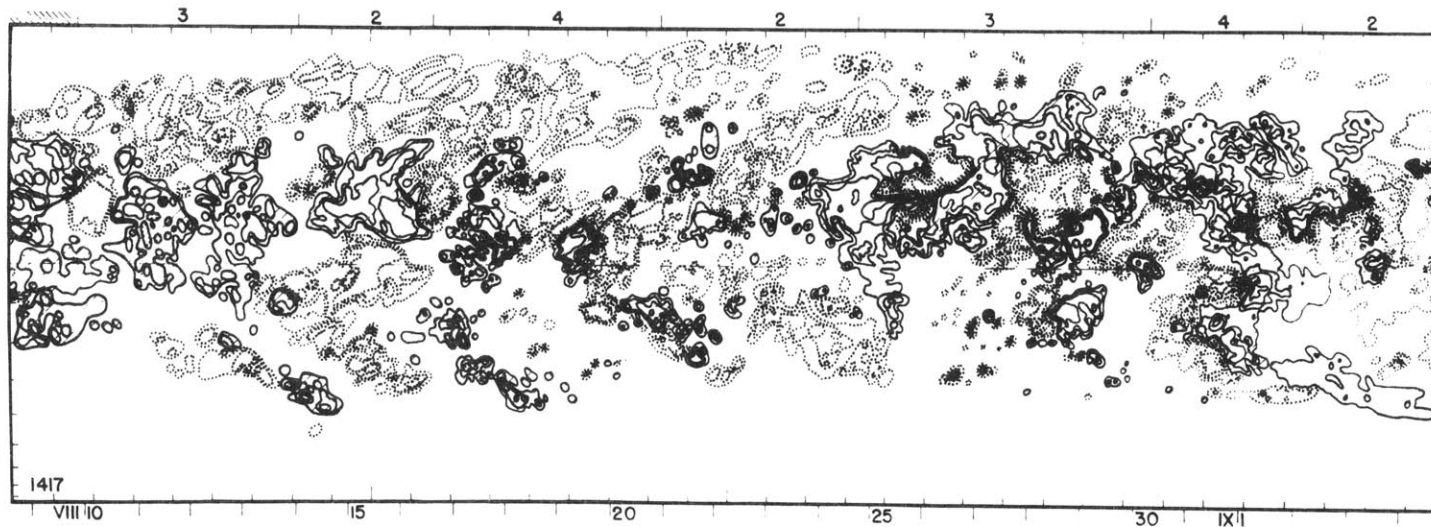


Fig. A-6. Synoptic chart of solar magnetograms, of Bumba and Howard (1965), taken at activity maximum (rotation No. 1417, August 1959). Ordinate is the sine of the latitude, with figure centered about the equator. Lower abscissa is plotted in days of the month. The chart represents one complete rotation. Numbers at the top edge indicate the quality of the magnetograms from which the chart was drawn, with 4 the best. Solid contours and cross hatchings represent positive polarity, (toward the observer) dotted contours represent negative polarity. Isogauss contours are for 2, 6, 10, 15, 25 gauss. On the abscissa, 10° intervals of solar longitude are also plotted. Because the synoptic patterns show many persistent features on successive rotations, we can think of the time coordinate as longitude, plotted increasing to the left. Therefore in a qualitative sense this chart represents the large scale fields everywhere on the sun at a given time. With longitude as a coordinate, the solar rotation is from right to left in the figure. Thus the magnetic field patterns are tilted up stream away from the equator. For further discussion, see Chapter 1.

(By permission from the Astrophysical Journal, Vol. 141,
University of Chicago Press).

SOLAR MAGNETOGRAM AT ACTIVITY MINIMUM
(Bumba and Howard, 1965)

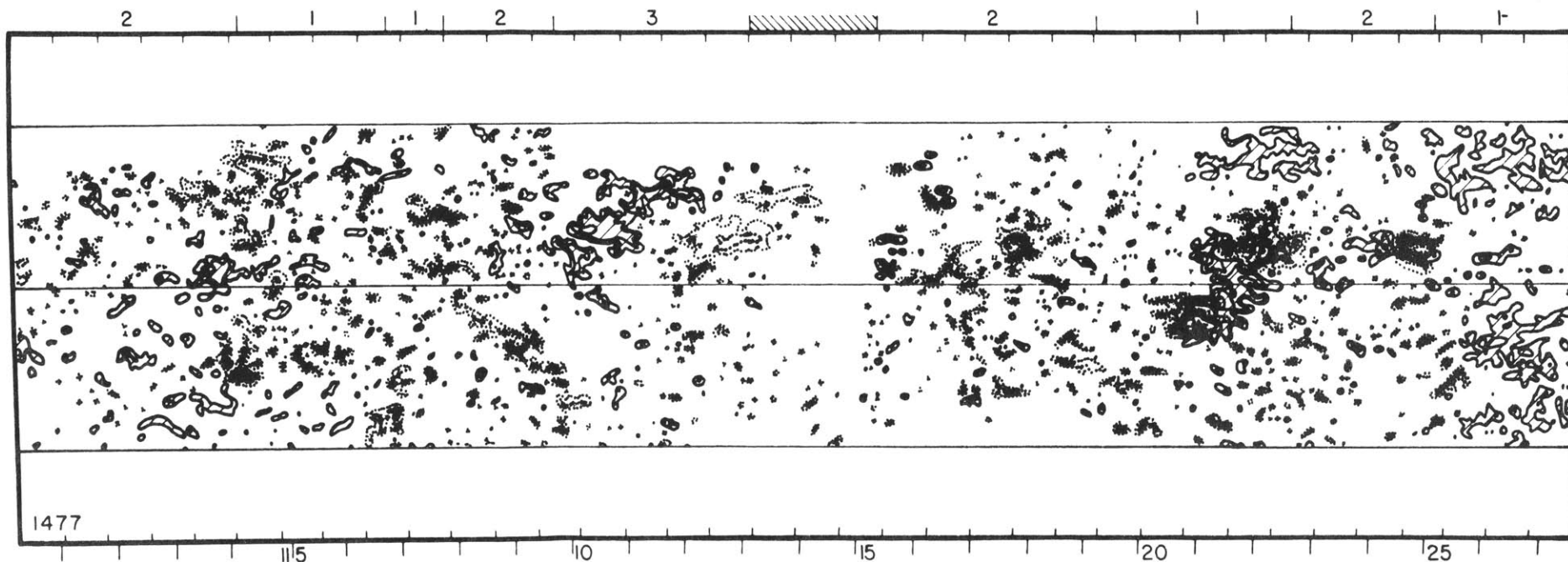


Fig. A-7. Synoptic chart at activity minimum (rotation No. 1477, February 1964). Details same as Fig. A-6, except that no data is plotted poleward of 40° , due to the weakness and therefore lack of reliability of the fields there.

(By permission from the Astrophysical Journal, Vol. 141, University of Chicago Press).

Biographical Note

The writer was born on May 28, 1941 in Hartford, Connecticut. Until college he lived in Storrs, Connecticut, still his family home, where he attended local schools. He studied at Harvard College from 1958 to 1962, receiving a B.A. degree magna cum laude in Physics in June 1962. He was enrolled at M.I.T. in September 1962, and received an M.S. degree in Meteorology in February 1964.

He is presently engaged to Miss Susan Hartman of Harleysville, Pennsylvania, and plans to be married on July 9, 1966. Beginning in September 1966, he will be an assistant professor in the Department of Astro-Geophysics at the University of Colorado in Boulder, Colorado.

Publications by the author

"On the vertical transport of angular momentum in the atmosphere".
Pure and Applied Geophysics 57, 161 (1964).

"On the mean meridional circulation in the presence of a steady
state, symmetric, circumpolar vortex". Tellus 16, 160 (1964).

"The mean meridional circulation of the southern hemisphere inferred
from momentum and mass balance". Tellus 17, 277 (1965).

(With V.P. Starr) "Energetics of the solar rotation". Astrophysical
Journal 141, 1119 (1965).

(With V.P. Starr) "On the structure and energetics of large scale
hydromagnetic disturbances in the solar photosphere". Tellus 17,
334 (1965).