THE MAGNETIC HISTORY OF SOLAR ACTIVE REGIONS

Thesis by

James Marshall Mosher

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1977

(Submitted November 1, 1976)

Acknowledgements

I wish to thank my thesis advisor, Robert B. Leighton, for his confidence and support. I wish also to thank Harold Zirin for many pleasant hours spent at Big Bear, and Bob Howard for providing access to data in the Mount Wilson Archives.

A particular debt is owed to Tom Pope, of Lockheed Solar Observatory, whose questioning, prodding, and encouragement are largely responsible for the successful completion of this project.

Finally, I must thank my father, not only for typing the entire manuscript, but also for helping to record and reduce data whenever such help was needed.

Abstract: An attempt has been made to use recent magnetic observations to trace the history of a typical solar active region from birth to death. By comparing the short-term motions to the long-term spreading, it is demonstrated that the decay process is dominated, over periods ranging from days to months, by a random walk of field lines, with a diffusion constant of roughly 200-400 km^2/sec . While the interaction between diffusion and differential rotation dictates the geometric pattern of the decaying region; the actual quantity of surviving flux appears to be less, and its ultimate annihilation more thorough, than would have been expected. This probably indicates a continued subsurface coupling between opposite polarity features. In addition, the long-range agreement between theory and observation is considerably improved by postulating the existence, in the middle latitudes of each hemisphere, of a systematic, poleward-moving meridional flow of about 3 m/sec .

The outlook for being able to make continued progress towards the understanding of basic solar phenomena by further efforts in this direction is promising.

TABLE OF CONTENTS

1. Introduction

1.1 Background

The first comprehensive attempt to describe the magnetic history of a solar active region was apparently that of Cowling (1946), who, upon the basis of the early Mount Wilson observations of sunspots (Hale and Nicholson, 1938), developed two now well-published curves showing the growth and decay of "30 and 50 day spots". These curves indicated a decay much more rapid than that which could reasonably be expected on the basis of Ohmic losses, and led to the suggestion that in all probability the bulk of the field was being eroded away by convective motions in the atmosphere.

In the years since Cowling's study the accepted picture of active region structure has undergone a dramatic evolution, and it is now well understood that the spots themselves represent only a tiny portion of the total magnetic picture. In particular, it is now understood that the bulk of the field in a decaying region is represented not by the spots, but by the 'calcium plage', which was very difficult to observe with the old visual technique, but easily monitored by modern magnetographs, (Beckers, 1971). These modern instruments are capable of providing a detailed quantitative history of development for individual active regions in terms of the flux, area, and separation between the polarities. In addition, the perfection of high resolution photographic cancellation

techniques (Leighton, 1959; Leighton, et al., 1962) has provided a much clearer picture of the photospheric velocity fields under whose influence the field is assumed to move.

There is a certain inspiration to re-evaluate our understanding of the development of solar active regions in years of low activity, since it is mainly then that they can be regarded as isolated features seen against a "quiet" background (see, for example, Babcock and Babcock, 1955; Bumba and Howard, 1965ab). For the most part, however, these modern re-evaluations have concentrated on the quantitative description of a few individual regions, with a more qualitative overview; tending to abandon the original effort towards developing a solid quantitative description of the average region.

The present paper is intended to suggest the potential for such a study based on the wealth of detailed magnetic data now available. The results are drawn in part from an observational program undertaken at Big Bear during the summer of 1972, but rely primarily on information derived from the monthly reports of the National Oceanic and Atmospheric Administration (Solar-Geophysical Data) published during 1974-1976. Data based on these published magnetograms, particularly with regard to the measurement of flux are obviously rather crude; but even without reference to the primary data they seem sufficient to indicate certain broad

areas in which the present understanding of solar processes is likely to be adequate to explain the observations, and others in which it is not.

While many of the conclusions may be overturned by new observations, or even by a more careful analysis of the existing ones, one senses that it is, on the whole, at least a step in the direction of progress.

1.2 Overview

The solar active region is, basically, a magnetic phenomenon; which, at the photospheric level, can be thought of as a bipolar injection of field whose properties are, at any moment, characterized by a total flux and the area over which it is spread, and a separation between the two polarities. In general, the pattern of development is one of rapid growth followed by slow decay. The field seems to grow by eruption, and decay by spreading.

Many unknowns go into determining the nature of the growth phase; and in general, the factors controlling the strength of the new fields, the dimensions of the eruption, and the time scale are not well understood at present; nor, even is the nature of the instability permitting the transition from spot to plage. In the following pages we shall offer some suggestions, but the ultimate answers are still far from clear.

By contrast, a remarkably wide range of observational aspects of the decay phase can be brought together under the general umbrella of a single unifying concept: the 'random walk' (Leighton, 1964), whereby the individual flux tubes in a decaying region are assumed to move freely in response to the essentially random pattern of supergranular motions to which they are subjected, the fields 'adding' in areas where like polarities overlap and 'cancelling' in areas where opposite polarities overlap. While there may be some esthetic difficulty in accepting the degree of subsurface tangling which such a model would seem to predict, it is nonetheless becoming increasingly apparent (as we shall attempt to demonstrate) that the growth of an active region in area and separation (as well, presumably, as its decline in flux) over periods of many months, can all be explained by a single diffusion constant in the range $D = 200\text{-}400$ km^2/sec.

Moreover, this diffusion constant appears to be the same as that implied by the direct observation of the motion of individual field patches over periods of 24 hours and less; which, in turn, is consistent with the kinds of motions that might be expected due to a bending of the field lines in response to the changing supergranular network.

If one accepts the general validity of the random walk or diffusional picture of active region decay, then certain

consequences flow from it: one can, for instance, estimate the "lifetime" of the average flux tube in regions in various states of decay; calculate the width of the neutral channel separating the two polarities; predict the rate of magnetic energy dissipation (which one might reasonably expect to be related to flare production, and other manifestations of active region "activity"); or calculate the large scale pattern of fields which would result as the residue of many individual regions erupting over the course of a cycle.

Where these predictions fail to coincide precisely with observation, that disagreement can provide important insight into the nature of the solar fields by demonstrating the necessity of particular mechanisms whose reality would otherwise be difficult to establish. A case in point is the evidence for a systematic meridional flow which grows out of an apparent disagreement between the observed and predicted large scale fields. A more careful analysis of the data might well reveal similar discrepancies over other time scales.

We shall begin by discussing the growth of active regions, which, in the present content is primarily of interest in terms of establishing the initial conditions of the diffusion problem. In the subsequent sections we will explore the areas in which the predictions of that model can be compared with observation, and what the nature of that comparison seems to be.

2. The Growth Phase

2.1 Introduction

Since the solar active region is a magnetic phenomenon, it is natural to divide its development into a period of growth and a period of decay according to whether the net surface flux is increasing or decreasing with time. Such a definition is essentially the same as one based on spot area, flare activity, or X-ray intensity, but should not be confused with actual physical growth in the sense of increasing area: the area occupied by the active region increases during both the growth and the decay phases, at nearly the same rate.

By net flux we are referring to $\int B \cdot dA$ over one polarity. The gross geometry of an active region is that of a kinked tube with both ends disappearing beneath the surface. In the strictest sense, the net flux through the surface is zero; although obviously the 'actual' flux can either increase or decline depending on whether new lines are added to or removed from the tube.

Key points in the following description are the following:

1. The growth phase of spot-producing regions is short-lived, ranging from 1-10 days, with 4 or 5 being typical.

2. The flux increases rapidly ($\sim 2 \times 10^{16}$ Mx/sec)

attaining a final value which is roughly proportional to the maximum sunspot area (as first pointed out by Sheeley, 1966).

3. The separation between the centers of weight of the positive and negative fields increases very rapidly ($\gtrsim$.5 km/sec) during the first day, then more slowly. A gradual increase in separation continues even during the decay phase, but at a much slower rate ($\sim$.015 km/sec). The transition appears to be rather abrupt, but a relatively large range of separations are encountered at the start of the decay phase (30,000-80,000 km), with little systematic dependence on region "strength".

4. The area occupied by each polarity increases, typically, at $\lesssim 3 \times 10^3$ km^2/sec. The area over which it is spread increases by $\sim 7 \times 10^3$ km^2/sec with no obvious transition from a "growth" to a "decay" pattern.

A more precise description of the growth of a typical active region could certainly be formulated, but it should be recognized that certain observational difficulties arise because of the closeness of the relevant time scale to the disk-passage time of the region: if its birth is observed near disk center, the maximum phase will occur near the limb; whereas if the early stages of decay can be easily observed, then the birth will have likely occurred behind the limb.

In addition very high resolution observations of the events immediately preceding the birth of a substantial active region are very difficult to obtain because the very low probability of such an event happening in any small pre-selected area ($\sim 2 \times 10^{-7}$/sec of useable time for the typical $4' \times 6'$ Big Bear frame).

2.2 A Description of Emerging Flux at High Resolution

Of the several examples of emerging flux observed with the videomagnetograph at Big Bear, the best documented case is probably McMath Region 11972. This particular region had the happy chance of erupting in a relatively quiet area about 20° east of center (on July 24, 1972), so that its development during the first few days could be studied in considerable detail. Figure 1 provides a sample of the daily survey pictures. The region was already prominent in H-alpha at the time of the first observation -- with the characteristic pattern of dark strongly aligned arch filaments and bright emission at the footpoints -- but a search of the full-disk Singer Link coverage indicated that the birth could not have taken place more than about two hours before. The axis of the initial arch filaments is highly inclined (50°) with respect to lines of constant latitude and in a direction opposite to that normally associated with active regions -- that is, with the preceding end closest to the pole, and the following end closest to the equator. A pair of dark, granule-

Figure 1: This sequence of photographs illustrates the development of McMath Region 11972 during the first seven days of its life. Reading from left to right, the pictures indicate its appearance in H-alpha, in FeI λ5324 (magnetic cancellation), and in the continuum 13 Å from H-alpha. In the videomagnetogram, positive polarity features appear lighter than the average gray background. The splotchiness evident in some of the pictures is due to imperfections in the original photographs.

With the clock in the upper left-hand corner, north is at the top, and solar rotation from left to right. The group is at a latitude of 7° in the northern hemisphere, and crosses the central meridian on July 26.

Each row presents pictures as nearly simultaneous as possible. The times are as follows:

(a) July 24, 1972	~1825 UT	(f) July 27, 1972	1702 UT
(b) July 24, 1972	~2220 UT	(g) July 28, 1972	1909 UT
(c) July 25, 1972	1602 UT	(h) July 29, 1972	~2145 UT
(d) July 26, 1972	0105 UT	(i) July 30, 1972	1703 UT
(e) July 26, 1972	~1725 UT		

The width of each frame is approximately 2.8×10^5 km.

(all photographs courtesy Big Bear Solar Observatory)

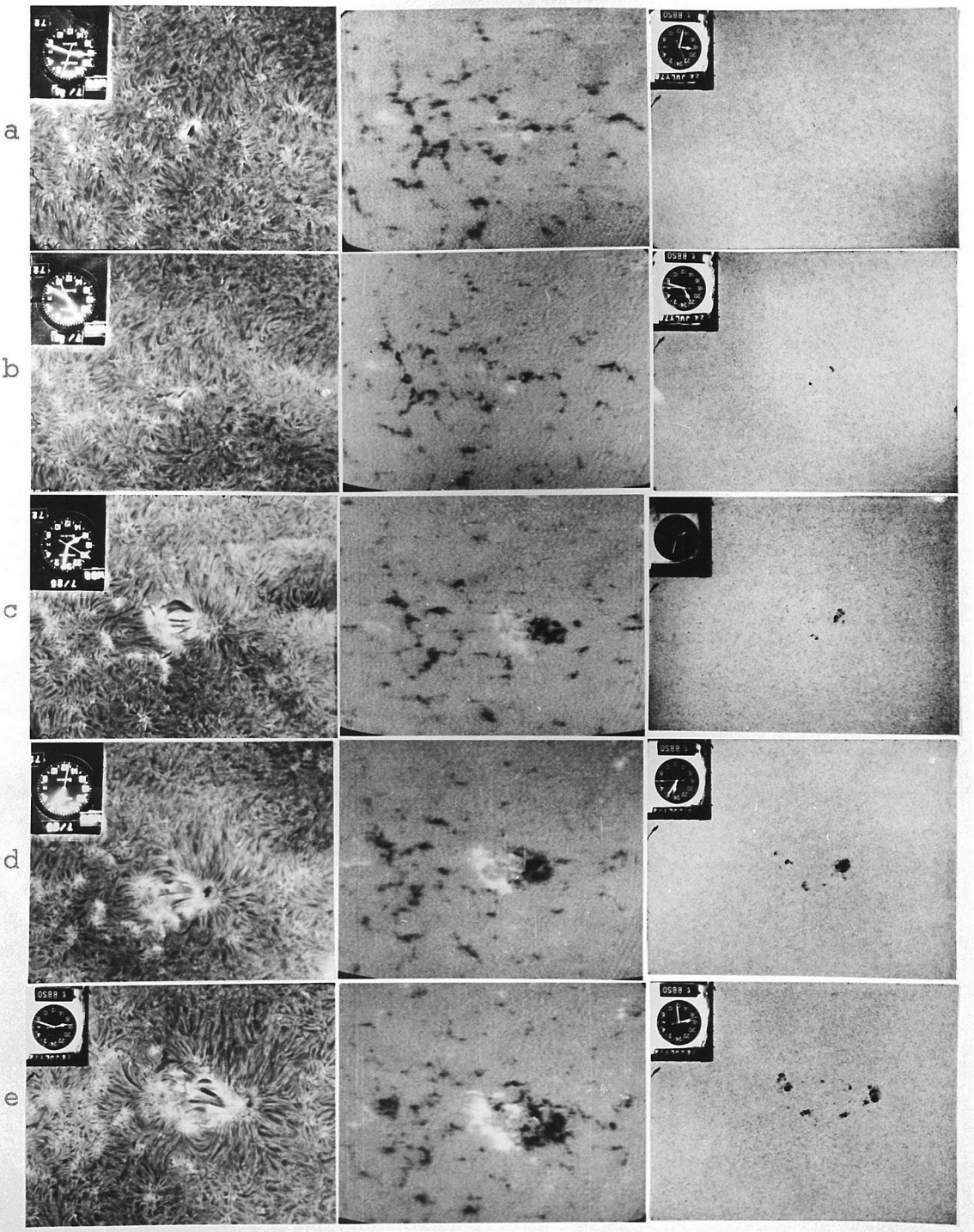


Figure 1

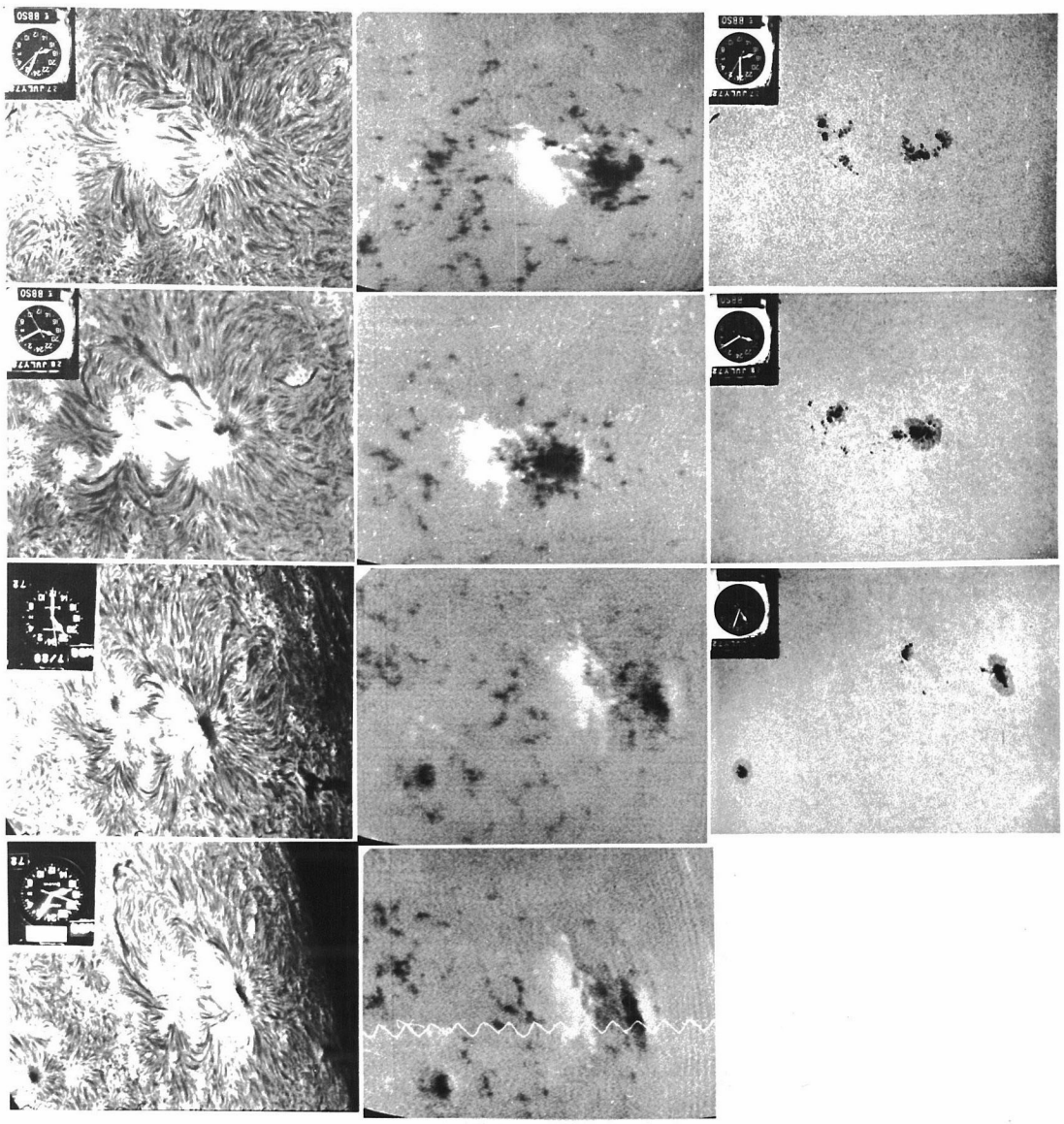
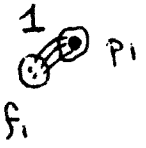


Figure 1 (contd.)

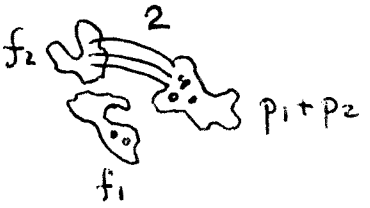
Figure 2: The growth of Region 11972 can be thought of as consisting of at least four major sub-eruptions. The simple bipolar configuration of the mature region (July 27-30), appears to result from a merging of the contributions from the sub-eruptions, as illustrated in the present figure. The solid contours represent the outlines of the magnetic field, while the filled-in areas correspond to the white-light spots (with stippling when a penumbra is present). The eruptions are labelled 1-4, and indicated, when active, by connecting lines corresponding to the H-alpha arch-filaments. The growth in spot area after July 26 seems to result more from a merging and 'condensation' of existing fields than from the eruption of new ones.

The numbers in parentheses after the time refer to Figure 1. The drawings are to scale, but enlarged.

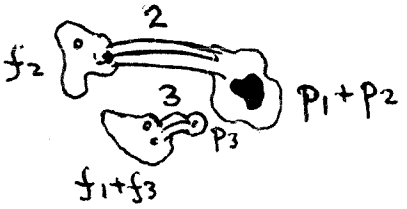
7-24 1900 UT (1ab)



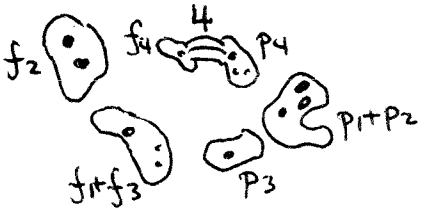
7-25 1600 UT (1c)



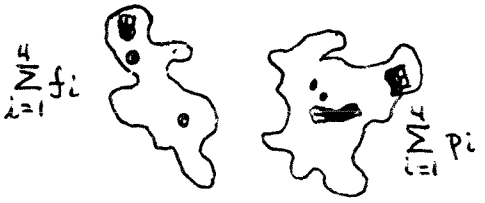
7-26 0100 UT (1d)



7-26 1730 UT (1e)



7-27 1700 UT (1f)



7-28 1900 UT (1g)

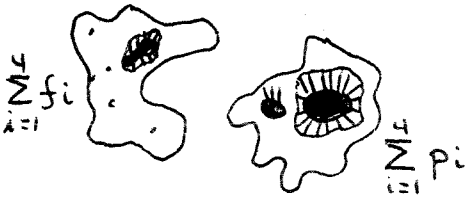


Figure 2

sized sunspot pores is already visible in the preceding end at the time of the first high-resolution observations. Similar pores form in the following end an hour later.

The locations of the pores are easily recognizable in the magnetograms, although the field appears to be considerably more extensive than the pores. For example, on the first high resolution magnetogram, taken at 1837 UT on the 24th, the preceding end is already marked by a strong patch of negative polarity covering 2×10^7 km^2 (the smallest resolved features measure about 2 television lines on a side, or 4×10^6 km^2 in area). The area of the underlying granule-sized pores is 1×10^6 km^2. In other words, the pores have the appearance of islands in a larger pool of relatively strong field. The granulation has a broken or "disturbed" appearance in most of this area.

At the following end the disparity in areas is even greater since there are no pores at all at the time of the observation being discussed. The field does, however, have a strong pore-like core ($\sim 1 \times 10^7$ km^2 in area) with a wispy, barely-detected tail ($\sim 4 \times 10^7$ km^2 in area) trailing off towards the preceding end. The disappearance of the diffuse component seems to be associated with the formation of pores in the following end at $\sim$ 1940 UT. Otherwise, the field does not appear to undergo any remarkable change in connection with the creation of the pores. This leaves one with

the distinct impression that the pore has been produced by a "condensation" of existing field, rather that by the eruption of some new kind of structure. In fact, the spots seem to be continually encroaching on the surface fields, commanding a larger and larger fraction of the total area.

Actual measurements of the spot and total field areas for Region 11972 are shown in Fig. 3. At the beginning of the growth phase, the field is completely of plage strength. By the end, nearly half the field area is accounted for by spots.

A second obvious difference in the character of the growth pattern of the spots, as opposed to that of the field in general, is that the spots (both umbras and penumbras) appear to grow by a discontinuous series of spurts, while the general field grows steadily. This behavior not only enhances the impression that the spots can be created or destroyed by a very minor condensation or expansion of the surface fields, but also permits one to estimate the strength of the plage field, since the amount of flux represented by spots of a given area is presumably well-known. For example, an increase of 3×10^7 km^2 in the area of the p-spot umbra between 24 and 33 hours, is offset by an (anomalous) increase of at most 6×10^7 km^2 in the area of the following plage; and between 72 and 100 hours (when observations are unfortunately rather difficult because the region is approaching the limb),

Figure 3: The growth of spot and network fields in Region 11972. The open circles represent the total area occupied by field on the videomagnetograms. The solid circles are the umbral areas as seen in the continuum, with upward extensions (shaded area) indicating the penumbra. Preceding and following fields are shown separately. The zero of time is 1600 UT on July 24, 1972.

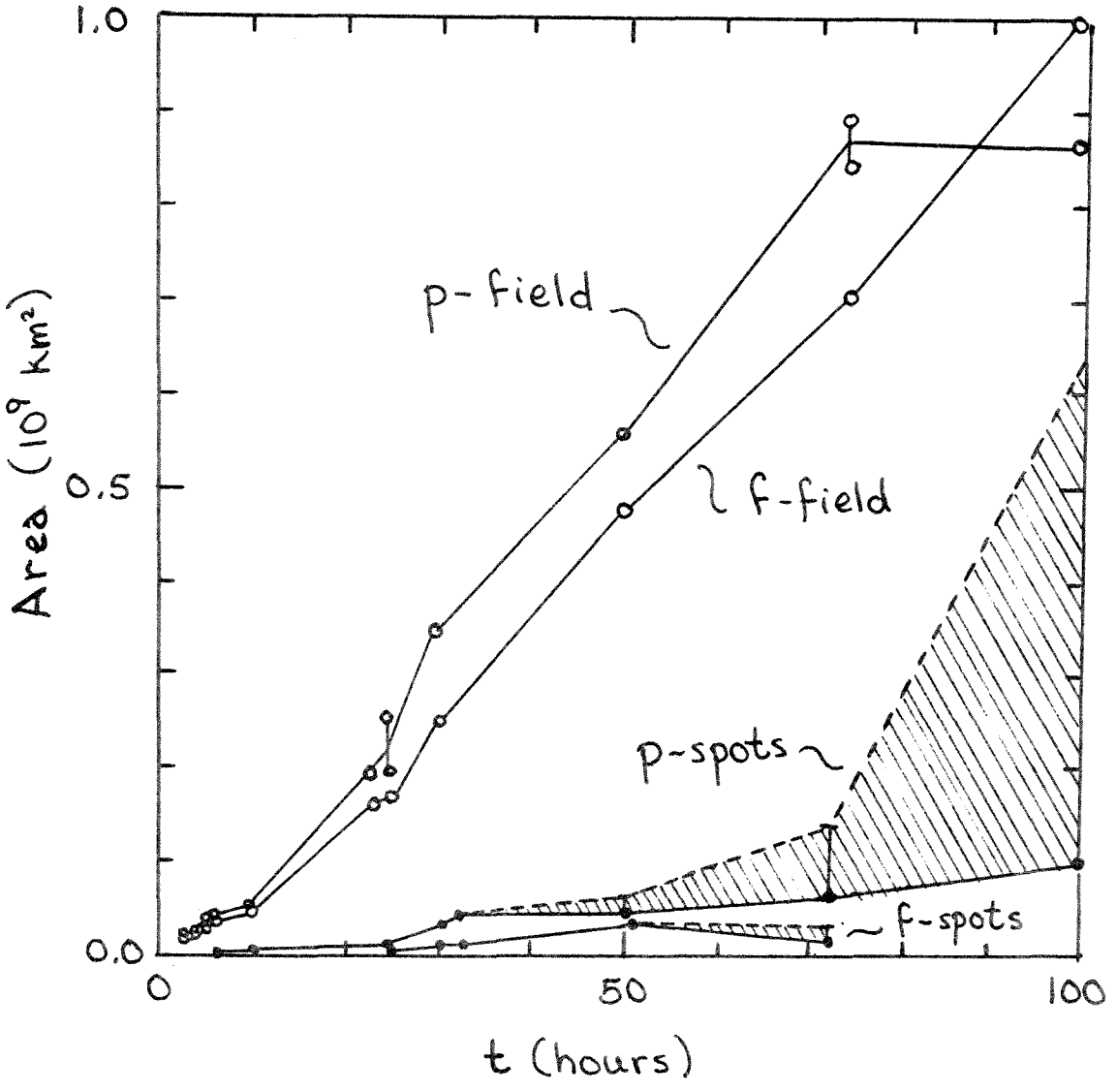

1.0
0.5
0.0
Area (10^9 km^2)
p-field
f-field
p-spots
f-spots
0
50
100
t (hours)

Figure 3

an increase of $\sim 5 \times 10^{8}$ km^{2} in the area of the p-spot penumbra (plus the complete disappearance of the f-spots) is accompanied by little or no change in the area occupied by the following field. These observations, and others like them, suggest that the strengths of the plage and spot fields cannot be very different.

An independent estimate of the strength of the plage fields can be made by determining the flux (see Appendix I). The videomagnetograms are presumably reasonably accurate as far as indicating the area occupied by field; the Mount Wilson data are presumably reasonably accurate as far as indicating the total flux. The flux in Region 11972 was estimated from the NOAA bulletins on two occasions: July 26 ($T \approx 50$ hours) and July 28 ($T \approx 100$ hours). If the results for the p- and f-ends are averaged, one seems to have $\sim 0.8 \times 10^{21}$ Mx in spots and $\sim 2.4 \times 10^{21}$ Mx in plage (at each end) on the first occasion; $\sim 3.7 \times 10^{21}$ Mx in spots and $\sim 5.5 \times 10^{21}$ Mx in plage on the second ($\pm \sim 50\%$). This implies a growth rate of $\sim 3 \times 10^{16}$ Mx/sec. Compared to the growth in area of 3×10^{3} km^{2}/sec, the average field strength would have to be 1000 gauss. The true plage field strength is likely to be even higher, since the finite resolution of the videomagnetograph undoubtedly exaggerates the true area of the field.

Of course the picture of continuous growth in the total flux is only an approximate one, relative to the pattern of growth of the spots. When examined in sufficient detail, it too is also discontinuous. Fig. 2 indicates how the growth of Region 11972 can be thought of as consisting of at least four distinct magnetic sub-eruptions over the course of 3 days. Each of these sub-eruptions is marked by arch-filaments in H-alpha, and the fields which they contribute seem free to merge with those already existing so that at the end of the growth period a single, simple bipolar structure results. The individual footpoints seem to strengthen as they pull apart, but this is probably just because the lines (which must initially lie horizontally in the surface) are twisting into an angle more favorable for observation. Any actual amplification of the fields on the surface would be resisted by eddy currents. The latter eruptions tend to be more generally east-west than the initial one, so that the region as a whole is brought around from its original high inclination to a more normal one. When viewed at low resolution it would appear to revolve through about 60°.

The motions of individual features in emerging regions have been studied by numerous authors (see, for example, Frazier, 1972; or Vorpahl and Pope, 1972). Velocities of 0.3 - 2 km/sec are encountered. In the present context we are less concerned with the details of the eruption process

Figure 4: Growth in separation for Region 11972. The open circles are based on optical observations, and the crosses on the magnetic data. The lower horizontal scale is marked in days. The inset shows a magnified view of the growth during the first day (July, 24, 1972). Its scale is in universal time.

The last measurement is quite uncertain because the region is close to the limb. The curve should, in fact, probably be nearly horizontal after the 27 th .

The separation of the region is about 7.5×10^4 km on the next disk passage.

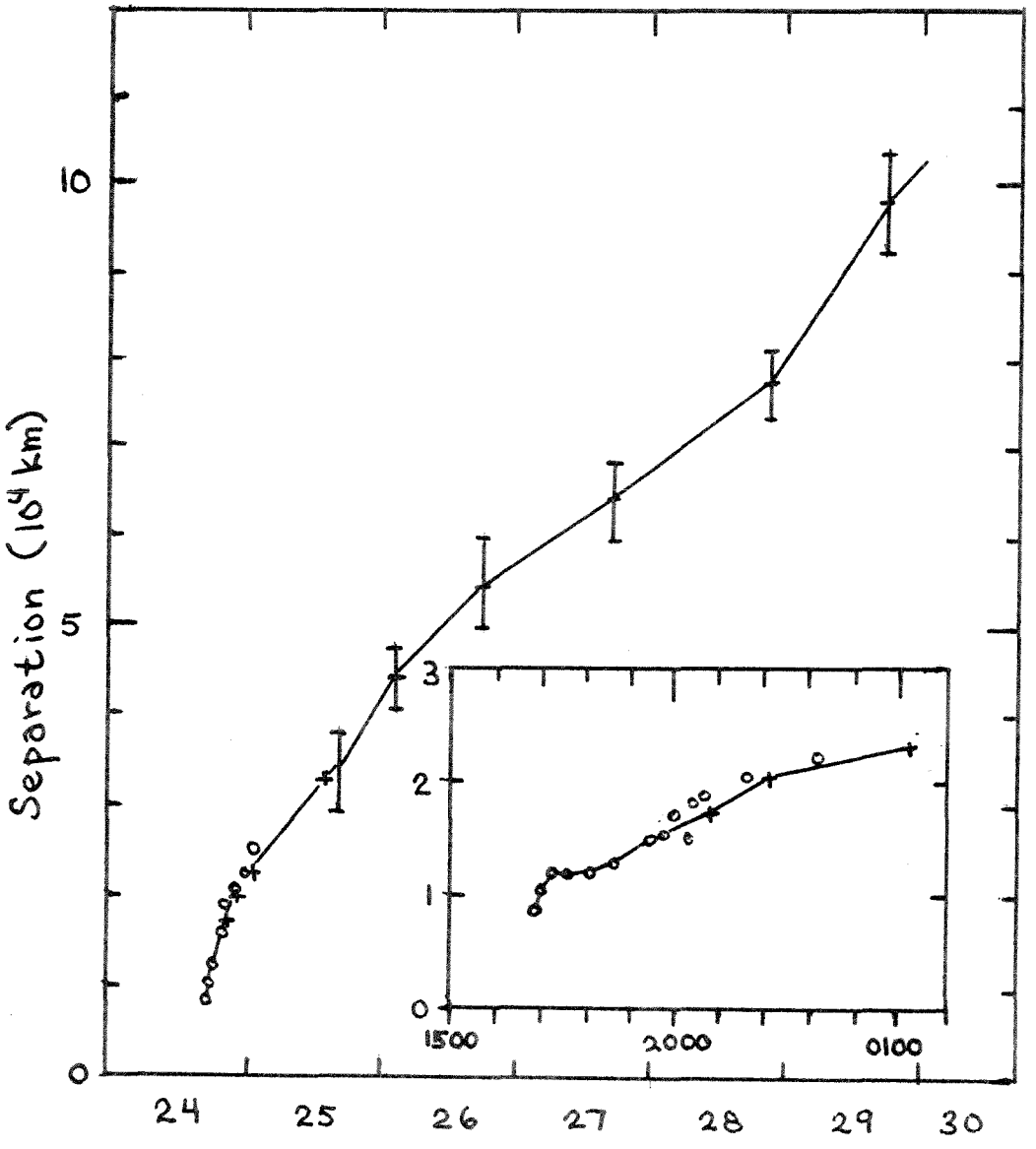


Figure 4

than with the overall picture -- in particular the flux and separation at the end of the growth phase.

Figure 4 indicates the growth in the separation between the centers of weight of the p- and f-plages over the first five days. As was noted by Harvey and Martin (1973) in connection with the smaller 'ephemeral' regions, the rate of expansion is most rapid at the outset and appears to decrease continuously throughout the life of the group. A growth rate of 0.5 km/sec is observed during the first 10 hours. 0.2 km/sec is typical of the next few days. An expansion rate of several kilometers per second could easily occur during the first hour or so since the footpoints are already separated by ~10,000 km when they first become prominent enough to measure. Results for positions close to the limb are rather uncertain. When the region returns, the separation is about 75,000 km.

2.3 Comparison of Results With the Development of Other Active Regions

The growth pattern of several other randomly selected active regions was studied by means of the NOAA bulletins, using particularly the years 1974-1976, when magnetic data from both Kitt Peak and Mount Wilson were available. The results suggest that Region 11972 is rather typical.

2.3.1 Separation

Figure 5 presents measurements of the separation between

Figure 5: Growth in magnetic separation of selected regions during the first few days of their lives. The curves labelled with numbers are regions appearing during 1972, which were included in a series of surveys conducted with the Big Bear videomagnetograph, and do not, necessarily, appear in Appendix II . Isolated crosses indicate single measurements made at the first central meridian passage for regions whose birth had been observed some number of days before.

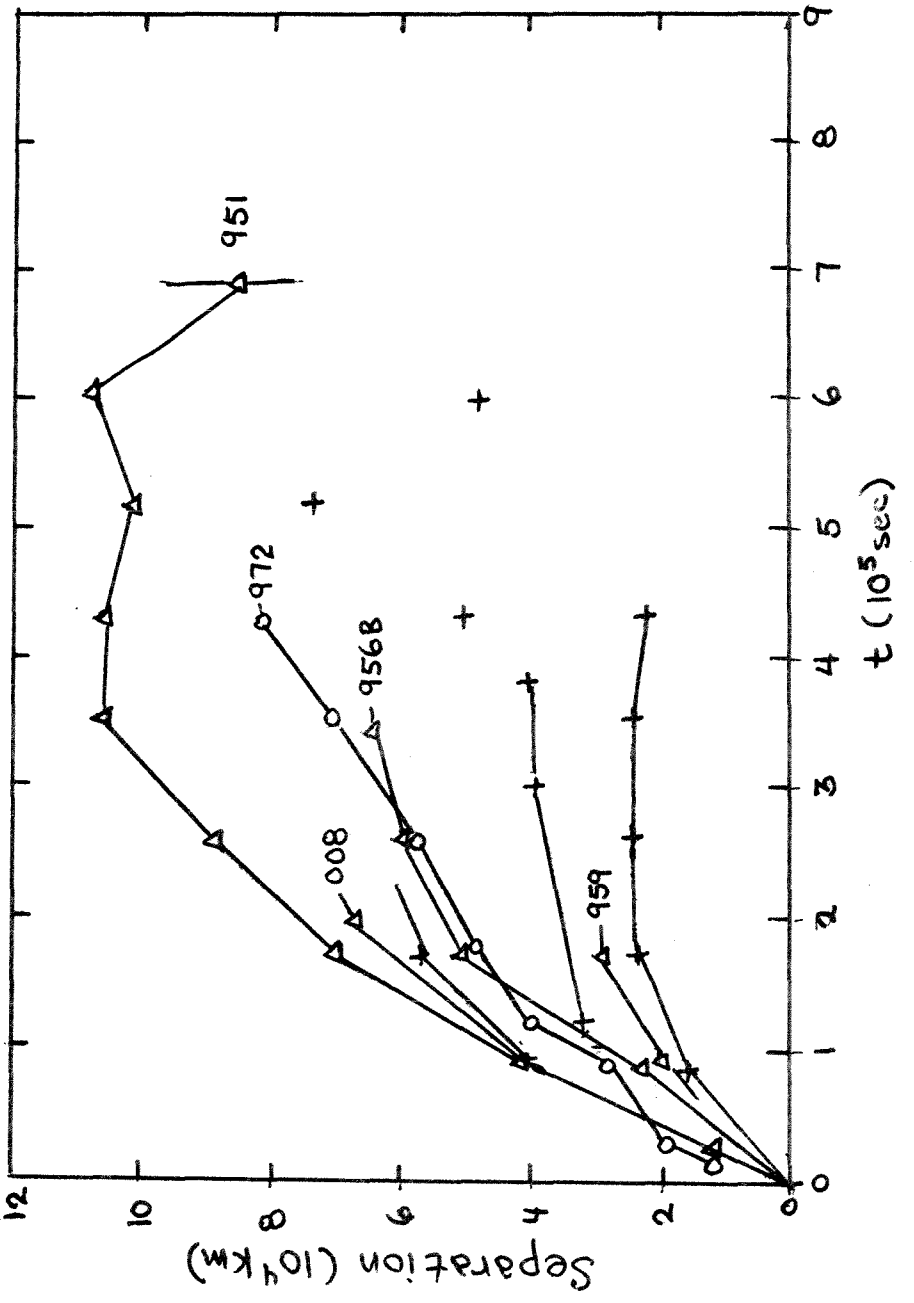

951
972
9568
008
959
Separation (10^4 km)
t (10^5 sec)
0
2
4
6
8
10
12
0
1
2
3
4
5
6
7
8
9

Figure 5

the peaks of the p- and f- polarity distributions as a function of the number of days since the birth of the region (the "birth" is defined as the moment of first appearance in H-alpha and magnetically). In some cases, a single measurement was made at the first central meridian passage. In others, daily measurements were made, and corrected for foreshortening. The graph indicates separation rates of from 0.1 - 0.4 km/sec, starting with an initial separation between 5000 and 10000 km. The initial expansion rate does not seem to provide much of a clue as to the eventual size of the region (in terms of total flux) although the larger ones seem capable of sustaining this rate for a longer period of time. Most of the regions shown appear to reach their "maximum" separation within 2-4 days.

The present results are not incompatible with those of Bumba and Howard (1965a), who studied the semi-major axes of the ellipses of K-line and H-alpha disturbance, and found numbers of 0.1 and 0.2 km/sec, respectively; although they indicate an essentially linear growth over at least 7 days. Nor are they inconsistent with the results of Harvey and Martin (1973) showing ephemeral regions growing to typical "dimensions" of 30,000 km in times of between 6 and 24 hours, provided one realizes that the "major axis" measured by the latter authors is approximately twice the magnetic "separation" referred to above.

2.3.2 Area

Figure 6 presents measurements of active region areas over the same period. These areas should not be confused with the "areas occupied by field" referred to earlier. The present measurements refer, roughly, to the area of a smooth curve enclosing the "bulk" of the field. If a precise definition were required, it would probably correspond to something like the size of the smallest circle containing 90% of the field -- but such a definition is not needed, since in practice the boundary between active region and background fields is sufficiently sharp to make the measurement of such a quantity relatively simple and unambiguous. Indeed, since the display of the weaker fields is always poor, it would require careful reference to the original digital flux data, and a correction for the state of the background field existing before the appearance of the region, to be sure that one were really including 90% -- or any other pre-determined fraction -- of the field in the chosen area; and such a refinement would be considerably beyond the scope of the present study.

As always the measurements refer to one polarity only. The area of the total active region is approximately twice as great.

As indicated in the figure, a considerable range of growth rates is encountered: from as little as 2×10^3 km^2/sec

Figure 6: Total area dominated by each polarity versus time for selected regions during the first few days of their lives. Preceding and following polarities are indicated separately, so that each region contributes two curves. Isolated crosses arise from single measurements at central meridian passage.

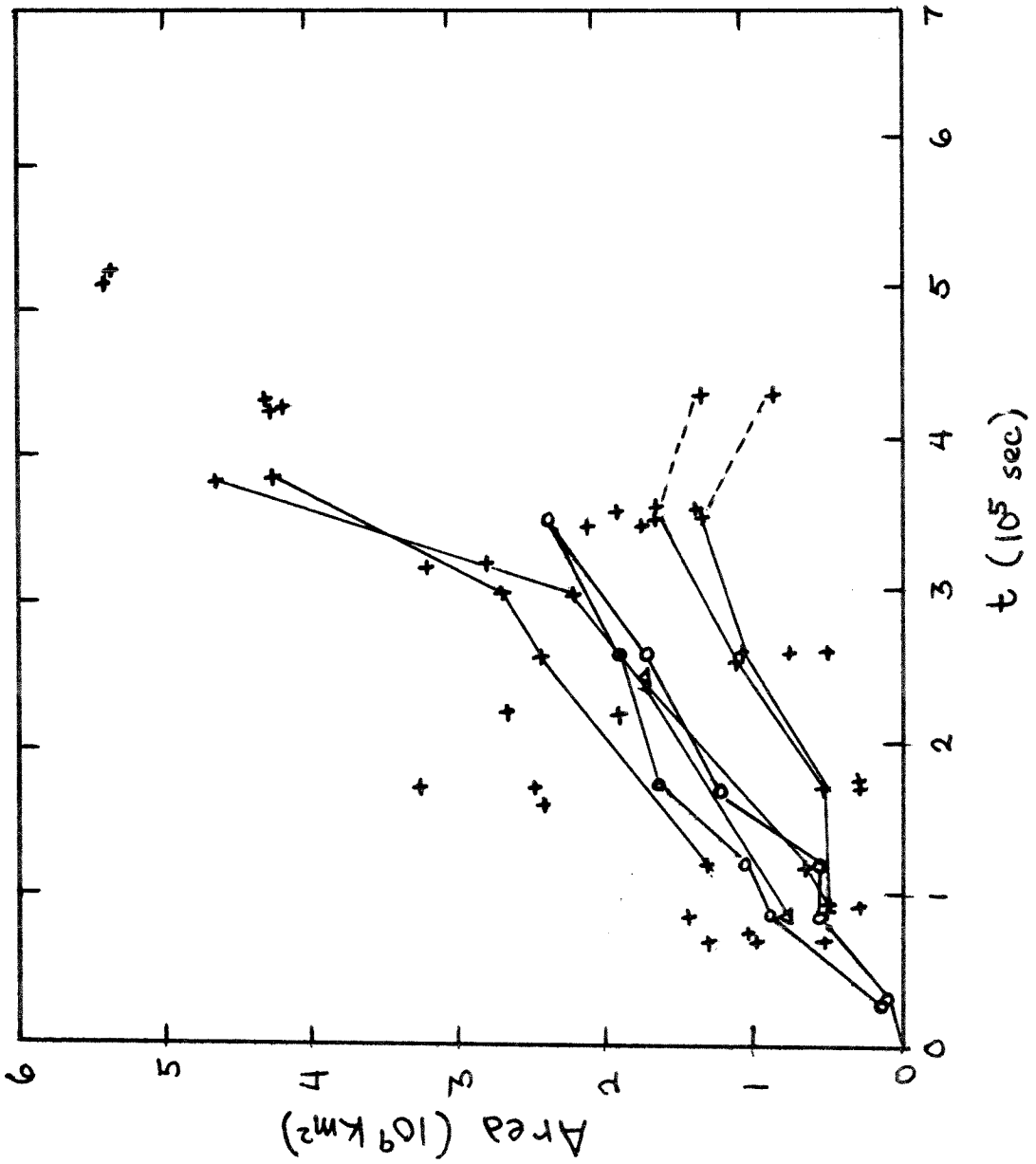


Figure 6

to as much as 2×10^4 km^2/sec. In general, the faster rates seem to be associated with the regions which generate the most flux, but the correlation is far from perfect.

The observations of Bumba and Howard (1965a) regarding the K-line and H-alpha ordering zones give values intermediate to those in the graph. If one assumes that their ellipses are about half as tall as they are wide, then the K-line areas correspond roughly to the lower envelope and the H-alpha areas to the upper envelope of the curves in the figure.

The curves of area versus time show no obvious transition from growth to decay. In particular, they do not exhibit the rapid rise to a plateau value characteristic of separation data.

2.3.3 Flux

Figure 7 gives the measurements for flux (see Appendix I). Since the method is crude, the curves are generally based on a single measurement made at the central meridian. The bulk of the regions seem to be growing at rates between 4×10^{15} and 3×10^{16} Mx/sec. Again, the higher rates tend to apply to the larger regions, and the smaller rates to the smaller regions; and again the correlation is not perfect.

Some of the growth rates undoubtedly appear artificially low because the region may have reached its peak before the time of the first measurement, but some of the low growth rates are also real. McMath Region 13340, for example, grows

Figure 7: Flux versus time for selected regions.
The measurements refer to the flux of a single plage, but in general, results for the two halves of the region have been averaged, so that a single point is given.

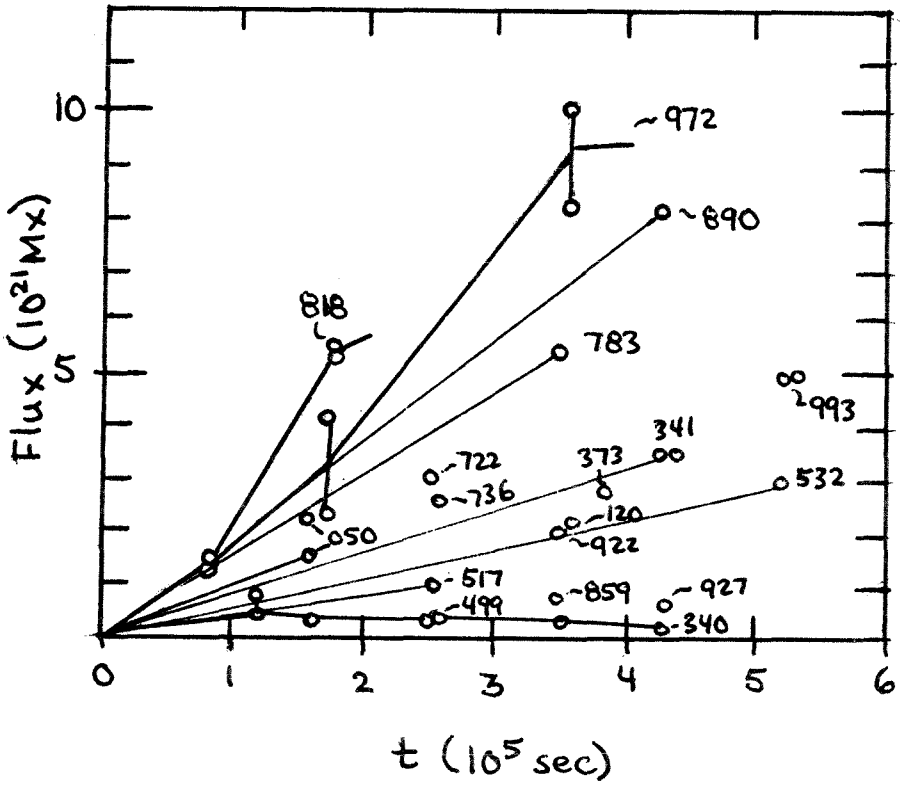

10
5
Flux (10^{21} Mx)
0
1
2
3
4
5
6
t (10^5 sec)
~972
~890
818
783
993
341
373
722
736
532
120
922
050
517
499
859
927
340

Figure 7

steadily over two days, producing small spots ($\sim$10 millionths of a hemisphere in area), but at a rate of only $\sim 1.5 \times 10^{15}$ Mx/sec. Region 13763 grows for two days at less than 10^{15} Mx/sec. On the other hand, Harvey and Martin, present an example of an ephemeral region which adds 10^{20} Mx (at each end) in 6 hours ($\sim 5 \times 10^{15}$ Mx/sec).

At the high end of the scale, there is little evidence for growth rates substantially higher than those shown. The reversed-polarity August flare region of 1972 (Zirin and Tanaka, 1973), for example, attained a maximum source strength of about 4×10^{22} Mx in $\sim$20 days (2.3×10^{16} Mx/sec). The largest "normal" region considered in the present study, McMath 13790, was born on the backside and had reached 3×10^{22} Mx by the time of first central meridian passage. Assuming that its growth took place over 1-3 weeks, a rate in the range $(1.5\text{-}4.4) \times 10^{16}$ Mx/sec would be implied. Similarly, Cowling's (1946) study of the development of recurrent spots indicated growth rates of 2.4 and 3.6×10^{16} Mx/sec over 4 and 10 days, respectively, for the "30" and "50" day varities.

We conclude that a considerable variety of eruption rates is encountered with something like 10^{16} Mx/sec being typical. Larger regions can be produced by higher than normal eruption rates, or by a moderate eruption rate maintained over a longer than normal period of time, or by a combination of the two effects. On average, the larger

regions grow more rapidly, and sustain their growth over a longer period than the smaller regions, but the pattern is not inviolable. Improved measurements might reduce the scatter, but it is unlikely that they would remove it entirely.

2.3.4 Source Flux Versus Sunspot Area

Sheeley (1966), in an investigation of active region magnetic fields, proposed the rule that the peak flux (of each polarity) is directly proportional to the total area of p- and f-spots (umbras plus penumbras) in the region at its point of maximum development; or more precisely:

$$\left(N_o / 10^{21}\ Mx \right) = 1.2 \left(A_M / 10^{18}\ cm^2 \right)$$

where N_o is the source flux and A_m is the spot area.

A similar relationship (Gnevishev, 1938; in Allen, 1973, §88) exists between the area of a spot group and its lifetime:

$$T\ (\text{days}) \approx 0.12\ A_{Max}\ (\text{millionths of a hemisphere})$$

The development time is presumably some more-or-less fixed fraction ($\sim 1/3$) of this total lifetime.

We have not attempted to verify Gnevishev's result, but it is quite easy to check the validity of "Sheeley's rule" using the previously calculated fluxes, since the NOAA bulletins also include a report of the sunspot areas.

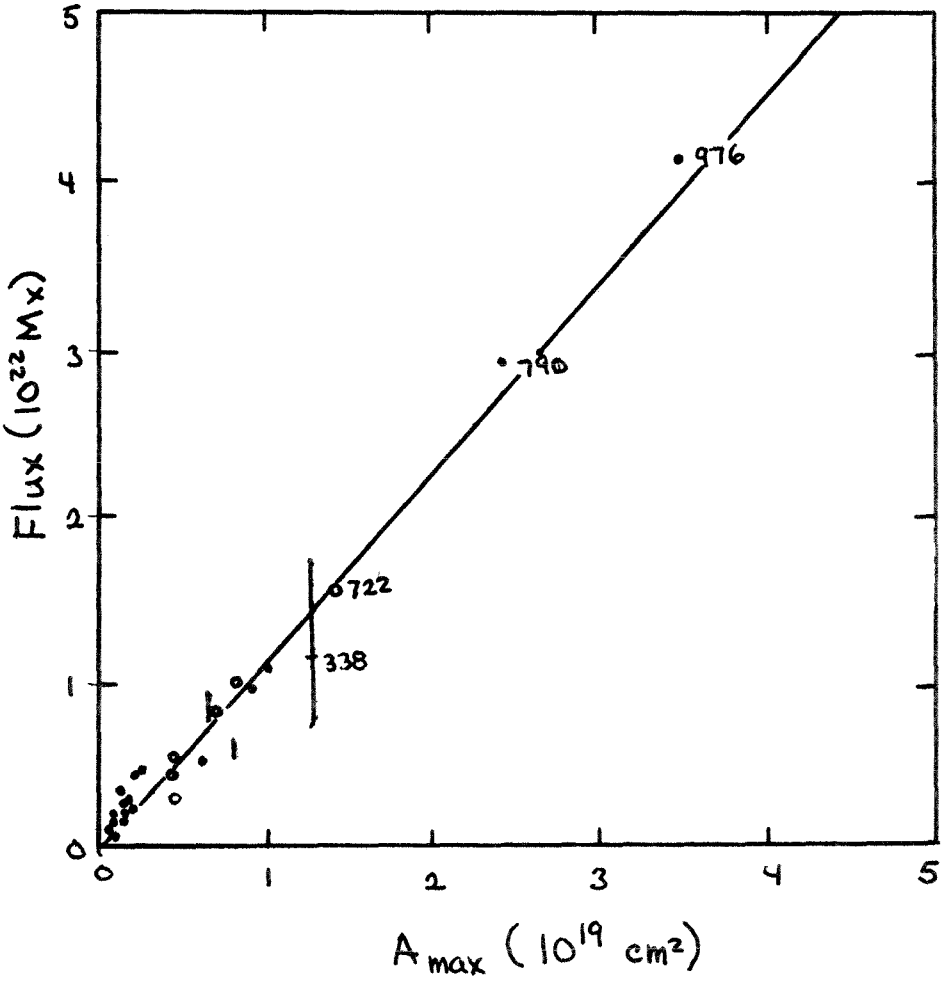


Figure 8: Plot of flux versus spot area for selected regions near the point of maximum development, illustrating "Sheeley's Rule" .

Figure 8 indicates generally excellent agreement (which may be somewhat surprising since the present results, based on the Mount Wilson contour plots, include a factor of 2 correction for line weakening, while Sheeley's, based on the density of photographic cancellations, do not). There are some deviations in the sense of a few regions with $N_o \approx 10^{21}$ Mx that don't quite attain the required maximum sunspot area; but the general trend is certainly there. It should also be noted that the intercept is not precisely at the origin: spotless regions can have fluxes of up to about 10^{20} Mx.

If most spots have central fields of 2000-3000 gauss, then it is easy to show that "Sheeley's rule" corresponds to a situation in which about half the flux is in spots and half in the plage. This agrees well with one's general impression of the relative areas of spots and plage in high-resolution magnetograms of a wide variety of regions at "maximum development".

2.3.5 Magnetic Separation versus Sunspot Area at Maximum Development

One might hope that a similar simple relationship would exist between the magnetic separation (center to center distance for the two polarities) and the spot area at maximum development. Then all the important magnetic parameters describing the growth of a region: flux, separation and development time could, on the average be specified simply

by giving the sunspot area. This would be a great bonus for empirical models of the solar cycle (such as those of Leighton, 1964, 1969), since accurate sunspot data exist over many cycles. The predicted pattern of large scale field could be computed with considerable precision.

Unfortunately, Figure 9 indicates that the relationship between separation and spot area is not very uniform. Some large regions (in terms of flux) are well separated, but others seem to be quite compact. Small regions tend to have smaller separations, but again, the systematic trend tends to be smaller than the scatter for individual examples.

Groups with spot areas $\gtrsim$100 millionths seem to have separations mainly in the range 60,000-80,000km. Smaller regions are mainly in the range 40,000-60,000 km, but the smallest ($\lesssim$ 10 millionths) can be as small as 20,000-30,000 km (or even 10,000 km for a short-lived ephemeral eruption). A clearer trend would no doubt emerge if more examples were included, and also if greater care were exercised in tracing the development of the individual regions, so that one could be certain the separations given truly represented the spacings at the moment of transition from growth to decay.

The Zurich spot classification system (Waldmeier, 1947) includes typical bipolar groups ranging from Classes C and D, with areas $\sim 10^{19}$ cm^2 and separations $\sim$75,000 km, to Class F, with an area $\sim 1.5 \times 10^{20}$ cm^2 and a separation

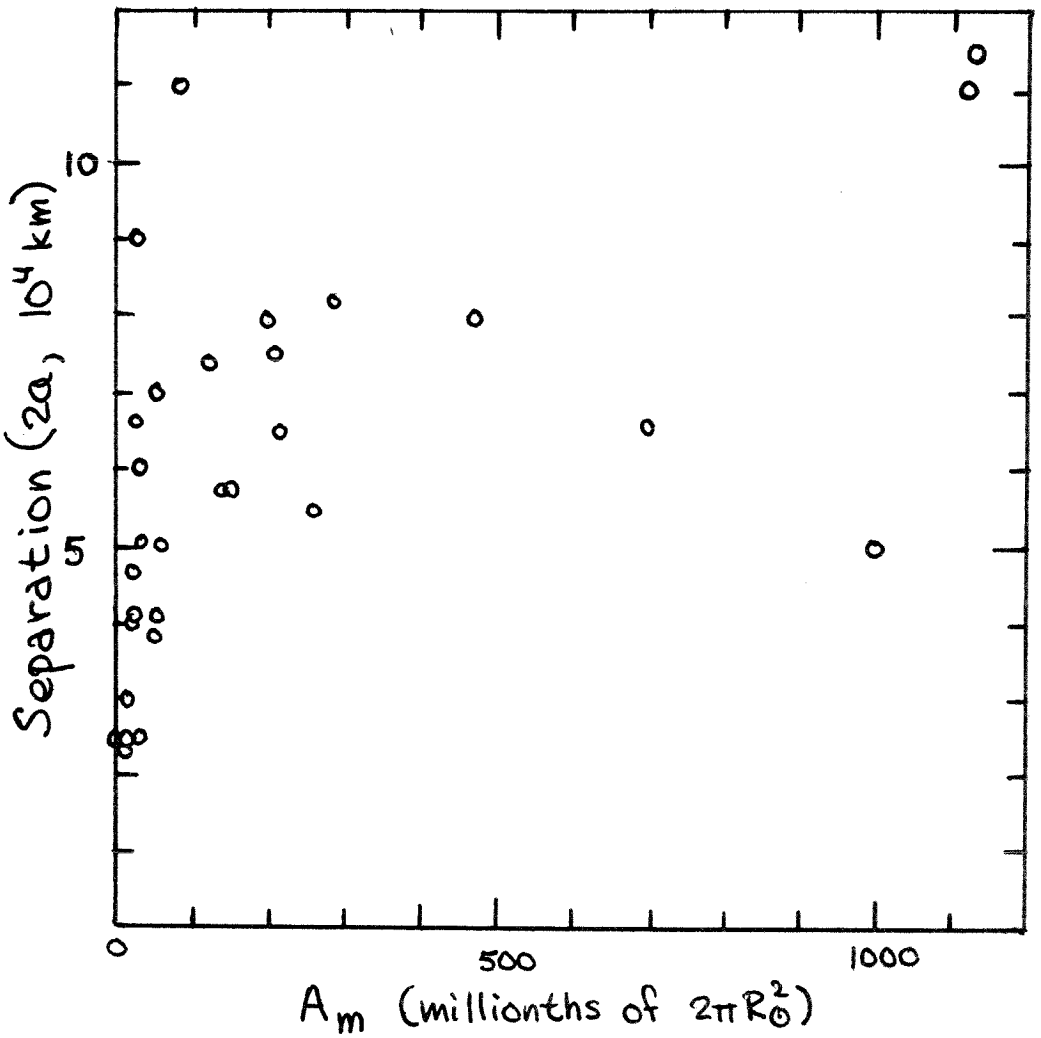


Figure 9: Magnetic separation versus area of spots at maximum development

(1 millionth = 3.04×10^{16} cm^2)

$\sim$180,000 km.

It is worth noting that the sunspot observations (Greenwich, 1925; Waldmeier, 1941) show expansion rates $\sim$.1-.2 km/sec (lasting at most 5-10 days) in good accord with the results of § 2.3.1 . The sunspot observations are somewhat deceptive, however, in that after the rapid rise to maximum separation, they tend to show a slight decrease in separation as the spots dissipate. The magnetic observations invariably show that the average separation between the remnants of the two polarities continues to expand, albeit at a much reduced rate (ephemeral regions may violate this general principle, but their decaying fields are so weak that accurate observation is not possible at the present time).

2.3.6 Summary

Based on the foregoing results, we can state that while individual regions may grow in rather different ways, there is a reasonably simple and consistent average pattern of development. The typical region appears first as two isolated magnetic footpoints, of negligible total flux, with a separation of slightly under 10,000 km. During the next few days it adds flux at $\sim 10^{16}$ Mx/sec, and separates at $\sim$0.2 km/sec. The regions which will eventually exhibit the largest flux tend to have the largest initial growth rates in terms of separation, area and flux, but the correspondence is not

perfect.

The peak flux attained is proportional to the area of the spots, with values ranging from $\sim 2 \times 10^{20}$ - $\sim 4 \times 10^{22}$ Mx being encountered for the regions involved in the present study. 10^{22} Mx is "typical", but there is no definite limit at either end of the scale. The peak separation has a weaker dependence on the spot area, values from 40,000-80,000 km being typical of a fairly wide range of smaller groups. The decay phase presumably sets in immediately after the peak values have been attained. Due to the inadequacy of the present source of data, we have not demonstrated in detail that the onset of the decline in flux is perfectly simultaneous with the decline in area of the spots, but it seems unlikely that it would be otherwise.

When examined at high resolution, the growth phase is seen to consist of a series of discrete sub-eruptions. The fields contributed by each individual sub-eruption merge to form the final configuration. Spots seem to form by a local 'condensation' of the erupted fields.

There is some hint in the NOAA bulletins that the eruption of the largest regions may be preceeded by a general "disturbance", with a number of tentative, short-lived bipolar eruptions (see, for example, the area of McMath 13722 on June 13 and 14, 1975); but it is not entirely clear that this is really related to the eventual eruption of the region.

2.4 Inclinations

There is considerable interest in the inclination of the magnetic axes of regions relative to lines of constant latitude, since, as pointed out by Leighton (1964), it is only the north-south component which, in the long run, contributes to the formation of large-scale fields; and the classic results are somewhat contradictory. Everyone agrees that the groups are arranged, on the average, with the preceding end closest to the equator, but not everyone agrees by how much.

Joy (in Hale et al., 1919) and Brunner (1930), basing their conclusions on extensive studies of drawings of "sunspot streams", found an average inclination of about 6° (with some dependence on latitude). Butler (1922), basing his conclusions on an equally extensive study of K-line spectroheliograms found 15°.

The present results are shown in Figure 10. While a conclusion based on thirty-one examples obviously cannot have the validity of one based on thousands, the average magnetic inclination of $6.3^{\circ} \pm 3.6^{\circ}$ certainly seems to be much closer to Joy and Brunner than to Butler. The discrepancy could arise either because of a difference in the age of the regions studied, or because of a peculiarity in the part of the cycle used, but a plot of inclinations on successive rotations (not shown) indicates no obvious trend

Figure 10: Inclination of the magnetic axis of selected regions with respect to lines of constant latitude.

(a) Tilt versus flux at maximum development (logarithmic scale).

(b) Tilt versus latitude.

A positive inclination is one in which the preceding plage is closest to the equator.

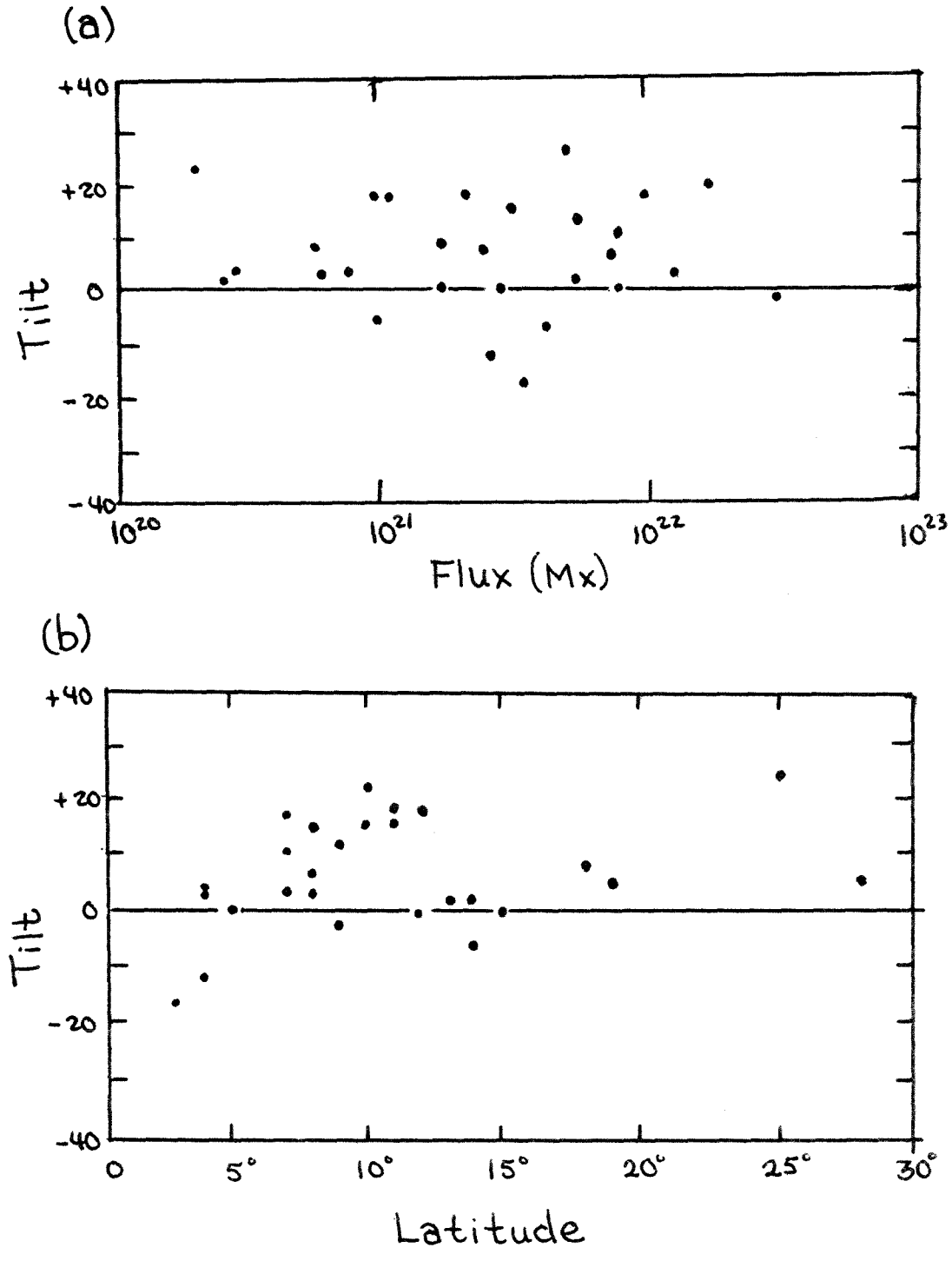

(a)
+40
+20
0
-20
-40
Tilt
10^{20}
10^{21}
10^{22}
10^{23}
Flux (Mx)
(b)
+40
+20
0
-20
-40
Tilt
0
5°
10°
15°
20°
25°
30°
Latitude

Figure 10

for the average inclination to either increase or decrease with age; and the former authors found only slight variations in the average angle from year to year. We conclude that the appearance of the calcium plage must have some tendency to exaggerate the actual inclination of the magnetic axis.

2.5 Adoption of a Standard Case

For purposes of the diffusion problem to be considered in Section 3, it is desirable to establish a set of initial conditions corresponding to an average or "typical" active region (in the strictest sense there is no real 'average' region, since an indeterminate number of ephemeral eruptions would have to be included in the average).

From the preceding discussion we conclude that the typical region whose decay we are going to be studying could be represented by the injection of 10^{22} Mx (of each polarity) with a separation of 60,000 km and an inclination of 6°. A typical latitude of eruption is $10\text{-}15^{\circ}$ during the period studied.

These initial conditions can be generalized by assuming a flux proportional to the maximum sunspot area according to "Sheeley's rule"; leaving the separation and inclination essentially unchanged. The present case corresponds to a region with a sunspot area of about 250 millionths (of a solar hemisphere), developing over $\sim$4 days.

2.6 Interpretation of Observations of Emerging Flux

The mechanisms underlying the development and eruption of solar fields are not particularly well understood. The size of the eruptions suggests an association with the supergranulation, but, at least according to Glackin (1975) there is no persistent relationship between the point of emergence and the form of the network. The size and strength of the flux ropes is presumably determined by couplings rather deep in the convective zone.

Bouyancy and/or convection may be important in getting that field to the surface, but an order of magnitude calculation suggests that the generally assumed bouyancy force (Parker, 1955) would be sufficient to pop a sunspot-sized flux rope through the surface in a matter of a few minutes.

If we picture the erupting segment of field as being a cylinder of radius r and length L (Figure 11) then the bouyant force, obtained by assuming that the magnetic pressure inside the tube is compensated for by a reduced density, would be:

$$F_{bouyant} = (\rho_{out} - \rho_{in}) \cdot vol \cdot g \approx \frac{B^2}{4\pi \langle v^2 \rangle} \cdot \pi r^2 \cdot L \cdot g$$

where $\langle v^2 \rangle$ represents the mean-squared velocity (thermal plus turbulent) which generates the normal fluid pressure (assumed to be the same inside and out).

The bouyant force is resisted only by the tension in

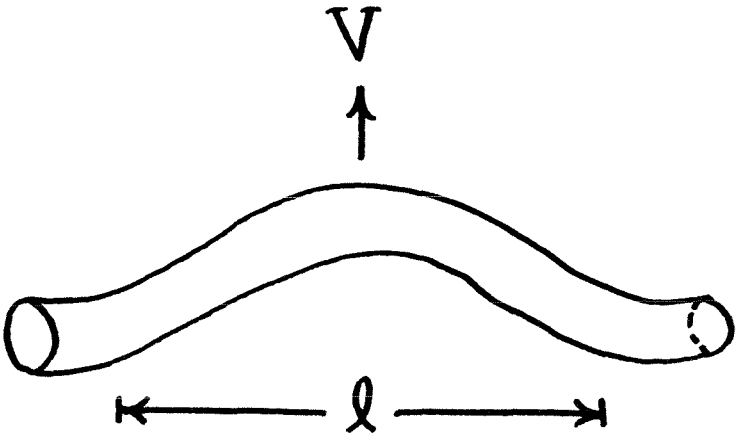


Figure 11: Illustration of bouyancy calculation

the field lines and by the drag due to the fluid streaming around the rising tube. We can write this, approximately, in the form:

$$\frac{B^2}{4\pi\langle v^2\rangle}\cdot\pi r^2 Lg \approx \rho V^2 2rL + 2\left(\frac{B^2}{4\pi}\right)\pi r^2$$

where the first term on the right is the drag, and the second is the tension. V is the velocity of ascent.

Now if we assume:

$$L \approx 30{,}000 \text{ km}.$$
$$v \approx 12 \text{ km/sec} \quad (T = 6000^\circ K)$$
$$g = 0.274 \text{ km/sec}^2$$

then it is easy to see that the bouyant force far outweighs the maximum possible restraint due to tension in the field lines. Ignoring that term, we find a terminal ascent velocity of:

$$V \approx \left(\frac{B^2}{8\rho\langle v^2\rangle}\right)^{1/2}\sqrt{rg}$$

Now, it is probably safe to assume that $(B^2/8\rho\langle v^2\rangle) \geq 1$, so that for a tube of sunspot-like width, $r \approx 3000$ km:

$$V \geq \sqrt{rg} \approx 30 \text{ km/sec}$$

At such a velocity, the entire tube would emerge in a little over 3 minutes. To obtain a more reasonable time (on the order of a few days) one would have to assume either that the rope is highly fragmented (so that it would present a large drag surface with a minimum of bouyancy), or else that the individual "strands" corresponding to the sub-

eruptions are released gradually, over the required period by some, as-yet-unexplained mechanism. Of course, the bouyancy calculation may also be wrong. Perhaps the magnetic pressure inside the tube is offset by a slight deficit of turbulence.

Intuitively, one would expect that other factors tending to limit the rate of flux eruption might include the time required to drain the photospheric material from the rising loops, and the time required to erode away the constricting currents to the point where they can conform to the 'vacuum' condition of the corona.

Roberts (1970) has calculated the draining time for a flux tube, and finds $\sim$1/2 hour. This seems to be related to the visibility lifetime of arch filaments, but, since the result is not particularly sensitive to the assumed cross-section of the tube, could not be considered helpful in explaining the eruption rate of the region as a whole.

From a different point of view, Leighton (1969) has proposed an empirical model of the solar cycle in which a characteristic eruption time τ appears. This constant plays a dual role: representing both the eruption time for a single region, and (what in the smoothed model is the same thing) the time between eruptions in a given sunspot-wide zone of latitude. The value of τ is adjusted to make the period of the solar cycle come out right. Values of about 6 months are

required, which seems long indeed compared to the observed eruption time, even if one assumes that a considerable fraction of the process is taking place invisibly beneath the surface.

It is, nonetheless, interesting to note that one <u>can</u>, roughly, reproduce the statistics of sunspot eruptions by assuming the active latitudes to be occupied by a series of sunspot-wide (say 1°) active "bands" which are continually in the process of producing an eruption at one longitude or another. Thus, for example, in an average "active" year one might have 30 bands (occupying $\sim 15^{\circ}$ in each hemisphere). If each band produced a region every two weeks, then there would be something like two new regions per day, which is about what is seen (Weart, 1972; Glackin, 1973). Years of high activity would require more active bands, and years of low activity fewer.

It is not entirely clear what would cause an active band to erupt for two weeks at one longitude (we are assuming that the initial stages of the eruption are not seen in the photosphere), and then pack up and move elsewhere, but the following suggestion may be helpful. In §6.3 we will find that the predicted rate of flux annihilation in a region at its point of maximum development, (when the two polarities are first starting to really diffuse into each other) is comparable to (but slightly slower than) the observed rate

of eruption ($\sim 10^{16}$ Mx/sec). Since there is no a priori reason to expect such a coincidence, it leads one to suspect that the development of a configuration favorable to the recombination of fields may play a role in arresting the instability which produced the eruption in the first place. Given the observed initial separations, the diffusion model would explain why times on the order of a week are required to develop such a configuration.

3. The Decay Phase

3.1 Theoretical Considerations in Field Dispersal

The fields which erupt onto the solar surface find themselves embedded in a conducting plasma, which, for the most part, tends to resist any sudden change in either their strength or configuration. While the exact coupling between the field and the plasma is no doubt very complicated, it is thought that, to a reasonable degree of approximation, their interaction can be expressed by means of a "simplified magnetohydrodynamic equation" of the form:

$$\frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \times (\vec{v} \times \vec{B}) + \frac{1}{4\pi\sigma} \nabla^2 \vec{B} \qquad (3.1)$$

(see, for example, Cowling, 1953) where σ is the conductivity (in electromagnetic units). The equation is a simple consequence of Maxwell's equations, under the assumption that the conductivity is uniform and isotropic, and that "displacement currents" (due to changing electric fields) can be neglected. Since the very presence of a field (as in spots and faculae) obviously modifies the conditions of the plasma, the assumption of uniform conductivity is suspect. Nonetheless, equation (3.1) defines a kind of nominal behavior against which the actual observations can be compared.

The significance of the two terms in equation (3.1) can best be appreciated by considering them separately. When the conductivity is high we have:

$$\frac{\partial \vec{B}}{\partial t} \approx \vec{\nabla} \times (\vec{v} \times \vec{B}) \tag{3.1a}$$

It can be demonstrated that this corresponds to a situation in which the flux through any loop moving with the fluid is constant. Such a condition can be most easily satisfied if one pictures the field as being composed of a fixed number of "field lines" which are transported, bodily, by the motions of the fluid; and, hence, it is referred to as the "frozen-in" condition.

At the surface, the field lines are primarily directed in the radial direction, and the movement of the lines in response to turbulent horizontal motions present in the photosphere can be thought of as a two-dimensional random walk (Leighton, 1964). In an averaged sense, the dispersal of the field imposed by the frozen-in condition can then be expressed by a simple diffusion equation of the form:

$$\frac{\partial n}{\partial t} = D \nabla^2 n \tag{3.2}$$

where n is the number of field lines per unit area (i.e., the flux density or field strength). The diffusion constant, D, has the dimensions of a change in area per unit time. In the case of a random walk it would be on the order of the

mean-squared step length divided by the step time. Taking $L \sim 15{,}000$ km and $\tau \sim 20$ hours, for the supergranulation; or $L \sim 600$ km and $\tau \sim 8$ minutes for the granulation, we would estimate:

$$D_{convective} \sim L^2/\tau \sim (1-3) \times 10^3 \ km^2/sec \ . \tag{3.3}$$

The second term in equation (3.1) dominates when the conductivity is low:

$$\frac{\partial \vec{B}}{\partial t} \simeq \frac{1}{4\pi\sigma} \nabla^2 \vec{B} \ . \tag{3.1b}$$

This term describes the effect of Ohmic losses. In other words, if one were to imagine a sunspot-like field, with lines confined to a cylindrical tube, the field would have to be supported by a current flowing in a solenoidal fashion at the surface. The current gradually decays because of the finite resistivity of the medium, but as it decays the field strength inside the tube decreases. This, in turn, sets up eddy currents in the surrounding plasma, which try to resist the change, but are only partially effective. As a result, the field fuzzes out with the magnetic energy going into heat.

Equation (3.1b) states explicity that this is a diffusion-like process, which again can be represented by a fixed number of field lines (preserving $\int B \cdot dA$ through the surface). Flux 'disappears' only where oppositely directed field lines

happen to overlap. Assuming a surface conductivity of $\sigma \approx 2\times10^{-8}$ emu, the Ohmic diffusion constant would be given by:

$$D_{ohmic} = \frac{1}{4\pi\sigma} \approx 4\times10^{-4}\ km^2/sec . \tag{3.4}$$

Evidently the expected transport of field lines due to convective motions far outweighs the fuzzing-out of field concentrations due to Ohmic losses (it is the neglect of the convective term which leads to statements that spots should last hundreds or thousands of years).

In summary, we find that the solar magnetic fields can, to first order, be described by a fixed number of lines being convected, deformed and transported by fluid motions. The eruption of a new region could be thought of as an organized convective process; the dispersal of an old one as a random convective process. The most important factor left out of the present discussion is probably the possibility that the lines might 'resist' the fluid motions, and simply allow the plasma to flow <u>around</u> them (which would not violate the frozen-in condition). This possibility will be considered further in § 5.2 .

Leighton (1964) has already demonstrated that the diffusional interpretation of active region decay is qualitatively in agreement with the observed pattern of spreading, and quantitatively as well, at least as far as order of

magnitude. The data available at the time did not really permit any closer quantitative comparison. The data available now do, and in the following sections we wish to determine just how far and how closely that comparison might be pursued. We will find a generally excellent agreement between the predictions of the diffusion model and average observed pattern of decay; but it is not entirely clear that these predictions could be distinguished from those characterizing a decay process of quite different origin. To demonstrate that the apparent diffusion is actually caused by a 'random walk', one would also have to be able to show first, that the observed daily motions of individual features are compatible with the required overall diffusion constant, and, second, that those motions could be produced by the available fluid motions. That will be the subject of §§ 4&5

For the present, it becomes necessary to define more explicitly the predictions of the diffusion model, and then to attempt to describe the pattern of decay which is actually observed.

3.1.1 The Diffusion Equation and the Definition of the Diffusion Constant

The diffusion equation, in its simplest form is that which we have enountered in the previous section:

$$\frac{\partial n}{\partial t} = D\nabla^2 n \qquad (3.5)$$

where n is a density of "particles" per unit area (in two dimensions). In the present context the "particles" are flux tubes (which can be assumed to be of some quantized strength -- say 3×10^{18} mx), so that the density becomes the local field strength and $\int n dA$ the total flux.

The diffusion equation is a mathematical way of expressing a process of smoothing: the function changes most rapidly (with time) in the places where it varies most rapidly (in space) -- and changes in such a way as to iron out the bumps.

The fundamental solution is that representing the release of a point source of flux, N_o, at $t=0$:

$$n(x,y) = \frac{N_o}{4\pi D t} e^{-\frac{(x^2+y^2)}{4Dt}} \qquad (3.6)$$

-- a Gaussian whose width is increasing with time. (Since the sun is in fact spherical, rather than flat, the final condition should actually correspond to one in which the particles are spread uniformly over the surface at a limiting density of $N_o / 4\pi R_\odot^2$; but in the present investigation we will not be employing time scales over which this effect becomes significant).

If thought of in terms of a probability distribution, then the probability of finding a particle within a distance $r \to r + dr$ of the origin is given by:

$$P(r)dr = \frac{n(r)\, 2\pi r dr}{N_o} = \frac{r e^{-\frac{r^2}{4Dt}}}{2Dt} dr \qquad (3.7)$$

The mean-squared distance increases linearly with time:

$$\langle r^2 \rangle = 4Dt \tag{3.8}$$

The area within which "two-thirds" (1 - 1/e) of the particles are confined -- which is also of interest -- is essentially the same except for an extra factor of π :

$$A_{2/3} = 4\pi Dt \tag{3.9}$$

In modelling the solar fields, it will be noted that the flux is not actually released all at once. In fact, in § 2 we found that new flux is being added even as the old is diffusing away. Thus we may wish to consider the effect which it would have if the flux were not released all in a single packet of size N_o , but rather over a time τ at a rate:

$$\frac{dF}{dt} = \frac{N_o}{\tau} \quad \text{for } 0 \le t \le \tau \tag{3.10}$$

The form of the resulting field configuration can be found by thinking of the release as consisting of a series of infinitesimal packets whose solutions can be superposed:

$$n(r,t) = \int_{\theta=0}^{\tau} \frac{1}{4\pi D(t-\theta)} \cdot e^{-\frac{r^2}{4D(t-\theta)}} \cdot \frac{N_o}{\tau} \, d\theta \tag{3.11}$$

At the end of the "timed release" the density is given by:

$$n(r, t=\tau) = \frac{N_o}{4\pi D\tau} E_1\left(r^2/4Dt\right) \tag{3.12}$$

Figure 12: Comparison of field distributions resulting from "normal" (instantaneous) and "timed" release of flux. The graph indicates a cross section along the x-axis. In the first case 10^{22} Mx is assumed to have been released instantaneously at $t = 0$, and allowed to diffuse at $D = 400$ km^2/sec. In the second case, the same amount of flux is injected continuously over $\tau = 4 \times 10^5$ sec. The graph is at $t = 4 \times 10^5$ sec.

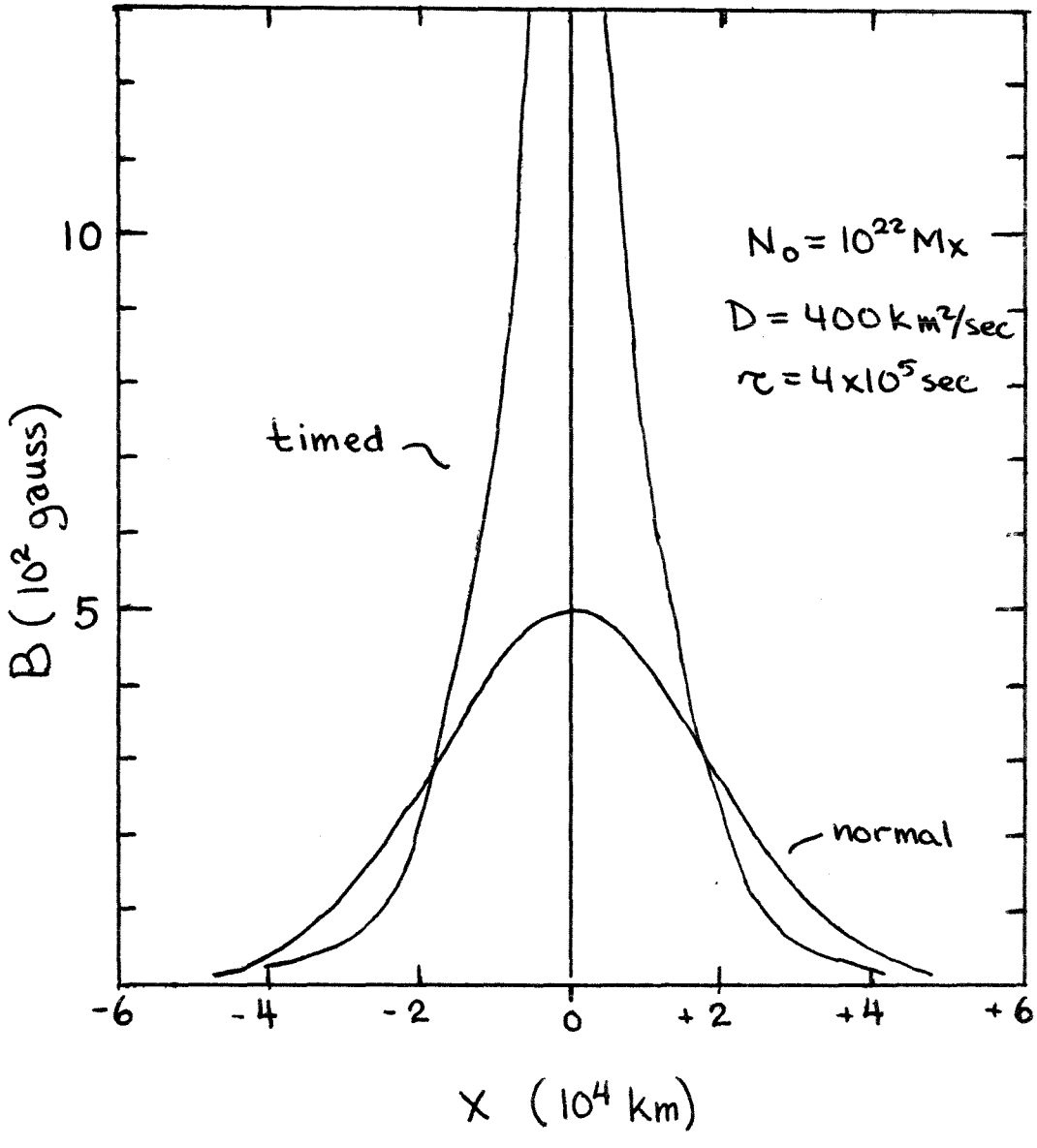

10
5
B (10^2 gauss)
$N_o = 10^{22}$ Mx
D = 400 km²/sec
τ = 4 x 10^5 sec
timed
normal
-6
-4
-2
0
+2
+4
+6
X (10^4 km)

Figure 12

where E_1 is the exponential integral (Abramowitz and Stegun, 1965; § 5.1.1).

Figure 12 indicates how the difference between this and the standard solution is mainly in a higher central density. This discrepancy damps out quickly for $t > \tau$.

3.1.2 Bipolar Solutions on a Flat, Non-shearing Surface

3.1.2a Discrete Point Sources

The simplest solution of the diffusion equation which might be hoped to approximate the dispersal of active region fields is that representing the simultaneous release of N_o particles of positive polarity at $+a$, and N_o particles of negative polarity at $-a$, along the x-axis. The form of this solution can be found by superposing two point-source solutions of the form given in equation (3.6). The result is:

$$n(x,y,t) = \frac{N_o}{2\pi Dt} e^{-\left(\frac{x^2+y^2+a^2}{4Dt}\right)} \sinh\left(\frac{xa}{2Dt}\right). \tag{3.13}$$

If N_o is identified as the source strength, in maxwells, and 2a as the initial separation (in centimeters), then $n(x,y,t)$ will represent the predicted field strength, in gauss. Of course, on the sun we expect the diffusing field to be concentrated by the supergranulation into a few patches of relatively high field strength, at the expense of larger

Figure 13: Typical solution to the diffusion equation with discrete bipolar sources.

(a) Isogauss contours for the standard $N_o = 10^{22}$ Mx and $a = 3 \times 10^4$ km region when $a^2/Dt = 1$ (i.e., an age of 52 days if $D = 200$ km^2/sec or 26 days if $D = 400$ km^2/sec).

(b) Cross section along the x-axis for the same. The dashed line indicates the predicted profile for an equivalent "doublet" source (§3.1.2b)

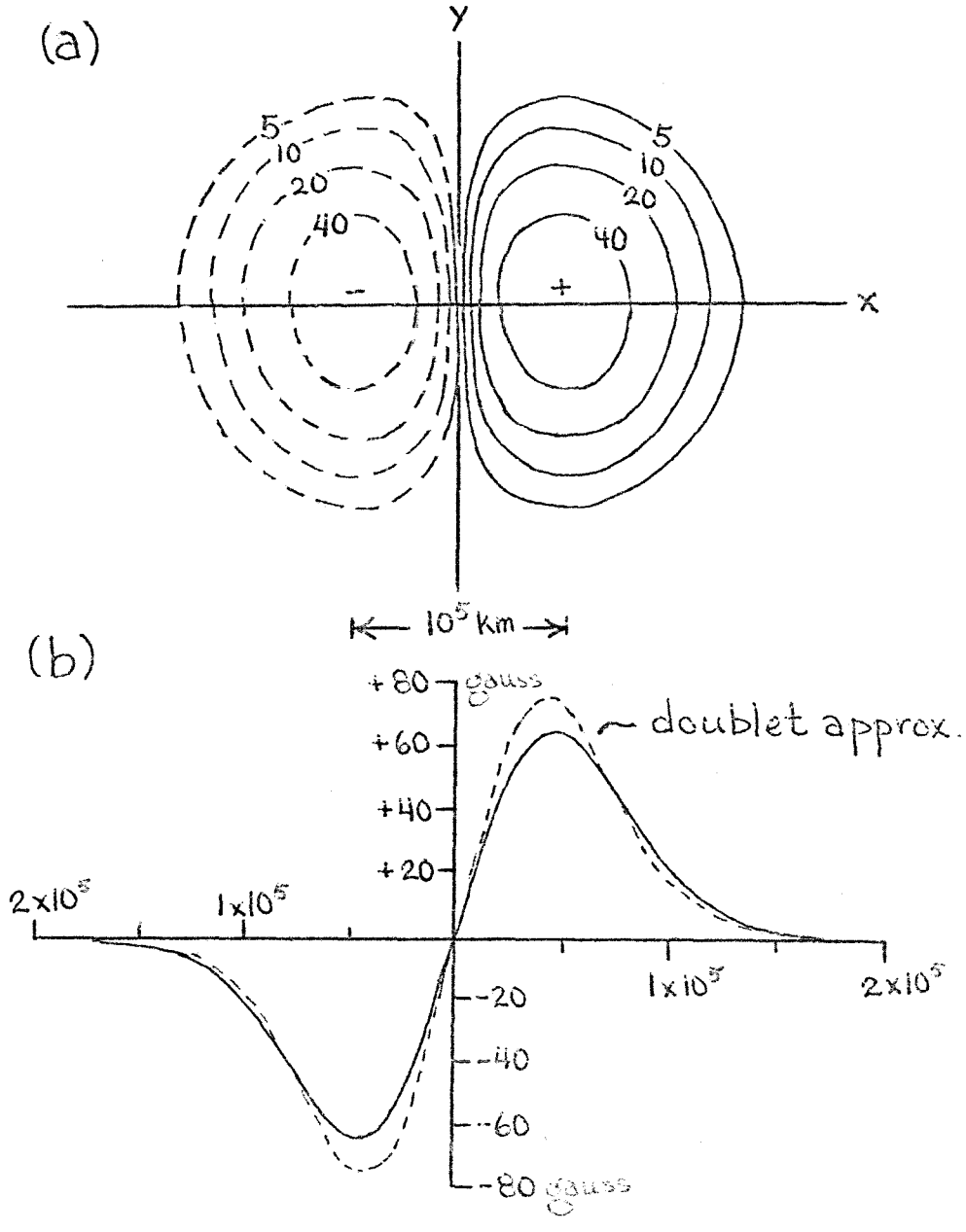

(a)
Y
X
5
10
20
40
−
+
10^5 Km
(b)
+80 gauss
+60
+40
+20
2×10^5
1×10^5
1×10^5
2×10^5
-20
-40
-60
-80 gauss
~ doublet approx.

Figure 13

voids with little or no flux: $n(x,y,t)$ should be thought of as the field strength averaged over several supergranules.

The form of the bipolar solution is illustrated in Figure 13; while Figures 14-16 indicate how the flux, separation and area evolve with time.

Explicitly, the net surviving flux (of one sign) -- obtained by integrating the predicted density over the half-plane $x \geq 0$ or $x \leq 0$ -- is given by:

$$N_+ = N_0 \, erf\left(\frac{a}{\sqrt{4Dt}}\right) \qquad (3.14)$$

where *erf* is the error function as defined in Abramowitz and Stegun (1965; § 7.1.1).

The point, $x_{max}^{(t)}$, of maximum field intensity (along the x-axis) can be found numerically by solving the equation:

$$\frac{a}{x_{max}} = \tanh\left(\frac{a\, x_{max}}{2Dt}\right) \qquad (3.15)$$

The separation, S, is given by:

$$s(t) = 2\, x_{max} \qquad (3.16)$$

and the maximum field strength by:

$$n_{max} = n(x = x_{max}, y = 0, t) \qquad (3.17)$$

The area inside contours of a particular field strength, or, alternatively, the area within which some fraction -- say two-thirds or 90% -- of flux is contained, could be evaluated numerically, by adding up the number of bins, on a

Figures 14-16: The predicted variations in flux, separation, and area with time for discrete bipolar sources.

The dashed lines indicate the corrections due to differential rotation. Since the exact correction depends on the the assumed tilt, separation, diffusion constant, and rotation law, these should, in general, only be regarded as schematic. The curves given here are based on Leighton's (1964) calculation for a source with $a = 6.66 \times 10^4$ km, having a preceding plage at 23.6° and a following plage at 24.8° latitude. He assumed $D = 1535$ km^2/sec ($T_o = 10$ years), and used Newton and Nunn's rotation law.

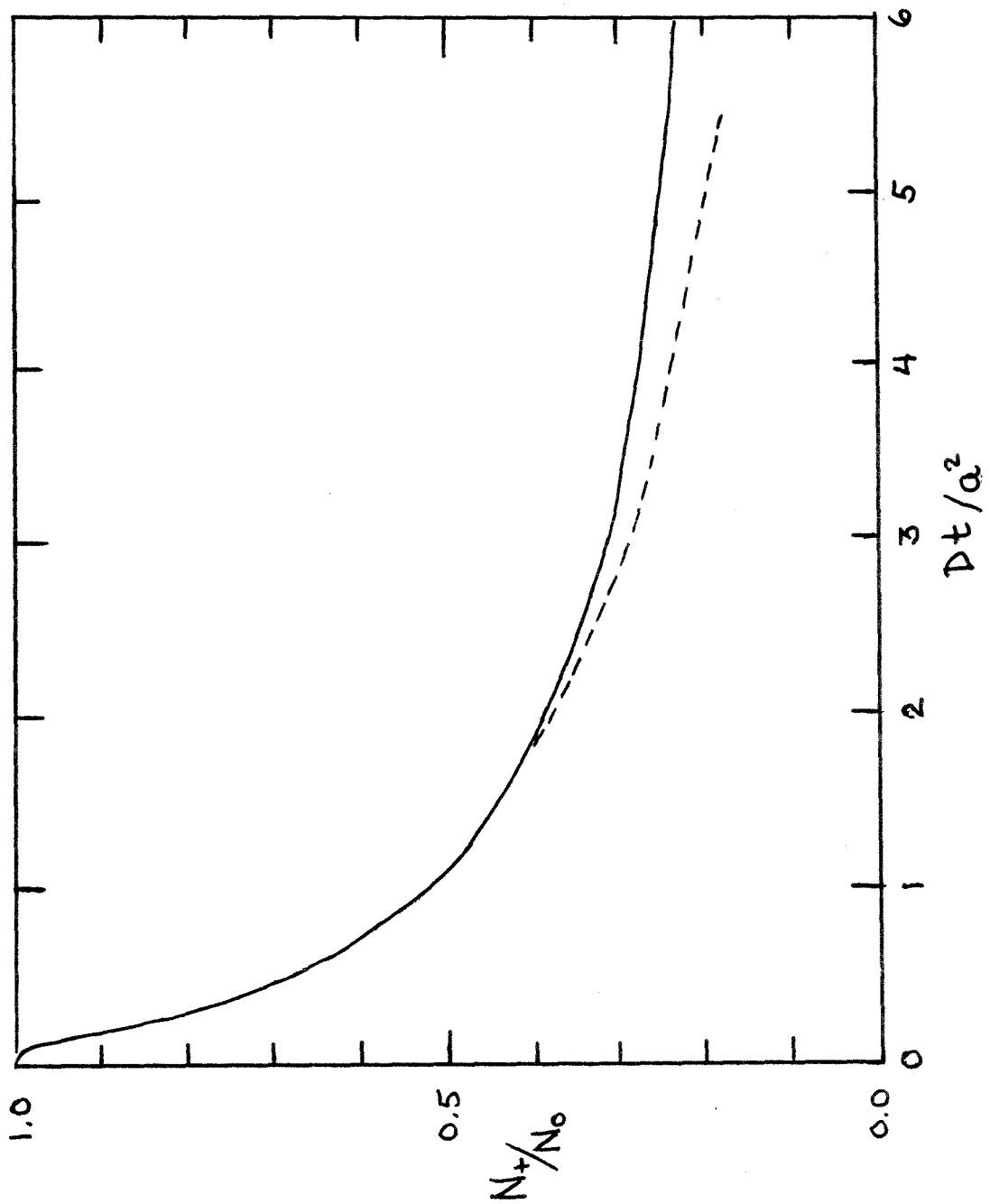


Figure 14: Flux vs Time

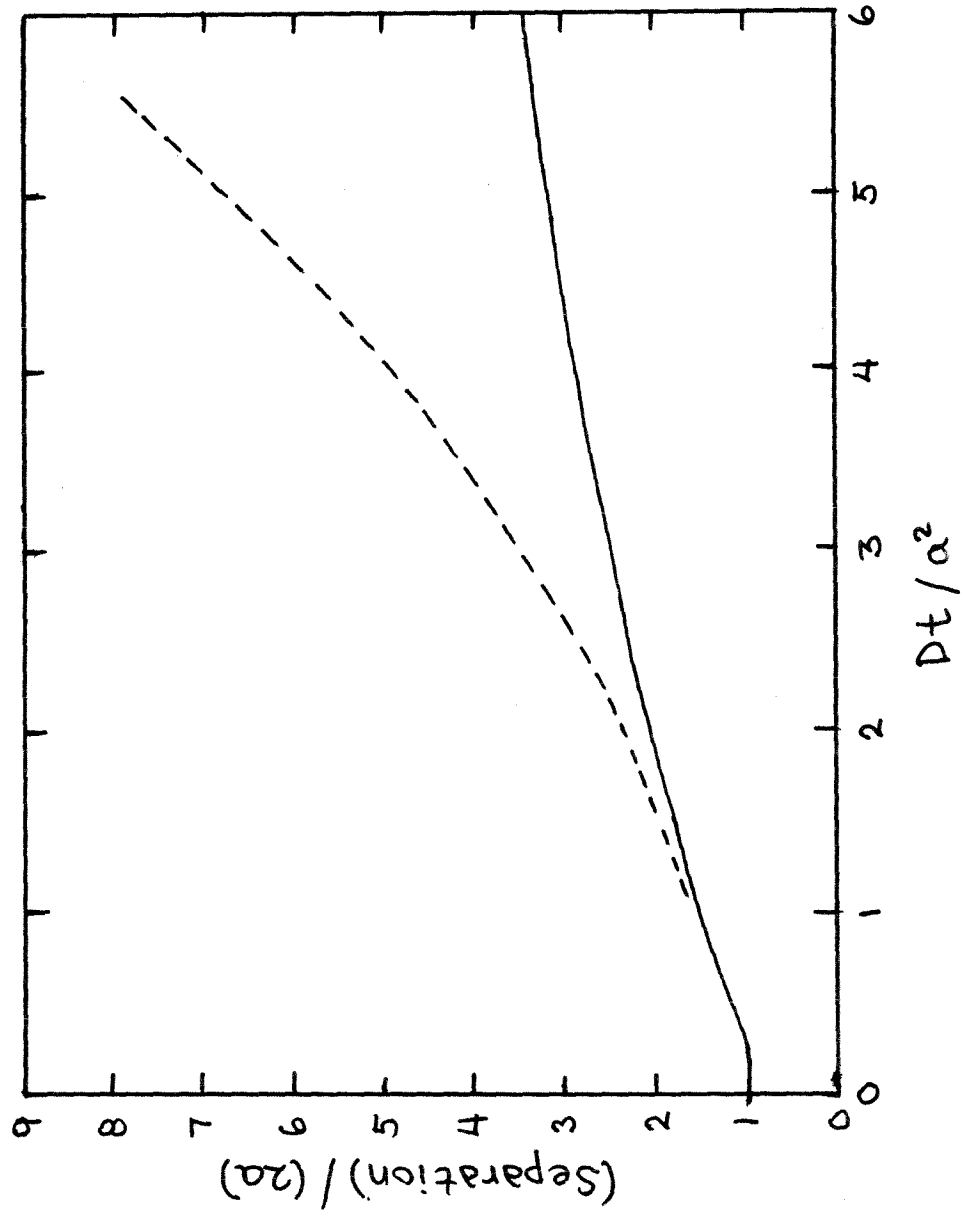


Figure 15: Separation vs Time

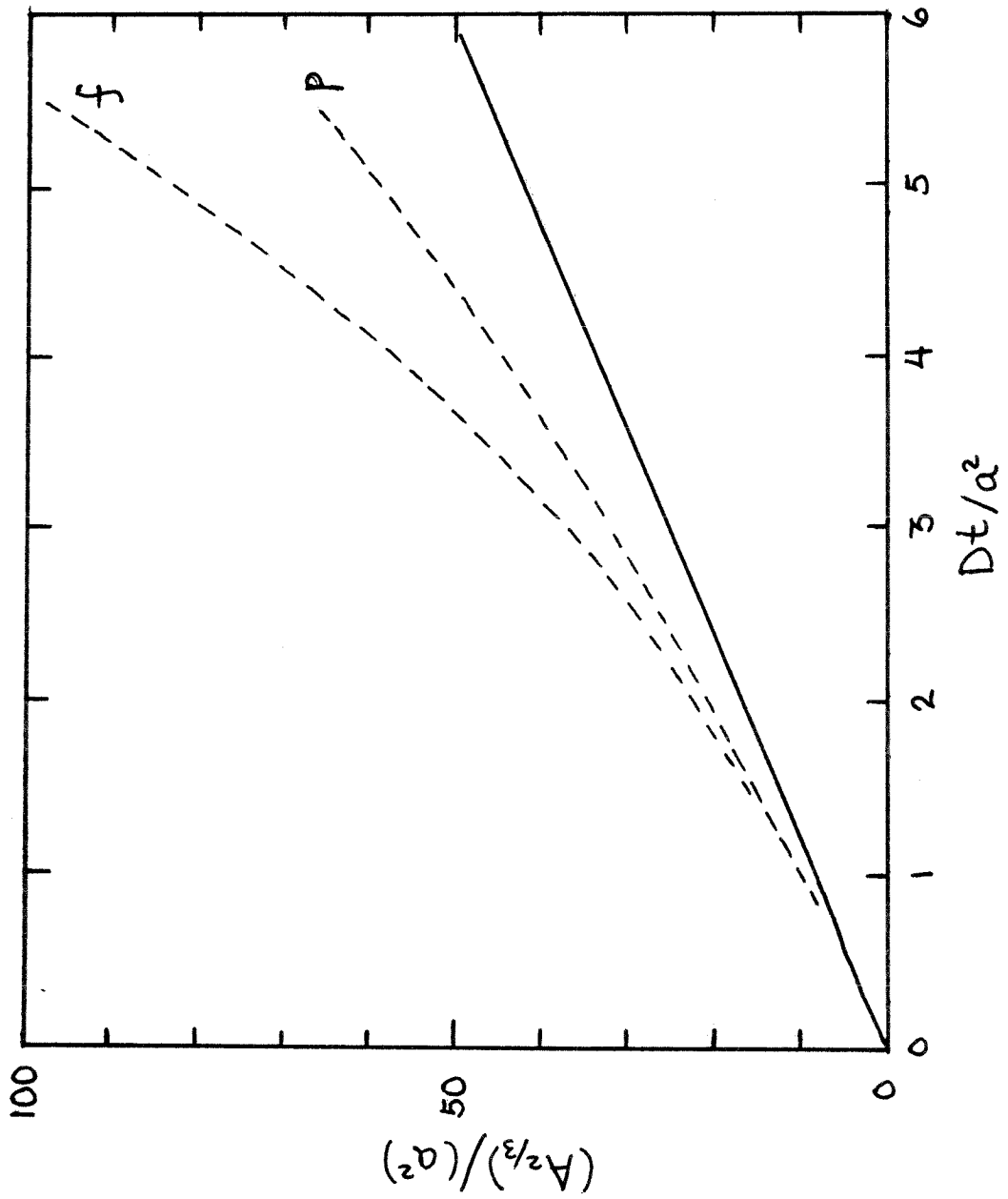


Figure 16: Area vs Time

printout, which satisfy the specified condition. Alternatively, one finds that for a wide variety of smooth, Gaussian-like forms the two-thirds area is given approximately by the total flux (or number of particles) divided by the peak field strength (or density). The 90% area is generally about twice as great.

Since the measurements of active region areas based on the NOAA reports are not particularly precise anyway, (see § 3.2.3b), the latter approach should be quite adequate for the present purpose. Thus, we will adopt:

$$A_{2/3} \approx \frac{N_+}{n_{max}} \qquad (3.18)$$

and

$$A_{90\%} \approx \frac{2N_+}{n_{max}} \qquad (3.19)$$

Fortunately, these equations do not have to be worked out separately for every choice of N_o, a, D, and t which one might wish to consider. It can be shown that equation (3.13) expresses the same geometric form for any combination of constants having a particular value a^2/Dt , provided that the distances are expressed as fractions of the fundamental separation distance (more precisely, half the separation): a . In other words, for a particular value of a^2/Dt the flux will be proportional to N_o , the peak field strength to N_o/a^2 , the separation to a, and the area to a^2 .

3.1.2b Doublet Source

A additional simplifying feature of the equations, which does not emerge very clearly from the preceding discussion, is that for times $t > a^2/D$, the predicted flux, area, and separation are relatively insensitive to the choice of a. Since for $a = 30{,}000$ km and $D = 400$ km^2/sec this asymptotic condition begins to set in at $t \approx 2\times10^6$ sec, or ~1 month; it will be quite useful in parts of the following discussion.

The doublet source approximation is attained by letting $a \rightarrow 0$, in equation (3.13), in such a way that the "doublet source strength", $N_o a$, remains constant:

$$n(x,y,t) \approx \frac{N_o a}{4\pi D^2 t^2} \times e^{-\left(\frac{x^2+y^2}{4Dt}\right)} \qquad (3.20)$$

The derived quantities corresponding to this long-time limit are then:

$$N_+ \approx \frac{N_o a}{\sqrt{\pi D t}} \qquad (3.21)$$

$$x_{max} \approx \sqrt{2Dt} \quad ; \quad s(t) = 2x_{max} \qquad (3.22)$$

$$n_{max} \approx \frac{N_o a}{\pi e^{1/2} (2Dt)^{3/2}} \qquad (3.23)$$

$$A_{2/3} \approx 2^{2/3}\sqrt{\pi}\, e^{1/2} Dt = 8.265\, Dt \qquad (3.24)$$

$$A_{90\%} \approx 2\, A_{2/3} \approx 16.5\, Dt \qquad (3.25)$$

It will be noted that, according to equation (3.24), A_{w_3} is a little over half what it would be for a unipolar source (see eqn. 3.9). The reason it is approximately half, is that the flux migrating in the $-x$ direction is annihilated by the opposite polarity half of the region. The reason that it is slightly over half, is that the annihilation of the central peak makes the outlying fringe fields more important.

It should also be noted that in practice the doublet source approximation, except inasmuch as it permits easy order-of-magnitue estimates, has rather limited application, since over the same time scales where it is valid, corrections due to the presence of differential roation are also important. In the absence of differential rotation, the doublet source approximation could be used up to $t \sim 0.2\ R^2/D$, after which the spherical nature of the diffusion problem would have to be considered.

3.1.3 Corrections due to Differential Rotation

In the presence of differential rotation, the diffusion equation has to be modified by adding a term representing the non-random transport of fields:

$$\frac{\partial m}{\partial t} = \nu \nabla^2 m - (\vec{v} \cdot \vec{\nabla}) m \qquad (3.26)$$

where $\vec{v}$ is the shear velocity. According to Newton and Nunn (1951) this varies with latitude (λ) according to the

form:

$$v = 2.77 \sin^2\lambda \ \frac{\text{deg}}{\text{day}} = 0.389 \sin^2\lambda \cos\lambda \ \text{km/sec} . \qquad (3.27)$$

Wilcox and Howard (1970) suggest that a slightly smaller amplitude may apply to active region fields. In any event, the effects due to differential rotation are, in general, rather complicated and difficult to anticipate.

Basically the differential rotation can be thought of as an additional mode of mixing, with low velocity amplitude but very long time constant, which dominates the ordinary diffusion effects at long times. The differential rotation operates both by pulling the sources apart (or together -- depending on at what latitudes they are placed), and by drawing out the diffusing patches into long skinny features, which, by virtue of their relatively great "surface area", exhibit more diffusion than would otherwise take place. Differentially, the process can be thought of as one of alternately allowing the region to diffuse for a monent, and then shearing; but over any finite amount of time, the final field cannot be obtained by simply shearing the field which would have developed in the absence of shearing.

The effects of differential rotation have been investigated empirically (by means of computer modelling) by Leighton (1964). The region which he considers, with an initial spacing of 12° (in longitude) and an inclination of $+6.54^\circ$, is about twice the size of the 60,000 km nominal

region which we have adopted in § 2.5. On the other hand, the diffusion constant, which he defines as:

$$D = R_\odot^2/T_o \qquad (3.28)$$

with T_o = 10 and 20 years, is about four times what we later will find seems appropriate (§ 3.2).

The fact that the assumed separation is twice, and the assumed diffusion constant about four times those in which we will be interested is fortunate, since it means that the value of a^2/D is about the same in the two cases; which in turn means that, in the absence of differential rotation, the expected field patterns would be about the same, except with half as great the dimensions. Since the 'horizontal' stretching due to differential rotation is roughly proportional to the 'width' of the region (in latitude), it would also be about half as great.

In other words, if we knew the fractional corrections for the solutions at $t = 0.5$ years with and without differential rotation assuming $2a = 133{,}167$ km and $D = 768$ or 1535 km^2/sec (i.e., $T_o = 20$ or 10 years -- which we can obtain from Leighton, 1964), then we can apply those correction factors just as well to the cases $a = 30{,}000$ km and $D = 200$ or 400 km^2/sec.

This has been done and the following results are obtained:

Table 3.1

$D = 200\ km^2/sec;\ t = 0.5$ years

	Without Diff. Rotation	With Diff. Rotation
Surviving Flux	$.290\ N_o$	$.219\ N_o$
Separation	1.64×10^5 km	3.03×10^5 km
Peak Field ($N_o = 10^{22}$ Mx)	11.1 gauss	5.2 gauss (p) 4.0 gauss (f)
Area (2/3)	$2.62 \times 10^{10}\ km^2$	$4.34 \times 10^{10}\ km^2$ (p) $5.66 \times 10^{10}\ km^2$ (f)
Area (90%)	$5.24 \times 10^{10}\ km^2$	$8.42 \times 10^{10}\ km^2$ (p) $1.17 \times 10^{11}\ km^2$ (f)

Table 3.2

$D = 400\ km^2/sec;\ t = 0.5$ years

	Without Diff. Rotation	With Diff. Rotation
Surviving flux	$.212\ N_o$	$.155\ N_o$
Separation	2.27×10^5 km	5.41×10^5 km
Peak Field ($N_o = 10^{22}$ Mx)	4.02 gauss	2.1 gauss (p) 1.4 gauss (f)
Area (2/3)	$5.30 \times 10^{10}\ km^2$	$7.63 \times 10^{10}\ km^2$ (p) $1.12 \times 10^{11}\ km^2$ (f)
Area (90%)	$1.06 \times 10^{11}\ km^2$	$1.51 \times 10^{11}\ km^2$ (p) $2.20 \times 10^{11}\ km^2$ (f)

As can be deduced from the tables, the correction factors are not very different for the two cases except as regards the separation between the peaks of the two polarities. For the assumed inclination of $+6.5^{\circ}$, it appears that the preservation of flux due to the pulling apart of the sources by shearing nearly cancels the enhancement of annihilation caused by the elongated contact surface, so that the net surviving flux at any time is nearly the same with or without differential rotation. The area and separation grow considerably more rapidly, however, and over six months the effect is more or less as if the diffusion constant had been doubled.

Figures 14-16 illustrate how the errors due to differential rotation accumulate with time. As pointed out by Leighton (1964), the separation, which would ordinarily grow $\propto\sqrt{t}$, develops nearly linearly with time. The predicted growth of area with time assumes a pseudo-parabolic form.

The exact corrections due to differential rotation, will, of course, depend on the exact initial conditions. Leighton's source is placed at a latitude of $\sim 24^{\circ}$. The shearing would be less if it were placed at 10°; and even less if we were to accept the differential rotation law of Wilcox and Howard (1970). We may, then, in a sense, regard the present results as representing an upper limit on the possible corrections due to differential rotation.

3.1.4 The Importance of Annihilation

We have assumed, implicitly, in the preceding discussion that the two halves of the decaying region are free to diffuse into each other. Since it is not obvious that this should be true one might wish to consider what would happen in the 'opposite' case -- i.e., if the two halves of the region were separated by an invisible barrier that prevented any exchange of particles between them.

The solution for the diffusion of a point source next to a barrier can be derived by imagining an 'image source' of equal strength (and the same polarity) placed symmetrically, an equal distance on the other side of the barrier. The image source replaces those particles which would otherwise cross the barrier, and maintains a zero density-gradient at the interface.

The solution (analogous to eqn. 3.13) obtained in this manner is:

$$n(x,y,t) = \frac{N_o}{\sqrt{4\pi Dt}} e^{-\left(\frac{x^2+y^2+a^2}{4Dt}\right)} \cdot 2\cosh\left(\frac{xa}{2Dt}\right) \quad (3.20)$$

where it is understood that the particles are positive in the $x > 0$ half-plane, and negative in the other.

The area covered by flux is about the same in this as in the former solution. The quantity of flux would remain constant, which is distinct from the former solution (in

which it falls off with time), but this might be somewhat hard to demonstrate observationally because of the difficulty of maintaining a uniform sensitivity as the field passes through various stages of fragmentation. The neutral line would also evolve differently: with a barrier, it would simply be sheared by differential rotation; with diffusion it develops less curvature. But the most important difference is in the behavior of the separation between the peaks of the two polarity concentrations. Equation (3.29), far from corresponding to a situation in which the two polarities move apart, actually describes one in which they move <u>together</u>. At times $t >> \alpha^2/D$ it indicates an $\approx$ Gaussian distribution of particles around the origin with the maximum concentration of each polarity occurring right at the barrier. Thus it is the interdiffusion and 'annihilation' of opposite polarity fields which causes the increase in separation in the random walk model of active region decay.

3.2 Description of Observations of Decaying Active Regions and Comparison with the Predictions of the Random Walk Model

3.2.1 Qualitative Description and Overview

The long term decay of active regions was studied almost exclusively by means of the full-disk magnetograms (Kitt Peak and Mount Wilson) published in the monthly NOAA Prompt Reports of Solar-Geophysical Data. The data obtained at Big Bear do not include a large enough field of view to provide an accurate picture of the interaction of the decaying fields with their surroundings; nor is the sensitivity quite as good for the weak fields. A total of 33 regions, observed on up to nine successive central meridian passages, are included in the study (see Appendix II). Many of these are the same as those discussed in §2.

Since a host of interrelated phenomena are being simultaneously investigated in such a study, it would seem useful, at this point, to present a brief summary of the key conclusions. Those points which are crucial to the evaluation of the diffusion constant will be elaborated in greater detail in §3.2.3 .

(a) During years of low activity, when the remnants of active regions can be followed against a relatively field-free background, a reasonably clear picture of evolution emerges, in which the large scale patterns of field can be thought of simply as arising by a superposition of fields

Figure 17: A series of enlargements from the Mount Wilson daily magnetograms showing the development of McMath Region 13790 over the course of six rotations:

(a) August 8, 1975 (first central meridian passage)

(b) September 4, 1975

(c) October 1, 1975

(d) October 29, 1975

(e) November 24, 1975

North is at the top, and the direction of solar rotation is from left to right. The boxes on the second and last pictures indicate the area assigned to the preceding and following plages. The Kitt Peak magnetograms are of considerable assistance in defining the exact boundary.

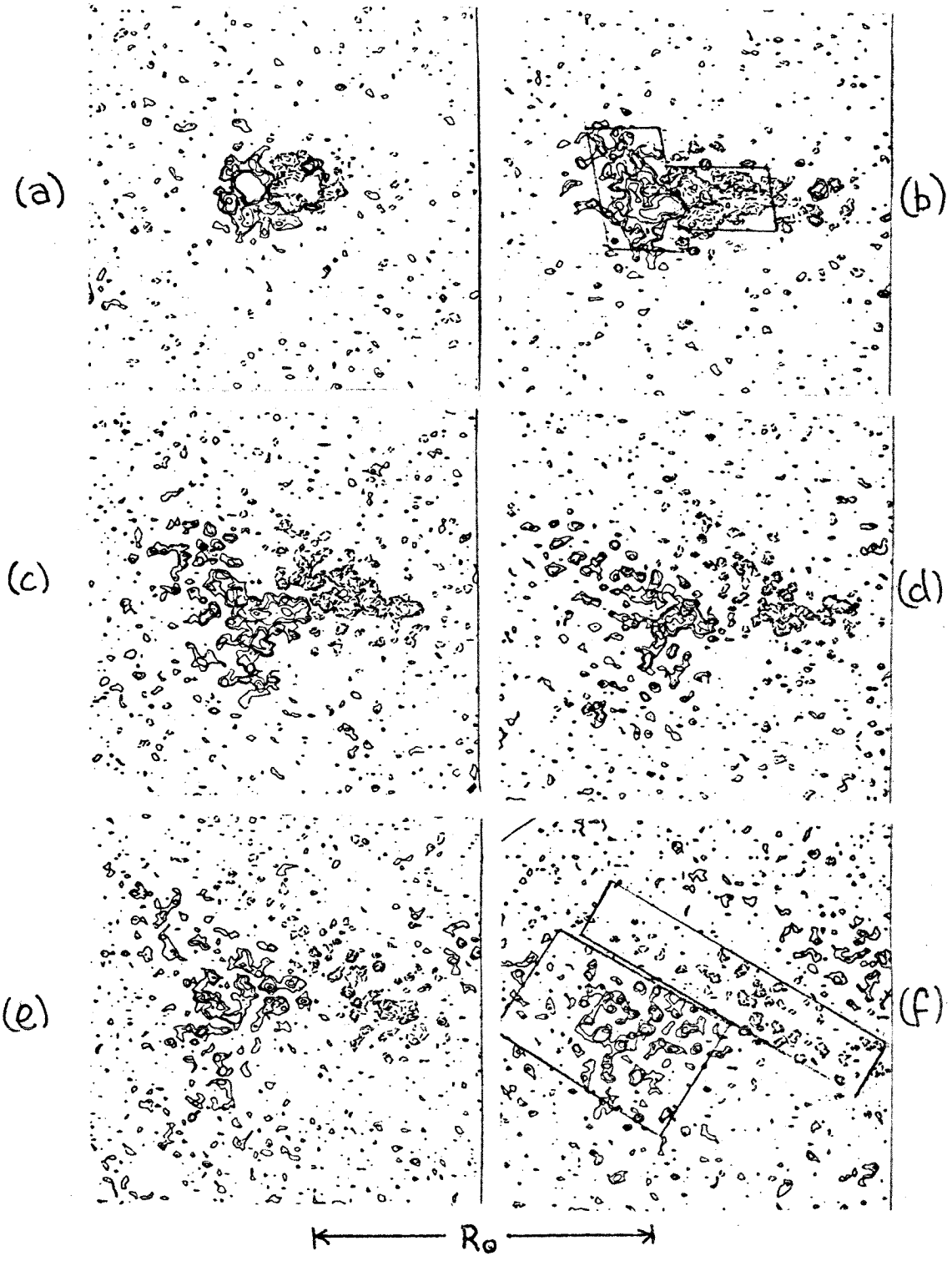


Figure 17

spreading out from the individual regions. It appears that a patch one polarity is free to merge with a neighboring patch of the same polarity, or, alternatively, to be swallowed by an encroaching field of the opposite sign.

(b) All large, weak, 'unipolar' areas appear to be the remnants of active region eruptions. In no case was it possible to find an extended unipolar area that could not be traced back to the eruption of a substantial active region; although the largest such areas may represent the sum of contributions from several (apparently unrelated) regions. The fine-scale mixed-polarity background which surrounds the unipolar areas most likely results from tiny, short-lived, localized, eruptions.

(c) The larger regions emerge in a well-known pattern with the center-of-weight of the preceding polarity being closer to the equator than the center-of-weight of the following polarity. The scatter is large enough, however, that even "reverse-tilted" regions are not uncommon (see § 2.4). As the region evolves the apparent inclination may change (often because of confusion due to small amounts of 'resurgent' activity), but the overall average seems to stay at about $+6^{\circ}$.

(d) The regions rotate differentially. Those above 15° drift towards lower Carrington longitudes, on successive central meridian passages, and those below 15° drift towards

higher Carrington longitudes.

(e) The form of the evolving regions is reminiscent of the curved and elongated forms predicted by Leighton (1964); although in the present (very limited) sample there seem to be two classes: one which develops almost exactly the curve predicted by Leighton, and another (with large positive inclinations) which seems to diffuse without shearing.

(f) Over those periods for which the field can be followed, the area dominated by each polarity increases more or less uniformly at $\sim 5 \times 10^3$ km^2/sec (slightly lower than the average rate of expansion observed during the growth phase). The expansion in area can be delayed by the presence of large long-lived spots (as in the p-ends of some regions).

(g) The separation between the peaks of the decaying p- and f-field concentrations increases at $\sim .015$ km/sec (about one-tenth the rate observed during the growth phase).

(h) The curves for the average area and separation as a function of time can be fit by a diffusion constant of $D \approx 400$ km^2/sec (ignoring differential rotation) or $D \approx 200$ km^2/sec (including differential rotation). The form of the fit is better when differential rotation is included.

(i) The lifetime -- that is the time over which the field remnants can be distinguished from the 'undisturbed' background -- is related to the source strength, $N_o a$, of the region. Again, the observed relation can be fit by a dif-

fusion constant in the range $D \simeq 200 - 400$ km^2/sec .

(j) The flux seems to fall off more rapidly with time than would be predicted by the diffusion model. This is attributed to the difficulties involved in trying to estimate the total flux of a highly-dispersed region from the kind of data used.

(k) The two polarities of the decaying region are separated by a relatively field-free "filament channel", which expands at ~ 400 km/day .

(ℓ) An H-alpha filament forms in the neutral channel on, typically, the first or second central meridian passage after the disappearance of the spots. Qualitatively, these seem weaker and less stable than the filaments which form between regions (no detailed study was made). Otherwise, the decay of a region is unspectacular in H-alpha.

3.2.2 The Problem of Resurgence

The study of active region decay is complicated by those not-infrequent cases in which the flux appears to have undergone a rejuvenation while on the backside; or, indeed, in which new, additional flux emerges before our eyes and joins with that already existing. In some instances, as when the new eruption is not well-centered on the old one, it is clear that this represents a new and unrelated eruption of field; but, since the initial emergence can itself be thought of as a series of discrete sub-eruptions, it is obvious that no

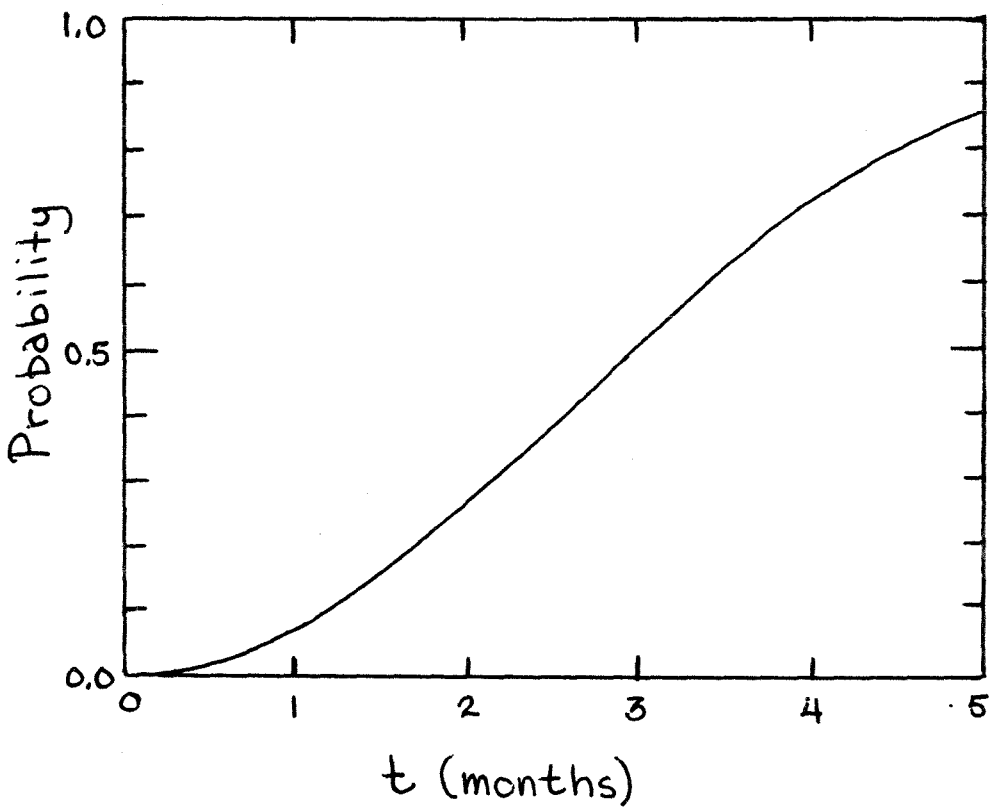


Figure 18: The probability of accidental resurgence within the boundaries of a decaying region (sunspot minimum).

definite line can be drawn separating genuine resurgences of growth from unrelated activity which serves only to confuse our interpretation of already-decaying fields.

In the present paper we shall accept the view of Howard and Edberg (1973) that most major "resurgences" of activity seen after the true initial 1 - 10 day growth period are nothing more than a reflection of the random probability of new region emerging within the boundaries of the old one.

The probability, P, of a secondary eruption occurring in time t within the bounds of a region expanding at rate $\dot{A}$ can be shown to go like:

$$P = 1 - \exp\left[-\tfrac{1}{2}\dot{A}\dot{N}t^2\right] \qquad (3.30)$$

where $\dot{N}$ is the number of regions erupting per unit area per unit time. Using $\dot{A} \approx 10^4$ km^2/sec, and $\dot{N} \approx 2\times10^{-18}$ km^{-2} sec^{-1} (corresponding to a typical solar minimum pattern of about 150 significant regions per year in $\pm 20^\circ$), it is not at all surprising to find that a considerable amount of searching is required to locate a region that can be followed for more than a few rotations without "outside interference" (see Figure 18).

3.2.3 Quantitative Results

3.2.3a Separation

Of the many parameters (such as flux, average field strength, area and separation) which can be used to describe

the state of a decaying active region, the easiest to measure, and the one which gives the most unambiguous results, is the separation between the peaks of the plus and minus field distributions.

The results for separations over periods of up to 1.6×10^7 sec (7 rotations) are shown in Figure 19. After a quick surge to an initial peak separation of $\sim$60,000 km, a gradual, and essentially linear, expansion at $\sim$.015 km/sec is seen. Of the 33 regions considered, about 10 survive long enough to contribute to this trend. There is a possibility that some of the short-lived regions may expand less rapidly during their decay, but in those cases the remnant fields are so weak that a definite judgement cannot really be made (cf. §§2.3.1 & 2.3.5).

In the absence of differential rotation, the data could be fit, approximately, by the predicted growth of a bipolar source with D in the range 400-800 km/sec , but the parabolic shape of the predicted curves does not seem particularly appropriate to the observed linear trend. A much better fit is obtained if one includes the effects of differential rotation as given in Table 3.1 and Figure 15. A diffusion constant close to 200 km^2/sec appears to be indicated.

Figure 19: Measurements of separation versus time. Each curve represents one of the 33 regions listed in Appendix II.

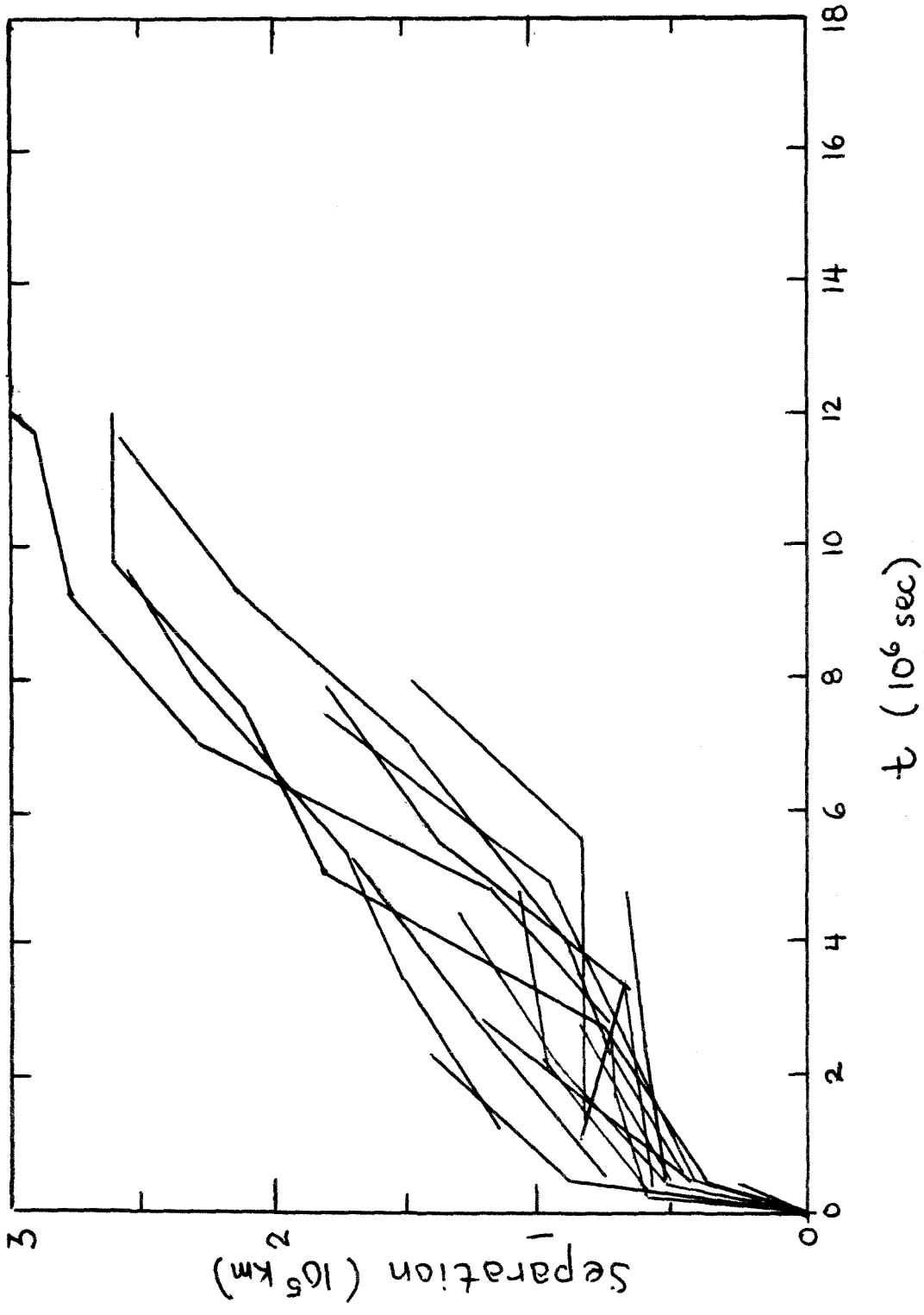


Figure 19

3.2.3b Area

The area dominated by field of a particular sign is much more difficult to estimate than the separation. For the present study, the intention was to measure the area over which the field appears to be 'noticeably' enhanced above the background level. A considerable degree of consistency is found when this estimate is made, independently, from either the Kitt Peak or the Mount Wilson data. Unfortunately, it is less clear how this measured area relates to the theoretical estimates developed in §3.1 .

The random walk model does not predict that the flux distribution will have a sharply defined edge. For example, if we assume a typical step length of 8000 km in 24 hours, then after, say, 4 rotations (10^7 sec), when the majority of the flux is confined within a circle of radius 8.6×10^4 km, an occasional line could have moved by as much as 8.8×10^5 km (110 steps). This inherent imprecision of the border, combined with the non-uniformity of the background field makes the estimates of occupied area highly uncertain. When a decaying plage is situated in the middle of weak pre-existing fields of the same sign it can seem to expand very rapidly. When it appears in the middle of an area of weak foreign polarity its growth can be retarded, or it can even seem to shrink. On the average one might hope that these errors would cancel, contributing mainly a large scatter to the

measured values.

Figure 20 presents the results. The areas are comparable to those reported by Bumba and Howard (1965b)* for unipolar magnetic regions; but significantly less than the size of the UMR described in Babcock (1963), and referred to by Leighton (1964), and others, which presumably represents a complex of activity resulting from contributions of several individual regions.

If one chooses to ignore the scatter, then it is fairly clear that there is an average long-term growth rate of about 5×10^3 km^2/sec . Assuming that these areas refer to $A_{90\%}$ as given by equation (3.5), a diffusion constant of $D \approx 300$ km^2/sec would be inferred. A somewhat lower diffusion constant ($D \approx 200$ km^2/sec) would be inferred if the trend towards increasing slope is attributed to differential rotation, as seems likely (see Figure 16). On the other hand a higher diffusion constant is again indicated if the observed areas correspond more nearly to $A_{90\%}$ than to $A_{1/2}$, as also seems likely, particularly in the more advanced stages of decay.

It will be noted that the lowest curves are produced by regions with long-lived p-spots (e.g., McMath 13338 and 13722). The presence of the spot seems simply to delay the release of the flux to the whims of the supergranular currents. Once the spots disappear, the normal pattern of expansion sets in.

---

*I am assuming that there is a factor of 10 error in the scale of their Figure 5.

Figure 20: Measurements of area versus time. Estimates for the preceding and following plages are indicated separately, so that each region contributes two curves. Region 13338 (lowest curve) has a large, long-lived p-spot, which appears to delay the expansion of the p-plage. The dashed portion of the curve indicates the lifetime of the spot.

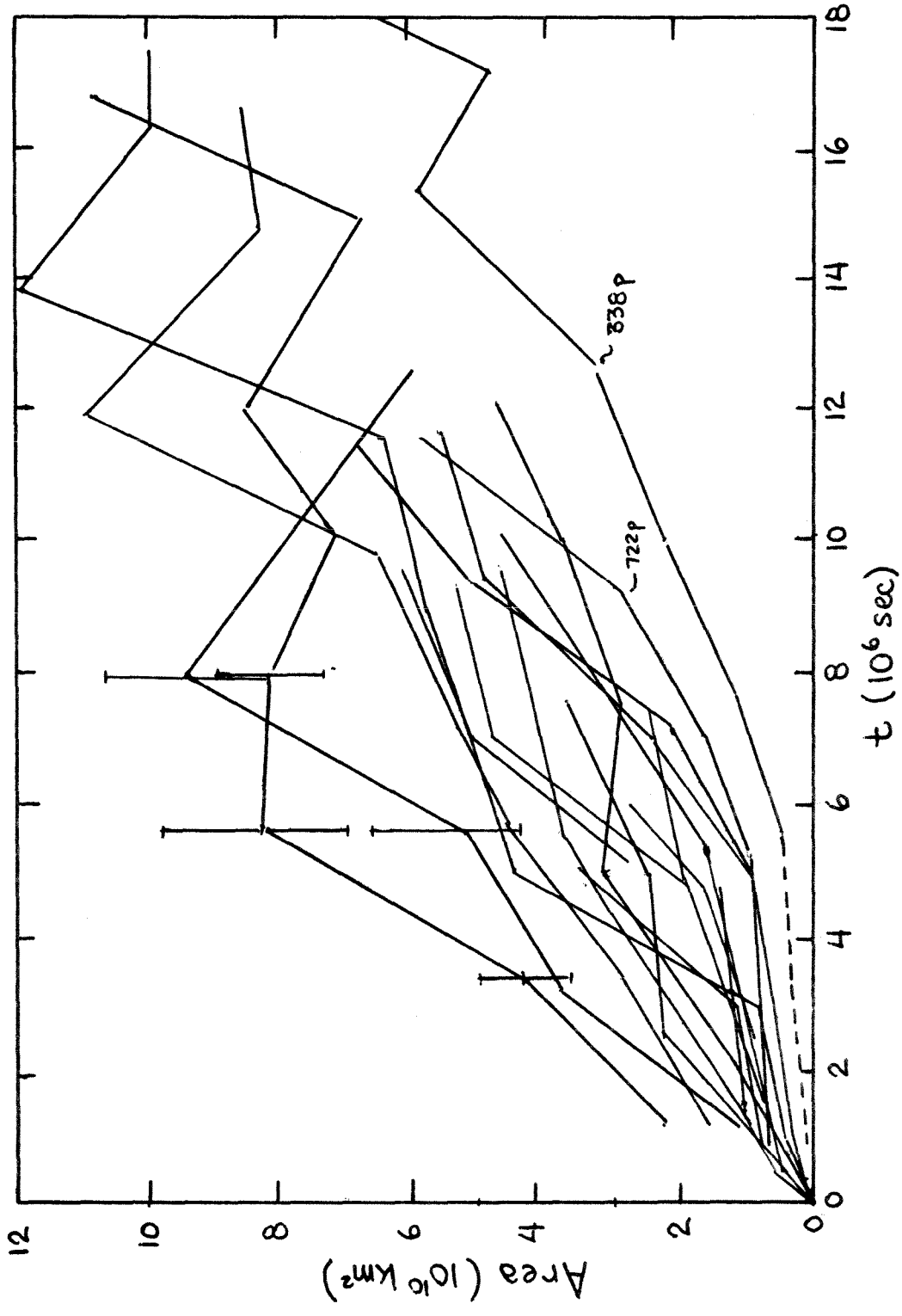


Figure 20

One might think that the measured areas would correspond more nearly to the area within which the predicted average field strength is above some fixed threshold level (comparable to the average background field -- say 1 gauss or 5 gauss), than to either $A_{2/3}$ or $A_{90\%}$. This possibility was considered in some detail, but it was found that the predicted curves did not add up very well to what was observed. In general, the area within a contour of a particular strength (in the smoothed model) would be expected to rise fairly rapidly to a mixiumum, and then fall until, ultimately, there is no field left above the chosen threshold. In the presence of differential rotation, the curves (at the 1-2 gauss level), reach higher peaks and terminate more rapidly (by factors ~ 2).

This behavior might well be observed if the area could be measured in a more objective manner (e.g., using the raw digital data), but it does not correspond very well to what is seen visually. Visually one sees the field weaken as it expands, fading (or blending into the background) more-or-less all at once at some 'maximum' area. One does not see it recede in area, or leave a 'shrinking' residue (except in those cases where it is obviously being squeezed out by a foreign polarity).

Part of the reason, obviously, is that in making the study visually one is continually lowering the threshold level of acceptance: when the region is young, only fields

above 5 gauss (nominal) are considered as being a part of it; when it is old, isolated fragments of flux averaging less than 1 gauss might be included. In addition, it should be noted that the graph includes only those regions which remain visible long enough to draw the curve. It is not known how to determine the area over which the remnants are spread when they are no longer strong enough to be distinguished from the normal background field. As a result, the latter part of the curve will be biased towards the larger source regions, and the early part towards the smaller source regions.

In summary, the measurements of area are rather subjective, and subject to considerable uncertainty, but they seem to indicate (if one includes the effects of differential rotation) a diffusion constant in the range $D \approx 200\text{-}400$ km^2/sec.

3.2.3c Lifetime

The lifetime of a magnetic region can be defined as the time over which its remnant fields can be distinguished from the general background. In many cases the lifetime cannot be determined (except as a lower limit) because of interference from nearby, or seemingly resurgent, activity (see §3.2.2). In those instances where interference is not a problem, a reasonably definite value can usually be arrived at, even though the idea is not entirely precise, and observed lifetimes can be found ranging anywhere from $\lesssim 1$ day (for an

ephemeral region) up to ~ 1 year (for the largest).

The predicted lifetime can be estimated, in the random walk model, by assuming that the lifetime is determined by the condition that the maximum surviving field strength fall below some threshold value. Since lifetime considerations will, inevitably involve the most advanced stages of a region's decay, we can use the doublet source approximation in our theoretical considerations. From equation (3.23) we know that the maximum field at time t is given, approximately, by:

$$B_{max} \approx \frac{N_0 a}{\pi e^{1/2} (2Dt)^{3/2}} \tag{3.31}$$

where N_0 is the flux (one sign) and a is the spacing parameter (half the separation) of the source field. If we demand that the peak field reach a threshold value B_{thresh}, then we obtain for the lifetime, T, that:

$$T = \frac{1}{2D} \left(\frac{1}{B_{thresh} \pi e^{1/2}} \right)^{2/3} (N_0 a)^{2/3} \tag{3.32}$$

Thus when plotted on log-log paper, one expects to find a linear relationship between the lifetime and the doublet source strength with a slope of two thirds. Figure 21 indicates that such a slope is indeed observed (the individual points are obviously rather uncertain, but because of the large range of values being considered it becomes relatively unimportant, as far as the slope is concerned, whether one

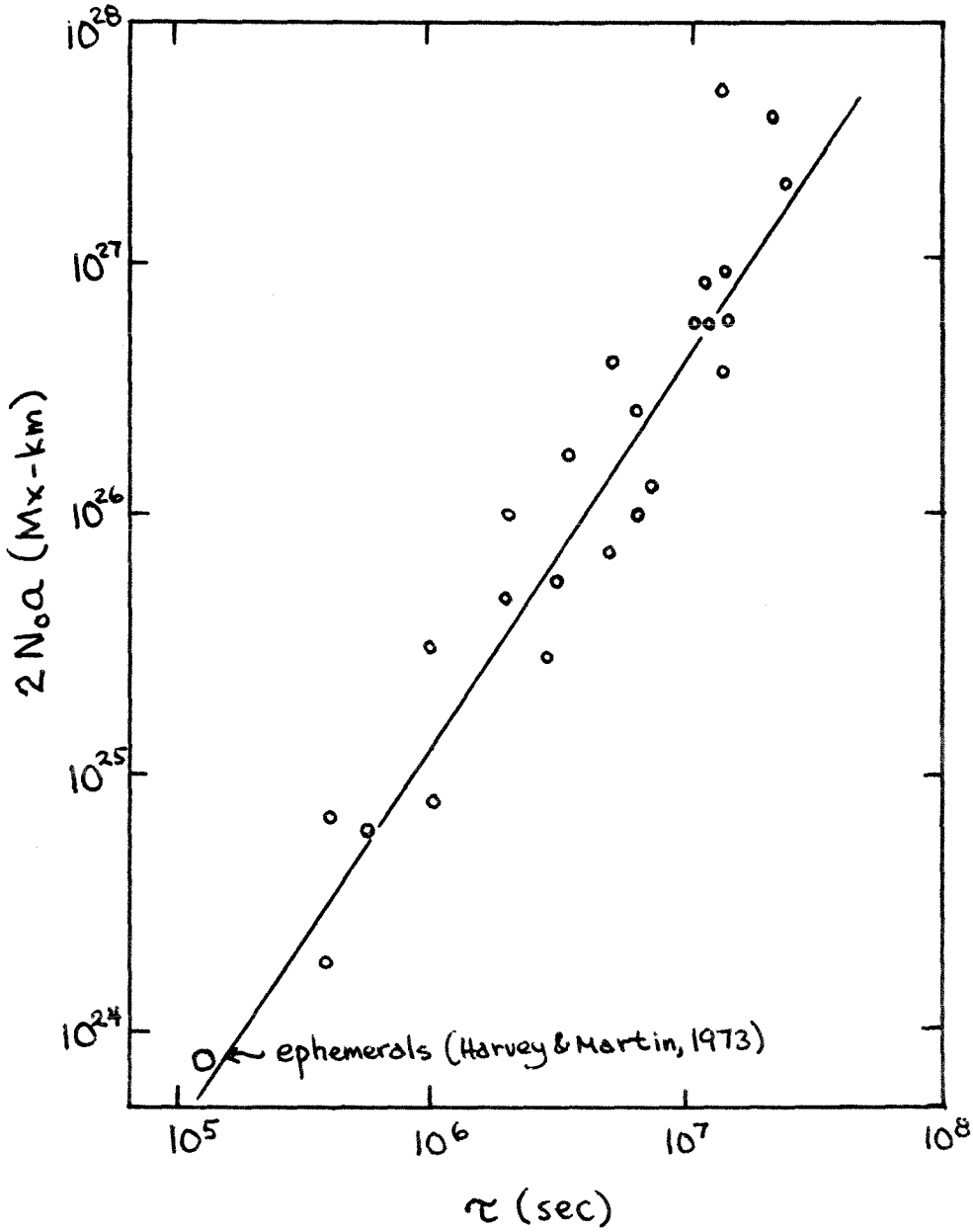


Figure 21: Measurements of lifetime versus doublet source strength. The data are listed in Appendix II.

says that a particular small region lasts 6 or 12 hours, or, alternatively whether one says a particular large region lasts 5 or maybe 7 rotations). The best fit to the points (with $T \approx 10^6$ sec at $2N_o a \approx 1.26 \times 10^{15}$ gauss-km^3) indicates:

$$B_{thresh} D^{3/2} \simeq 4.3 \times 10^4 \qquad (3.33)$$

where B_{thresh} is measured in gauss, and D in km^2/sec.

The only significant deviation from a simple power-law behavior seems to be among a few of the largest regions, which do not appear to last quite as long as might be expected by extrapolation from the smaller regions. The sense of the deviation is understandable for two reasons. First, the observations of the larger regions are more likely to be terminated by interference from neighboring activity, so that, on the average, one has a tendency to underestimate their lifetimes relative to the others (i.e., they have a higher effective B_{thresh}). Second, the lifetimes of the larger regions include increasing those times over which the corrections due to differential rotation become significant, giving them, in effect, a larger diffusion constant, as well (by up to a factor ~2).

Now, in order to extract a diffusion constant from the lifetime data, it is necessary to establish what the actual threshold field level is. On the Mount Wilson magnetograms,

as they appear in the NOAA bulletins, the lowest contour is labelled "5 gauss", but is is well known that a diffuse field does not have to be as great as 5 gauss to produce contours, since flux -- whatever its true average strength -- will be concentrated into a few isolated patches with dimensions $\sim 1''$ and strengths of, perhaps as much as 1000 - 2000 gauss (see, for example, Stenflo, 1973) -- and these are the objects represented by the contours.

The level which we want to use in equation (3.33) is neither the strength of the contours nor the true strength, but rather the average strength which a typical 'quiet' background field would have if the flux represented by the contours were spread out uniformly over the spaces in between. Unfortunately, this requires a knowledge of the amount of signal at levels below those which are normally plotted.

The fact that extending the plotting to "2 gauss" brings up a general, and partially random, splotchiness to the magnetograms (Howard, 1974a) and the fact that on some days there are obvious zero-biases apparent even at the 5 gauss level would suggest that the average strength of a field which is just beginning to disappear from the contour plot is somewhere in the range 2 - 5 gauss. This is somewhat higher than the average noise level, which we can estimate, by dividing the average daily flux (2.1×10^{21} Mx) by the area ($1.88 \times 10^{21} cm^2$), as being $\lesssim 1.1$ gauss in the 40-50° latitude zone

(Howard, 1974b), which should be typical of "quiet" fields.

It is reassuring that essentially the same lifetime is obtained independent of whether one uses the Mount Wilson or the Kitt Peak data (the present results are based on a combination of the two); but, unfortunately, this observation does not assist us in defining any more precisely the value of the threshold field. Without further information, about all we can say is that the (nominal) average field strength of a vanishing plage is probably between 1 and 5 gauss.

None of the numbers given above has been corrected for the factor of two line weakening correction (Howard and Stenflo, 1972); but the source strengths were. Thus it is the threshold strength in "actual" gauss that we want to use in equation 3.33 . Since one "Mount Wilson gauss" equals two "actual gauss", we have $B_{thresh} \approx 2 - 10$ gauss. This implies that $D \approx 260 - 770$ km^2/sec, depending on whether one accepts the higher or the lower threshold estimate. The correction for differential rotation would be quite small since the fit is biased towards the short-lived regions, but there is an intrinsic uncertainty of $\sim \pm 50\%$ due to the scatter in the points.

3.2.3d Flux

Of the primary measurements regarding the decay of active region fields, the ones which, at present, add up least well with the predictions of the random walk model are those regarding the amount of surviving flux; and this is hardly surprising considering the crudeness of the manner in which the flux has been estimated (based on the areas inside the contours on the tiny reproductions of full-disk magnetograms published in the NOAA bulletins) -- which may greatly underestimate the amount of field present in a diffuse and highly-fragmented region.

From § 3.1.2 we know that the flux should go like:

$$N_+ = N_0 \, erf\left(\frac{a}{\sqrt{4Dt}}\right) \tag{3.34}$$

or, asymptotically:

$$N_+ \approx \frac{N_0 a}{\sqrt{\pi D t}} \, . \tag{3.35}$$

Which means that one should be able to find one-tenth of the original flux surviving at times $\sim 30\ a^2/D$ (with some small correction due to differential rotation). For our nominal $a = 30{,}000$ km region, and $D \approx 200 - 400$ km^2/sec, this corresponds to something like 2 - 4 years.

In practice, it is difficult to find one-tenth of the original flux, on a normal Mount Wilson 5 gauss contour plot, after only 2 - 3 rotations. In fact, the flux represented

by the 5 gauss (and above) contours seems to fall off more-or-less exponentially, with half-lives ranging from 5.4×10^6 sec for a region with $N_o \approx 3 \times 10^{22}$ Mx (which can be followed for $\gtrsim 6$ rotations) to $\sim 1 \times 10^6$ sec for a region with $N_o \approx 3 \times 10^{21}$ Mx (which can be followed for only 1 - 2 rotations) (one rotation = 2.357×10^6 sec). If taken at face value, these data would imply diffusion constants upwards of 800 km^2/sec. Figure 22 shows an example of the kinds of discrepancies which are encountered.

The present results must be discounted, however, because of our incomplete knowledge of the amount of flux represented by signals below the 5 gauss contour (which could easily be the bulk of the field in a decaying region). Even with access to the raw digital data it is unclear that an entirely accurate estimate of the flux could be made, both because of one's uncertainty in knowing how to define the boundary of the region, and also because of a basic uncertainty as to whether the magnetograph really accurately represents the relative strength of fields exhibiting very different degrees of fragmentation, particularly when the apparent strength of those fields is close to the threshold of the instrument.

If it were possible to make these corrections with the required degree of confidence, and if it should still turn out that the rate of flux annihilation seems to be higher than could be anticipated on the basis of the apparent

Figure 22: Measurements of flux versus time for Region 13790. The dashed line connects the observations (based on the Mount Wilson 5-gauss contour plots); while the solid line indicates the predicted decline in flux for a region with the observed initial separation ($2a = 1.15 \times 10^5$ km), neglecting differential rotation. The points labelled 15851 and 15852 represent smaller regions whose total flux over 1-2 rotations was studied by Sheeley (1965) on the basis of photographic cancellations. They have been plotted on the same graph by scaling. In principle, all regions should lie along approximately the same curve if the fractional surviving flux is plotted against t/a^2 (cf. Fig. 14).

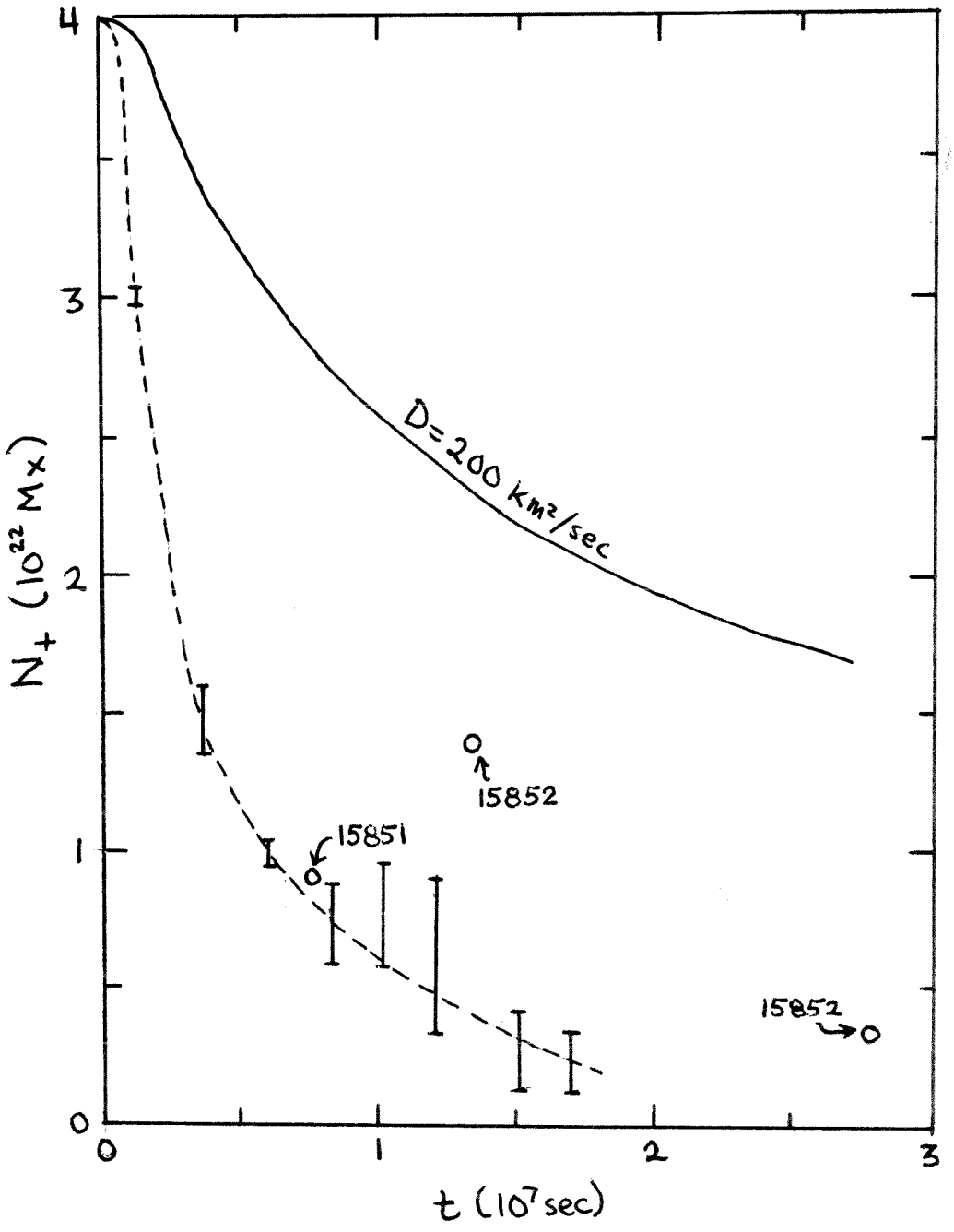


Figure 22

spreading of the fields; then this would seem to indicate the presence of some systematic tendency for field lines to recombine with definite partners, even though over periods of weeks, and even months, they may appear to be moving about at random. One could imagine this happening, for example, if individual regions (including the coronal and subsurface fields) were thought of as great rising toroidal loops, standing up vertically through the surface, and detached from the deeper "general" field which generated them. Over reasonable periods of time the fields, at their points of contact with the photosphere, might seem to be fraying apart and spreading at random; but over the long run, as each line of force is lifted out into the corona, two particular footpoints come back together and cancel, and the region as a whole would leave no lasting residue.

It is not meant to imply that there is anything in the present data which either requires, or even suggests, that such a modification to our thinking is necessary (see, however, §6.6). Indeed, one suspects that since the lifetime data work out rather well, and since the lifetime is itself determined by a combination of the area and the amount of surviving flux, there cannot be anything too seriously wrong with the rate of annihilation. The numbers are all uncertain by factors ~ 2, however.

3.3 Possible Discrepancies with the Predictions of the Random Walk Model

3.3.1 The Development of the Filament Channel

Just as one expects that in the outer fringes of an expanding bipolar region there will be a point at which the average field drops below the background level, so one also expects that between the sources there will be a zone where the opposite polarities have interdiffused to such an extent that, at least at low resolution, no average field would be detected.

Such a "neutral corridor" or "filament channel" is indeed observed in virtually every decaying region (except where disrupted by resurgent activity); but, rather surprisingly, at the present sensitivity it appears to be occupied not by the fine-scale salt-and-peppering of polarities which one would have anticipated, but rather to be almost totally free of field of either sign, leading to a nearly perfect segregation of the two polarities. If the random walk picture is at all correct, and if this trend continues at higher resolution and sensitivity, then it would indicate that the process by which opposite-polarity lines recombine in areas of overlap is very rapid and efficient -- otherwise a considerable intermixture of polarities would have to develop at the interface between the two decaying plages.

The main disagreement (if there is any) to the predictions of the model seems more to be in the manner in which the channel grows. Now, the present measurements were made rather hastily, and are based almost exclusively on the appearance of the regions on the Kitt Peak magnetograms, so it is quite possible that a uniform sensitivity was not employed, but they seem to suggest that the width of the channel grows in a manner more-or-less independent of the source strength of the region. That is, the width after one or two rotations can be just as great for a large region (that will go on for many more) as for a small region (which is in the process of dying, and should, one would have thought, have had a relatively wide channel).

In any event the results are given in Figure 23, where the average channel width is shown for regions aged 2 - 9 rotations. One seems to see a fairly clear linear trend at $\sim$470 km/day; although the scatter for the individual regions is large, and the measurement, in its present form, is rather subjective (it is supposed to correspond to the average distance between 'significant' opposite polarity fields at the point of closest contact).

The predicted neutral channel width can be determined easily enough by assuming that it represents the area in which the average field strength is below some fixed threshold level ($\sim$2 - 10 gauss, as in §3.2.3c). We know from

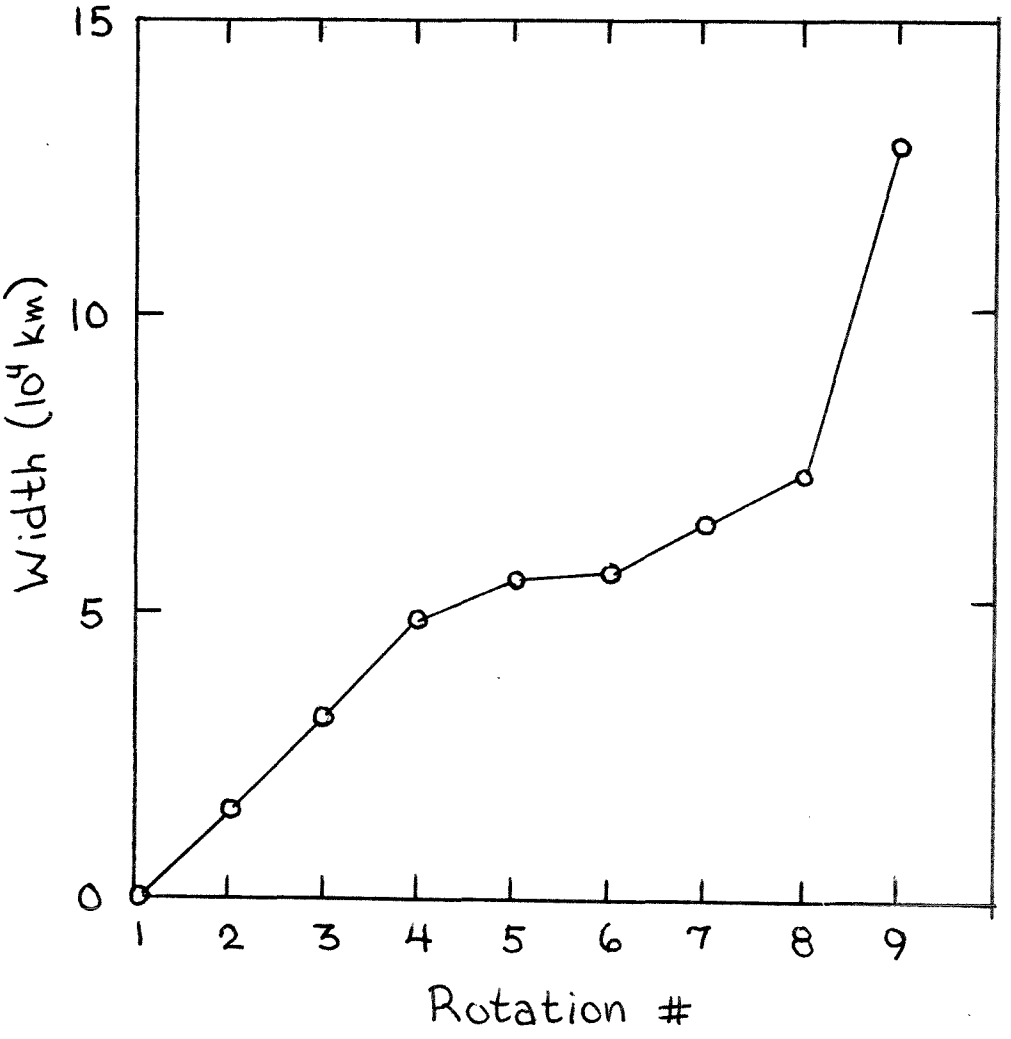


Figure 23: The width of the filament channel on successive rotations. Each point represents an average over all the regions surviving that long. The measurement is difficult and imprecise, and the scatter is large.

equation (3.13) that the field near the neutral line (at long times) can be approximated (cf. eqn. 3.20) by:

$$B(x) \simeq \frac{N_0 a}{4\pi D^2 t^2} x \qquad (3.36)$$

where x is the distance from the neutral line along the axis connecting the two sources. The width, w, will be given by:

$$w = 2x = \frac{8\pi D^2 B_{thresh}\, t^2}{N_0 a} \qquad (3.37)$$

This predicted quadratic growth, does not seem to be observed, but, then, the later points are biased towards regions with larger source strengths, and that could suppress such a trend even if it were present.

It should be noted that equations (3.36) and (3.37) are not valid for times such that $\sqrt{Dt} \lesssim a$, since the diffusion model densities begin looking like a real region only after the two polarities have started diffusing into each other. Indeed, in the approximation with discrete bipolar sources, the width of the neutral channel, as defined above, would actually <u>decrease</u> at first, beginning the quadratic growth only after reaching some minimum value. Thus the narrowness of the neutral channel in a weak, but not yet particularly well-diffused region, could easily be imposed by initial conditions beyond the scope of the normal 'random walk'.

If we assume that $B_{thresh} \approx 6$ gauss and assume that a point in the middle of the curve (say $w \approx 5 \times 10^4$ km at $t \approx 10^7$ sec) would correspond to our nominal region with $N_o = 10^{22}$ Mx and $a = 3 \times 10^4$ km; then we find that the width of the neutral channel would be explained by a diffusion constant of $D \approx 300$ km^2/sec, which is certainly not incompatible with the previous results. The predicted width would be slightly greater if the effects of differential rotation were included, but that is a correction small compared to the other uncertainties in the problem.

In summary, it seems that the development of the filament channel, which, at first, might seem to be a problem, is not likely to be a problem at all, although the exact pattern of growth will undoubtedly be sensitive to the exact initial distribution of field. More careful, and more objective measurements would be useful.

3.3.2 The Spatial Distribution of Field Fragments

A suggestion was made in §3.3.3b that when the magnetograms are compared visually an undisturbed field generally seems to fade more or less all at once at its maximum area (being visible on one rotation, and gone on the next). The same impression is had in K-line (Butler, 1924). By implication this might seem to mean that the flux winds up spread more-or-less uniformly over an available area, rather than having the fuzzy Gaussian-like distribution which character-

izes the random walk model. If such an impression was created it was unintentional. The visual impression is just as consistent with the 'top' of an extended distribution dipping below a threshold. The boundary of a dying region is never sharply defined.

It would be nice to be able to show explicitly that a cross section of field strength through a decaying region actually has the Gaussian-like shape predicted by the diffusion model (and, indeed, the width of such a curve would provide a far more objective measure of the 'area'), but it involves a process of spatial averaging which is beyond the scope of the present paper.

3.3.3 Shearing of the Neutral Line

A sketch of each region involved in the present study was made on transparent plastic, showing its general appearance on each successive central meridian passage, primarily with the idea that these transparencies could be laid on top of each other to produce a picture representing the average pattern of development (Figure 24). One of the most obvious characteristics of this average pattern is the shearing of the neutral line.

Shearing is expected when the diffusion takes place in the presence of differential rotation: the differential rotation tries to draw out the (initially vertical) line into a crescent (just as it would distort a line of constant

Figure 24: The average pattern of active region development as deduced from a composite of tracings from Mount Wilson and Kitt Peak daily magnetograms of 33 selected regions on successive rotations. Tracings from the southern hemisphere have been inverted so that they could be properly superimposed on those from the northern hemisphere. Thus, the poleward direction is at the top, and the equatorward direction at the bottom of each figure. Solar rotation is from left to right.

The regions are listed in Appendix II. Only four (McMath 13722, 13790, 13890 and 13818) survive the entire six rotations.

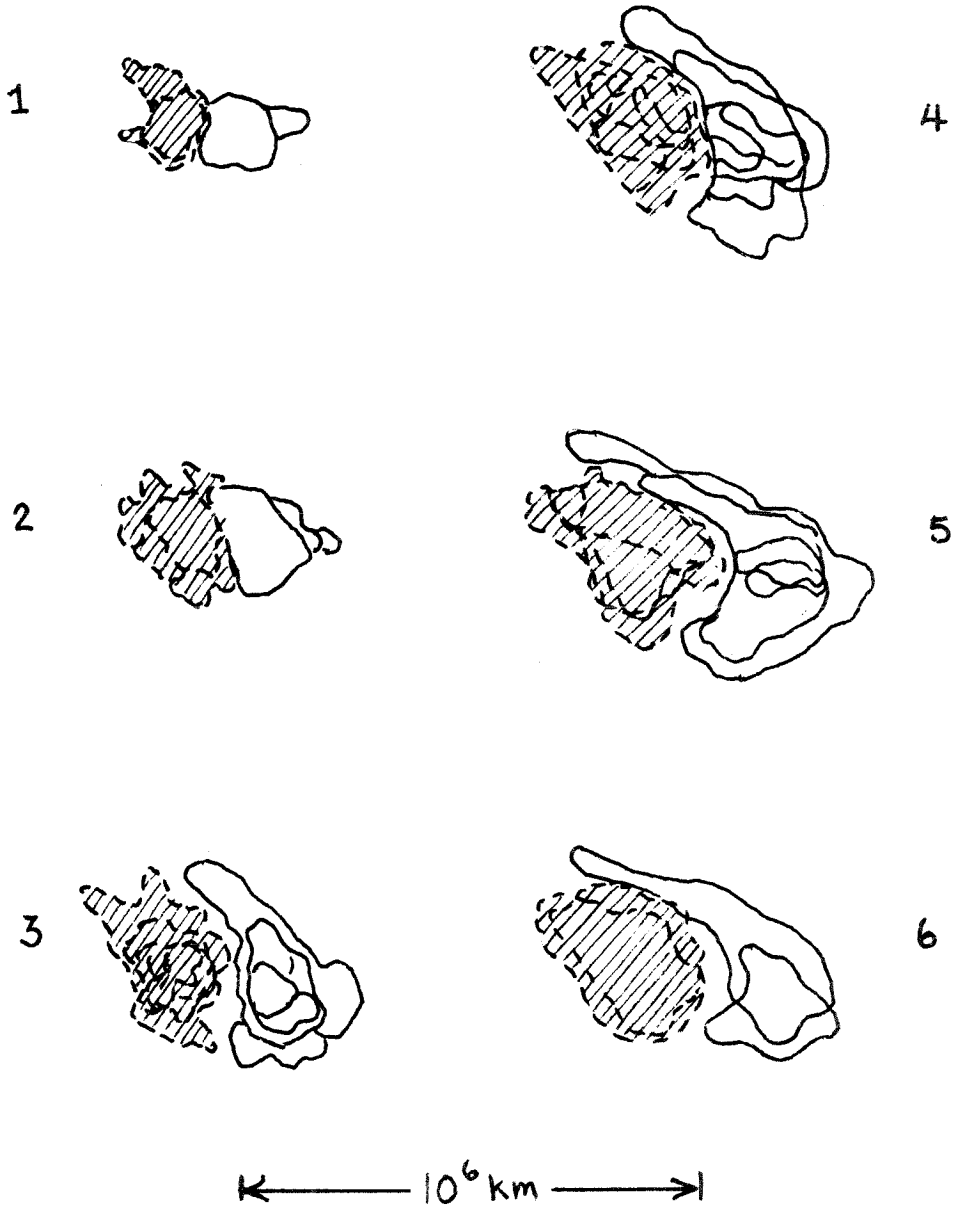


Figure 24

longitude); diffusion tries to straighten it out. Leighton (1964) seems to have shown that the bending is relatively insensitive to the choice of the diffusion constant, for using both $D = 768$ km^2/sec and $D = 1535$ km^2/sec he finds virtually the identical curvature after 0.5 years, but in both cases it is significantly less than might have been predicted by the Newton and Nunn formula which was put into the calculation.

In the present sample there appear to be two classes of large active regions. Some such as McMath 13790 and 13818 emerge in a east-west orientation, and develop, as they decay, almost exactly the neutral line curvature predicted by Leighton. Others, such as McMath 13890 and 13722 emerge with large positive inclinations ($\sim +30$-45°) and seem to evolve without shearing. The presence of these two long-lived components gives the hooked appearance to the composites on rotations 5 - 7 (Figure 24).

On the average, then, one would have to say that the shearing is somewhat less than predicted by Leighton, and much less than predicted by Newton and Nunn, even though the diffusion constant is small.

The present author does not entirely understand how changes in the initial inclination would affect the predicted neutral line curvature. Conceivably it could explain the presence of the two classes.

3.3.4 Lack of Long-term Residue

The most serious conflict of the simple random walk model with observation appears not to be in its description of the evolution of any individual region, but rather in its predictions regarding the long-term, cumulative effects of many eruptions.

If one thinks of the magnetic polarities as being averaged over narrow strips of equal latitude, then each region, as it comes up, with its preceding polarity slightly closer to the equator than the following one, will contribute a tiny north-south dipole moment which spreads out in latitude and weakens. The spreading is quite gradual, considering the large dimensions involved, and even over periods of several years one can think of the contributions from the individual regions as, essentially just adding (the two hemispheres are widely enough separated that they can be thought of as developing independently -- with only a slight cancellation at the equator).

These considerations lead to a very definite, and natural, prediction, that the latitudes immediately poleward of the active zones will develop a residual polarity (with the same sign as the following spots) which grows with every new cycle spot that erupts. Leighton (1964), assuming diffusion constants of $D = 768$ and 1535 km^2/sec, and making some reasonable assumptions about the amount of flux injected over the course

of a cycle, was able to show that the residual field would be on the order of a few gauss; which did not seem to conflict with any of the magnetic data available at the time.

Unfortunately, the strength of the field is inversely proportional to the diffusion constant, so that if one accepts the range of diffusion constants ($D \approx 200 - 400$ km^2/sec) suggested by the present measurements of active region spreading, a residual field (immediately poleward of the sunspot zones) on the order of 10-20 gauss would be anticipated, in the declining years of the cycle. This definitely does not seem to be in accord with the observations. Howard (1974a) finds that the average excess of following polarity at $40\text{-}50^\circ$ latitude is at most a few tenths of a gauss. It is only right at the poles that one sees the same polarity, consistently, day after day.

The lack of a substantial long-term residue suggests that the seemingly over-rapid annihilation of flux mentioned in §3.3.3d, and the somewhat shorter-than-anticipated lifetimes (requiring slightly higher than 'normal' diffusion constants) found in §3.3.3c, may be more than artifacts arising from inadequate data. The active regions seem surprisingly capable, over the long run, of erasing all record of themselves.

This problem will be discussed in greater detail in §6.6 .

3.4 Summary

No empirical model of any real phenomenon is expected to be perfect in every detail; those which coincide with reality on more than one or two points are generally regarded to be 'successful'. By that standard, we have found the simple bipolar diffusion model to be remarkably successful (at least at the present level of sophistication) at predicting the observed pattern of active region decay. Measurements of separation, area and lifetime are all consistent with a single diffusion constant somewhere in the range $D \approx 200 - 400$ km^2/sec. In addition, it seems likely that the width and curvature of the neutral channel, as well as the observed average spatial distribution of the field fragments, are compatible with the same picture.

The only cloud, of any consequence, looming on the horizon is the possibility that the rate of flux annihilation is higher than anticipated. Direct measurements are difficult and uncertain, but indirect evidence (stemming from the failure to appear of certain predicted long-term effects) certainly suggests that, in the long run, opposite polarity features have a significantly greater-than-random probability of recombining.

In spite of all this success, one would not be surprised if, when examined in greater detail, the data revealed that the diffusion constant required to explain, say, the separa-

tion, is slightly different from that required to explain the area, or the width of the neutral channel; or that a slightly different diffusion constant is needed for different-sized regions. Such deviations would simply indicate that there are processes other than completely random diffusion at work. All we can say at present, is that the numbers seem to be the same, at least within a factor ~ 2, and that, at least during those stages of decay which can be followed, diffusion is the dominant process.

A final point -- crucial to the acceptance of the random walk model -- and one which we cannot address adequately with the present measurements is that of whether the scatter which is found in the individual patterns of development (the part not caused by observational errors) can be properly explained by the variety of initial conditions (flux, separation, inclination, field distribution, and background) that are encountered. Intuitively, one feels that they are enough, but we obviously haven't proved it.

4. Fine-scale Motions in Decaying Regions

4.1 Introduction

The fact that the pattern of active region decay can be described as a diffusion-like spreading does not in any way prove that the underlying mechanism has to be a random walk of the field lines (even though we have used the words interchangeably in the previous section). Almost any spreading process would have similar gross properties. To show that there is really a random walk we have to try to examine the motion of individual magnetic features (over whatever period that is possible), and then ask ourselves: if those motions were extrapolated to what they would be after weeks or months, would they be sufficient to create the amount of spreading that is actually seen? Finally, should we find that they are sufficient, we would have to ask: could that motion be caused by the convective currents which are known to exist in the photosphere?

The latter question will be considered in § 5. For the moment we wish simply to discuss the motions, as they are observed, without worrying about what might cause them.

4.2 The Properties of a Random Walk

Just as two-dimensional diffusion was the framework within which we were able to discuss the evolution of the region as a whole, the two-dimensional random walk is the

most natural framework within which to discuss the expected motions of individual features.

In its simplest form, the two-dimensional random walk can be described as a process whereby an array of points is subjected to a pattern of motion such that in each interval of time τ, a particular point will undergo a linear motion by a distance L, in a direction which is random and uncorrelated with the direction of any of the previous steps. The most striking statistical property of the random walk, as of the diffusion process, is that the mean-squared displacement of a point from its point of origin increases linearly with time.

This property can be easily demonstrated, for if we assume that the displacement at time t is given by the vector $\vec{R}$, then at time $t+\tau$, the displacement can be written as;

$$\vec{R}(t+\tau) = \vec{R}(t) + \vec{r} \tag{4.1}$$

where $\vec{r}$ is the random vector of length L. By squaring equation (4.1) we obtain:

$$R^2(t+\tau) = R^2(t) + 2RL\cos\Theta + L^2 \tag{4.2}$$

where Θ is the angle between $\vec{R}$ and $\vec{r}$. Since Θ is equally likely to assume any value between 0 and 360°, if we average over a randomly-chosen selection of particles,

the cosine term will go to zero, and we are left with:

$$\langle R^2(t+\tau)\rangle = \langle R^2(t)\rangle + L^2 \tag{4.3}$$

In other words, the mean-squared displacement increases by L^2 in each time-interval of length τ. Comparing this with equation (3.8), we see that the random walk can be identified with a diffusion-like growth if we set:

$$D = \frac{1}{4}\frac{L^2}{\tau} \tag{4.4}$$

(this expression is sometimes given, incorrectly, with a factor of 1/2 instead of 1/4).

Over reasonable intervals of time it should be possible to represent the displacements of the individual particles by a probability function (cf. eqn. 3.7) of the form:

$$P(\delta)\,d\delta = \frac{2\delta}{\delta_0^2} e^{-\delta^2/\delta_0^2}\, d\delta \tag{4.5}$$

where $P(\delta)\cdot d\delta$ is the probability of having a displacement in the range $\delta \rightarrow \delta + d\delta$ (this is simply a Gaussian modified to account for the fact that there is more area available at the larger displacements).

Note: in the following discussion, the symbols δ, ℓ and r will be used interchangeably to represent the net displacement of a feature from its point of origin. Where the distinction is important, δ will be used in referring to the actual (or predicted) displacements, and ℓ in referring to

the measured (or apparent) displacements; but this convention is not used throughout. The symbol r will also, at times, be used to refer simply to distance. Hopefully the meaning will be obvious in context.

δ_o is the rms displacement, and for simple diffusion:

$$\delta_o = \sqrt{4Dt} \tag{4.6}$$

The fraction of particles which move less than any fixed amount, r , is given by:

$$f(\leq r) = 1 - e^{-r^2/\delta_o^2} \tag{4.7}$$

It is possible to exploit this property to construct a graph paper, which we might call "diffusion paper", analogous to ordinary "probability paper", on which the integrated distribution, when plotted, will produce a straight line. As we will have reason to use this technique shortly, it will be explained more fully in Appendix III.

Actually the properties of a random walk are not quite as simple as the discussion thus far may have implied. If one looks at the problem in too much detail, then it becomes apparent that the growth of $\langle r^2 \rangle$ with time depends not only on how far they move, but also on <u>how</u> they get from one point to the other (Figure 25). In general, the linear growth of $\langle r^2 \rangle$ is characteristic only of times long compared to the 'flight' or 'motion time', τ_1 $(0 \leq \tau_1 \leq \tau)$. Thus the most

Figure 25: The growth of $\langle r^2 \rangle$ with time in a random walk. In each of the four cases shown, the particles are assumed to move by a distance L in a time τ .

(a) and (b) both represent the average behaviour of a collection of particles starting at $t = 0$, but in the former case, the flights are assumed to be instantaneous (with the particles spending most of their time at the endpoints), whereas in the latter, the motion is assumed to be continuous (at a velocity $v = L/\tau$), giving a parabolic growth between endpoints.

(c) and (d) illustrate what happens when these two basic curves of growth are averaged over all possible initial phases (as we would expect to occur in the solar observations). Both patterns can be described as straight lines with slope $\frac{d\langle r^2 \rangle}{dt} = \frac{L^2}{\tau}$ for $t > \tau$, but in the latter case (d), due to a slow start, the actual displacements lag behind those of the former (b) by $\Delta\langle r^2 \rangle = \frac{L^2}{3}$.

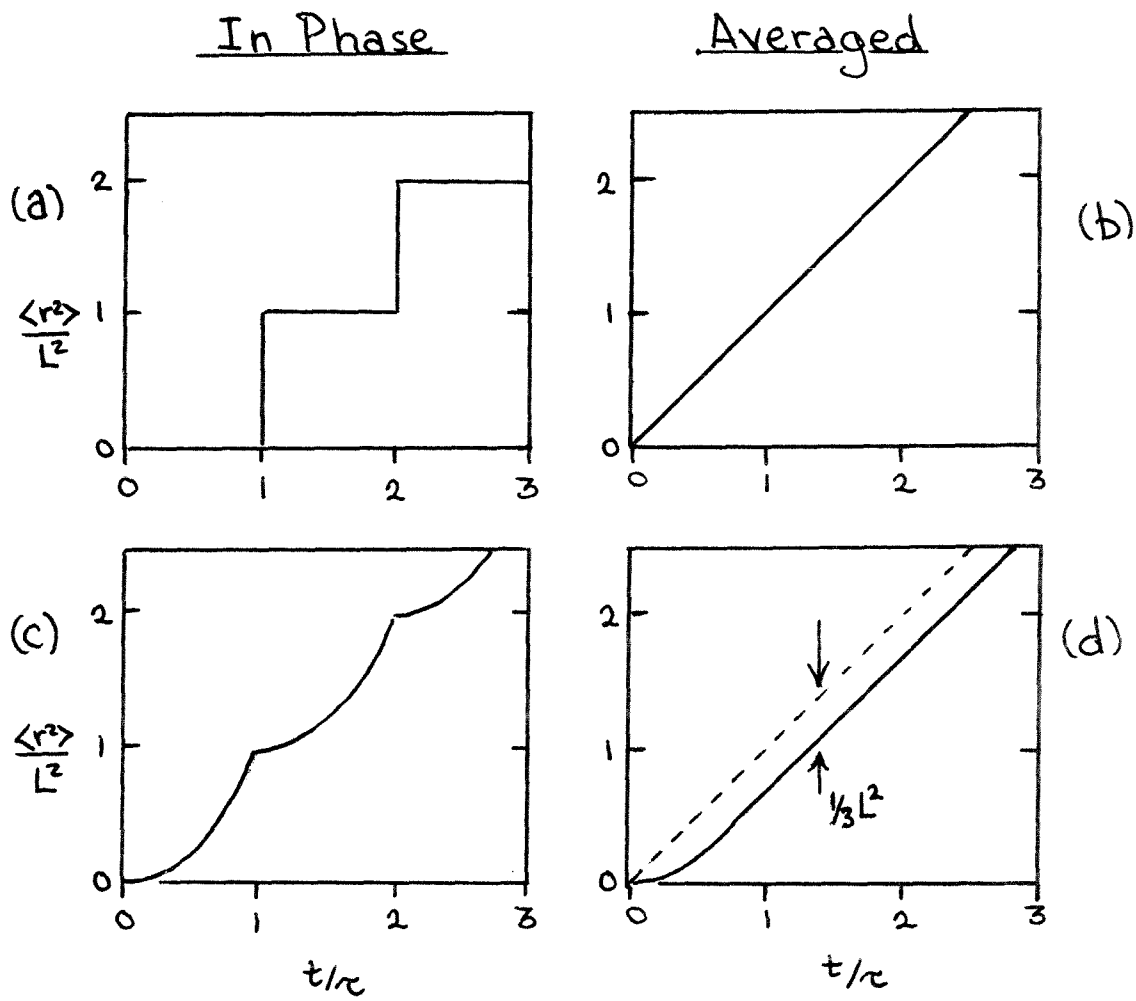


Figure 25

correct way to extract a diffusion constant from displacement data is to go by the slope at long times:

$$D_{eff} = \frac{1}{4} \frac{d\langle r^2 \rangle}{dt} \quad . \tag{4.8}$$

The diffusion constant will be somewhat underestimated if undue weight is given to the first few observations (as in the curved initial part of Figure 25d), when the random walk has not yet had time to 'randomize'.

If a variety of random processes, with different time scales, are involved, the curvature can continue for an extended period. Motions with a low velocity amplitude but long time constant (e.g., supergranulation) can eventually produce a larger diffusion costant than a similar motion with high velocity amplitude but short time constant (e.g., granulation) (see eqn. 4.4).

In other words, a process which is actually random (i.e., a random walk) can look non-random at short times. By the same token a process which looks random over the short run may become non-random when examined over the long. Figure 26 illustrates this possibility. Curve I is the standard random walk with its linear growth of $\langle r^2 \rangle$. For curve II this random motion has been superimposed on a steady linear expansion, which increasingly dominates the displacements at long times. The same upward curvature could also be caused by another component in the random walk, with a low velocity amplitude, but a long time constant

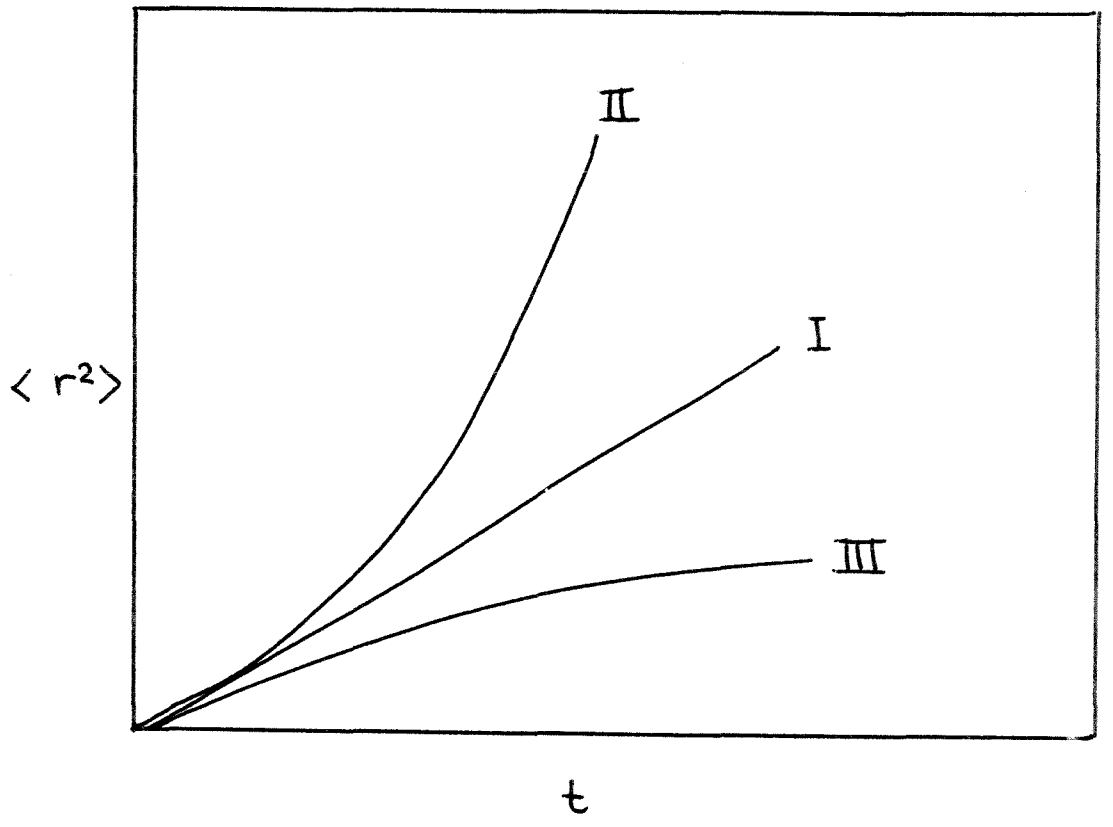


Figure 26: Characteristic patterns of $\langle r^2 \rangle$ with time.

(which might look like a systematic expansion over short times). Curve III is the case of the buoy tethered in the harbor. Over short times it seems to be moving about at random, but in the long run it can't ever get very far.

Any one of these behaviors, or some complicated combination of all three, could be found when the displacements of features on the sun are plotted.

4.3 Previous Work

The most thorough, and objective study of the motion of magnetic features in decaying plage regions, undertaken to date, is presumably that of Smithson (1972, 1973), in which the displacements were determined by marking, independently, the position of prominent features on magnetograms at two different times, 'rotating' the resulting dot patterns to a common time, overlaying, registering and measuring. According to Smithson (1973), for 21 'identifiable' magnetic features an rms displacement of 6870 ± 560 km was found in 24 hours, and for 16 'identifiable' features, a motion of 9118 ± 816 km in 48 hours.

These numbers are compatible with a linear growth of $\langle r^2 \rangle$ at $\frac{d\langle r^2 \rangle}{dt} \approx 500$ km^2/sec, corresponding to a diffusion constant $D \approx 125$ km^2/sec; which is somewhat lower than, but not entirely incompatible with our previous estimates of the diffusion constant (~ 200 - 400 km^2/sec) based on the long-term spreading rate of similar active region plages.

Furthermore, Figure 27, which is extracted from the data of Figures 6 and 7 in Smithson (1973), shows clearly that the observed distributions of displacements are not at all incompatible with the modified-Gaussian distribution expected for diffusion-like motions (eqn. 4.5), though perhaps somewhat truncated.

The total data seem also to suggest slightly higher rms displacements than those just quoted. The best fits to the histograms of Figure 27 (as determined by plotting the integrated distributions on diffusion paper) require:

$\delta_c = 8400 \pm 800$ km for the magnetic data (105 pts.)

and $\delta_c = 8700 \pm 800$ km for the K-line (126 pts.),

indicating $D \approx 210$ km^2/sec.

The problem with accepting these measurements as a definitive description of the short-term motions of magnetic features in a decaying plage is that about one-third of the points cannot be identified. Smithson himself suggests that this unidentified component may consist of flux which has been transported over distances in excess of 15,000 km, which could easily raise the mean-squared displacement by a factor of two or more. We do not necessarily accept this interpretation (an examination of his Figures 2-5 suggests that most of the 'unidentifieds', and apparent large displacements, may result from incomplete sampling around the edges

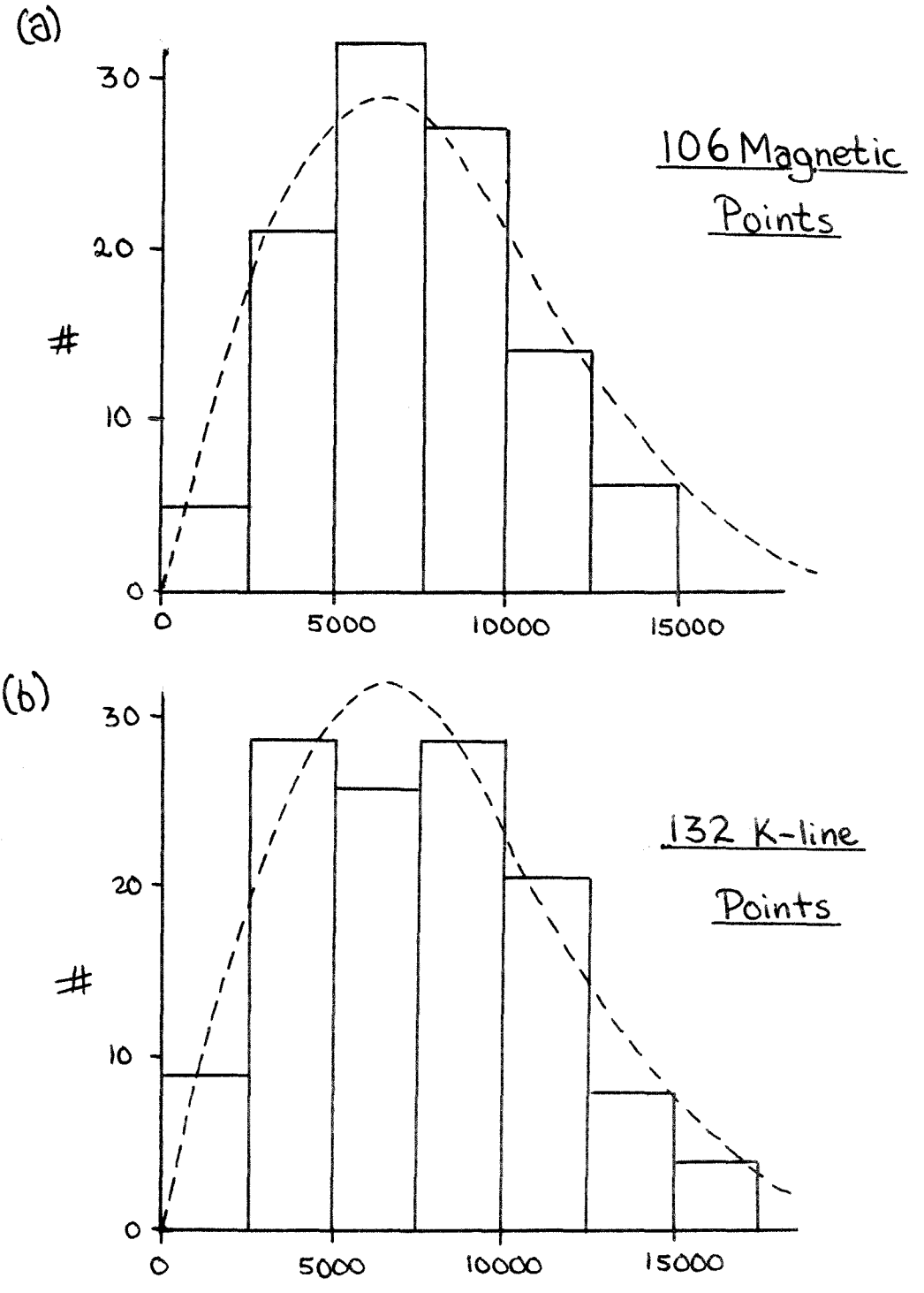


Figure 27: Smithson's 24 hour displacement data.

of the plage), and it seems important to attempt to extend the measurements to shorter times, where one can be more confident that definite, discrete features are being followed. Some additional comments about Smithson's data will be made in § 4.5 .

As far as shorter intervals of time are concerned, the only relevant study of which the author is aware is that of Schröter and Wöhl (1975), in which individual K-line brightness features were followed, in real time, over periods of 3 - 5 hours, by a method of photoelectric guiding. These authors chose to express their results in terms of a velocity distribution, which they found to have a 'Maxwellian' form, with $v_{rms} \approx 0.15$ km/sec. If we assume (and this is not entirely obvious) that on the average these velocities were obtained by taking the total displacement observed in ~4 hours, and dividing by the time, then we would infer:

$$(\Delta r)_{rms} = \delta_c \simeq 2160 \text{ km in } 4 \text{ hours.}$$

It is not entirely understood what size features this refers to, or exactly how the photoelectric system contends with short-term variations in form and brightness.

Finally, we have the statements of numerous observers (e.g., Dunn and Zirker, 1973; and Dravins, 1975) that, when examined at the highest possible resolution, the "magnetic filigree" is seen to be displaced by granule-like distances

in granule-like lifetimes -- say, ~600 km in 10 minutes. Again, there is some question as to the statistical significance of these numbers. I.e., are they peak values, or 'typical' values or rms values? and, exactly how were they determined? Even under the best seeing conditions the relative distances between features can change randomly, from frame to frame, by up to $\sim 1''$.

4.4 Attempts to Follow Individual Features

There are two basic reasons why the measurement of the displacement of solar magnetic features is a non-trivial enterprise. First, there is no way of distinguishing one field line from another. Second, the field patches are subject to short-term variations in 'visibility': a strong patch can spread and weaken, a weak diffuse patch can concentrate and strengthen; and patches of any strength can evolve both by merging and fragmentation, leaving the observer totally uncertain as to what actually happened in terms of the motion of individual field lines.

The detection of motion is least ambiguous when magnetic cancellations are used, but the resolution is poor ($\gtrsim 2''$), and internal changes in the form of a field patch are likely to take place in times comparable to that required for it to move by its own diameter. In addition, there are problems in dealing with marginal points dipping above and below the detection threshold of the instrument.

Higher resolution can be obtained by using observations (e.g., K-line or λ3840) in which the presence of the field is inferred from its heating effects. Unfortunately, along with the higher resolution comes more short-term variability. The apparent intensities both of the individual granule-sized field points, and also of the larger ($\sim$3000 km diameter) clumps into which they are organized can change erratically over times $\sim$10 minutes.

The present results (which are not much more satisfactory than the previous ones) are based primarily on a sequence of simultaneous magnetic and K-line observations made at Big Bear on July 2, 3 and 4, 1972 (Figure 28). The $4\hat{}\times6\hat{}$ field of view shows a "quiet" region, at a latitude of $\approx30^{\circ}$ N, crossing the central meridian on July 3. Approximately one-quarter to one-third of the frame is occupied by enhanced network of positive polarity, which is a remnant of the following plage of McMath 11857, born some $2\frac{1}{2}$ rotations earlier. Because of the relatively high latitude, the distortion of the field from day to day, due to solar rotation, is significant. To first order it is a simple linear shearing, with the equatorward edge of the frame appearing to rotate faster than the poleward side. When necessary, the shearing was removed by tracing the features at each latitude onto a sheet of transparent plastic which could be moved by a compensating amount.

Figure 28: Photographs of the weak field area studied on July 2-3, 1972.

Lower: Videomagnetogram (2345-0010 UT)

Upper: K-line filtergram (0010 UT; 0.6 Å bandpass)

The videomagnetogram is a photographic average of ten normal (256 frame) cancellations. North is at the top and solar rotation is from left to right. The width of each frame is approximately 2.5×10^5 km.

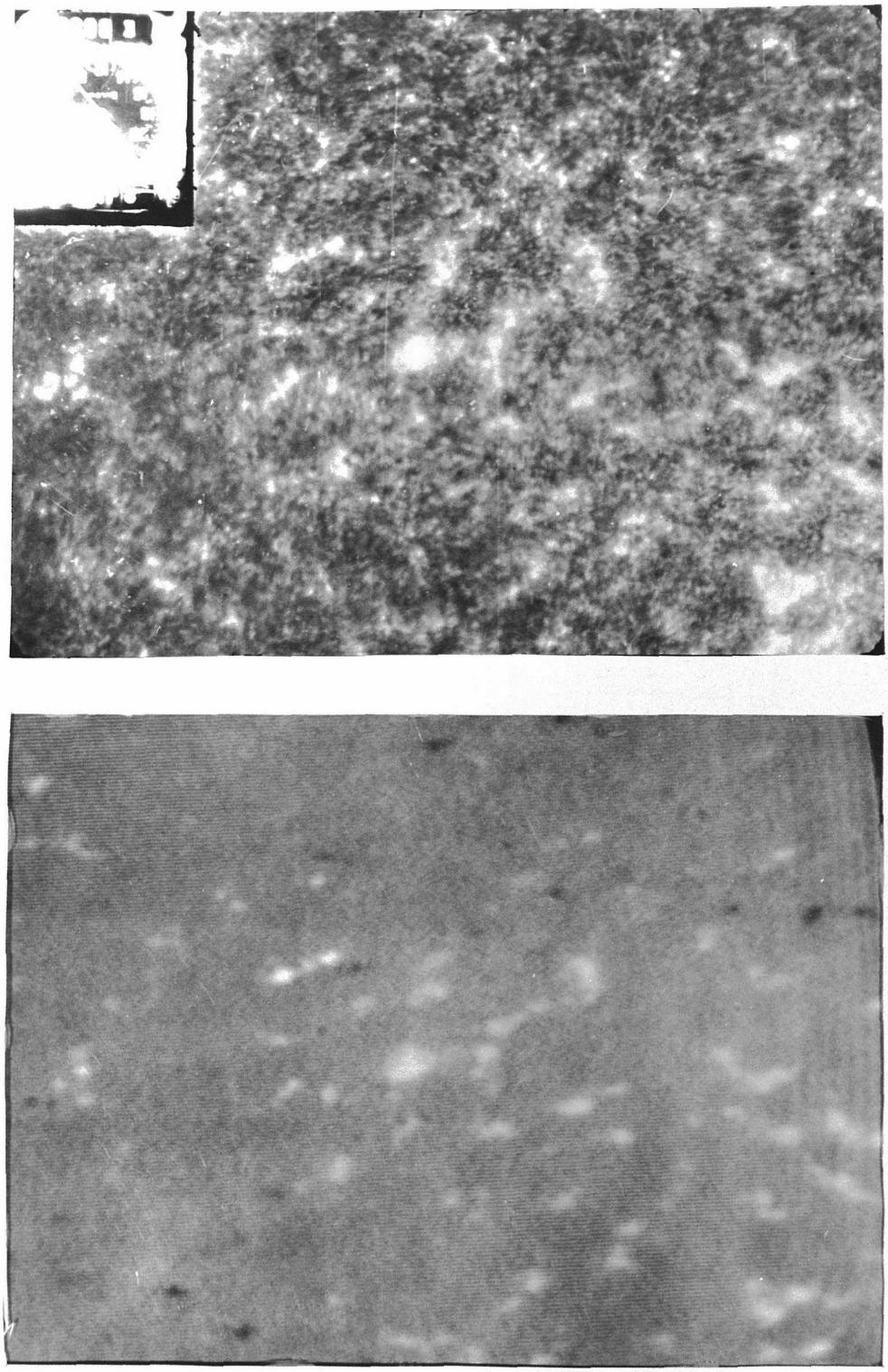


Figure 28

A number of attempts were made to extract the motions by direct measurement. These will be discussed only briefly, since widely different results can be obtained depending on the subjective judgements of the observer. An objective, and statistically more sound, procedure will be considered in § 4.6 .

4.4.1 Key Points: Separations and Displacements

In the first attempt, the total set of data for July 3 (i.e., magnetograms plus K-line filtergrams) was examined and an effort was made to define, subjectively, what appeared to be the motions of discrete "features", ignoring short-term variations in apparent strength, and going by the center-of-weight when fragmentation or merging seemed to occur. The position of approximately 30 prominent magnetic features was marked, in this way, on prints of four key K-line frames, at 1434, 2000, 2230 and 0100 UT, giving intervals of "5", 2½ and 2½ hours.

By overlaying the dot patterns (after correction for rotational shearing) lines could be obtained showing the path of each identified feature. The result is shown in Figure 29. The tails indicate the positions on the 1434 frame, and the subsequent dots correspond to the subsequent frames. Whether these curves even resemble the paths of individual field lines on the sun is open to considerable question; but one would suspect, if anything, that the finer features exhibit

Figure 29: The apparent motion of selected K-line features (July 3, 1972).

an even more random and erratic behavior, since a considerable degree of smoothing may be exercised when a frame is compared with earlier and later pictures, as was done here.

The mean-squared displacements are:

$\langle r^2 \rangle = (3.78 \pm .94) \times 10^6$ km^2 in $2\frac{1}{2}$ hours

$\langle r^2 \rangle = (8.78 \pm 2.2) \times 10^6$ km^2 in 5 - $5\frac{1}{2}$ hours

$\langle r^2 \rangle = (2.19 \pm .79) \times 10^7$ km^2 in $10\frac{1}{2}$ hours

which is quite consistent with the results of Smithson (1972, 1973) assuming a slight upward curvature to the graph of $\langle r^2 \rangle$ versus time. Excellent consistency is found when the first $2\frac{1}{2}$ hours is compared to the second $2\frac{1}{2}$ hours, and when the first "5 hours" is compared to the second.

It should be noted that here, as in Smithson's work, the absolute positions of the features on the sun is never determined. The somewhat arbitrary process of "registering" the dot patterns tends to produce a measured mean-squared displacement smaller than the true one by an amount of order (1/N), where N is the number of independent points in the sample. This is a correction small compared to the inherent uncertainties of the present study.

One of the most striking features of Figure 29, is that very few of the features are seen as executing a steady linear motion throughout the entire $10\frac{1}{2}$ hour period. Many appear to remain motionless, and of those those that move, most seem either to stop or to reverse direction at some point.

If we were to think of the supergranules as producing steady motions over ~20 hours, we would expect only about half the points to reverse direction, and the others to be steady. The observed behavior is clearly more erratic than that, and with (apparently) a relatively high proportion of 'stable' points. It is not, however, inconsistent with the idea that the magnetic features are wedged between supergranules, and that they can be disturbed by changes in any one of a number of neighboring cells. The maximum displacement in 10½ hours is 8400 km. The maximum displacement in 5 - 5½ hours is 5250 km. And the maximum displacement in 2½ hours is 4700 km.

A somewhat more objective measure of the degree of randomness in the motion can be obtained by considering the changes that occur in separations between particular pairs of points. Since there are ~500 different pairs in the sample, and since the measurements were being reduced by hand, only a small randomly-selected sub-sample was considered. A series of numbers, indicating the separation of each of these selected pairs on each of the four frames was obtained; and a scatter plot produced from them, showing the relative displacements in successive intervals of time (Figure 30). There is a slight tendency for points whose separation is increasing (decreasing) during the first 2½ hours to continue increasing (decreasing) during the second

Figure 30: Scatter plots showing the presence (or absence) of coherence in the apparent motions of the selected K-line features for 7-3-72.

(a) Change in separation over first 5 hours versus change in separation during second 5 hours.

(b) Change in separation during first 2½ hours versus change in separation during second 2½ hours.

All distances are in millimeters. 1 mm = 1050 km.

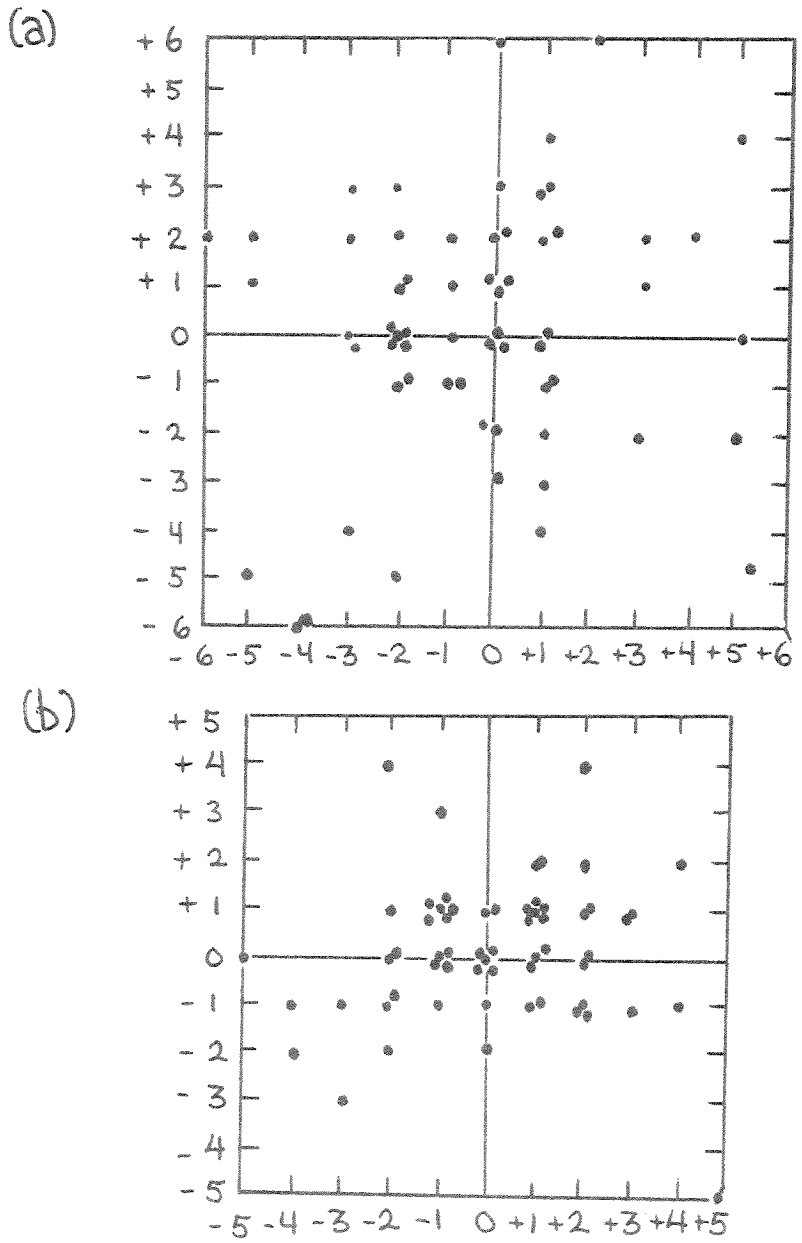

(a)
+6
+5
+4
+3
+2
+1
0
-1
-2
-3
-4
-5
-6
-6 -5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5 +6
(b)
+5
+4
+3
+2
+1
0
-1
-2
-3
-4
-5
-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

Figure 30

2½ hours, but there is no apparent tendency for points which moved apart (together) in the first "5" hours to continue doing so in the second. The 'flight time' is evidently rather short.

It is also possible to plot the change in separation against the separation, in an attempt to determine if there is any spatial cohesiveness to the motion. When this is done, we are unable to find any evidence that the relative motion of closely spaced pairs is any greater or less, on the average, than the relative motion of widely spaced pairs (such a correlation would be expected either if the fields were being moved about by a convective pattern larger than the supergranulation, or if there were a systematic expansion or contraction of the region). The method is not particularly sensitive, however, considering the small sample studied, and it would not be possible to rule out a systematic expansion (or contraction) on the order of a few percent, over the 10½ hours (but even if present, it could be caused by changes in focus). The best way to rule out the possibility of a true systematic expansion is to compare the diffusion constant for the 10 hour period ($\sim$150 - 200 km^2/sec for the present data) to that required to explain the spreading of a similar weak plage on successive rotations. If an expansion were present, it would be strongly felt in the long-term data and a much larger (effective) diffusion constant would be required.

A very careful reader may discover that the rms change in separation indicated by Figure 30 is approximately the same as the rms displacement indicated by Figure 29, and not enhanced by a square root of two (as one might have guessed since there is independent motion at two ends). This is an interesting point, but of such a mathematical nature that it will be relegated to Appendix IV. Separation data have the advantage that, to the extent that particular features can be identified, the effects of geometric shearing and differential rotation can be avoided by considering only the component of separation in the north-south direction.

4.4.2 Modification to Include Fragmentation

One has an uneasy feeling, in attempting to follow features "ignoring short-term variations", as just described, that the technique does not deal properly with the possibility of fragmentation, and that some field lines may move considerably more than the smoothed averages would imply. This is a great problem, since when fragmentation does occur, there is no way to be sure exactly what kind of a reorganization of the field lines was involved, nor that they will not (like the buoy in the harbor) return to re-form the same clump later on. Moreover, when one sees a feature fade at one point and something else strengthen nearby it is never clear if there was an actual transfer of flux between the two, or if the appearance is simply an accident caused by one point

(randomly) choosing to disperse a little, and the other randomly choosing to concentrate or emerge at about the same time.

Ignoring the problems, one can go ahead and try the experiment anyway, just to see what happens. Transparencies showing the appearance both of the K-line and the magnetic features, were made and laid on top of corresponding photographs taken at later or earlier times. What seemed to be the maximum reasonable amount of rearrangement of flux was estimated visually. Most of the time this was divided into two components: a motion of "clumps" (roughly the size of a supergranule wall, or "vertex point"), and an internal motion within the clumps. The two mean-squared components were then added to obtain a total mean-squared motion.

The results, which are at least as arbitrary and subjective as those obtained by trying to follow particular "key" features, are shown in Figure 31, along with the former ones. One would hope that the true rate of growth of r^2 with time is somewhere between the two limits, i.e.:

$$\left.\frac{d\langle r^2\rangle}{dt}\right|_{\text{observed}} \approx (5.5 - 17.0) \times 10^3 \text{ km}^2/\text{sec}$$

corresponding to $D \simeq (139 - 300)$ km^2/sec.

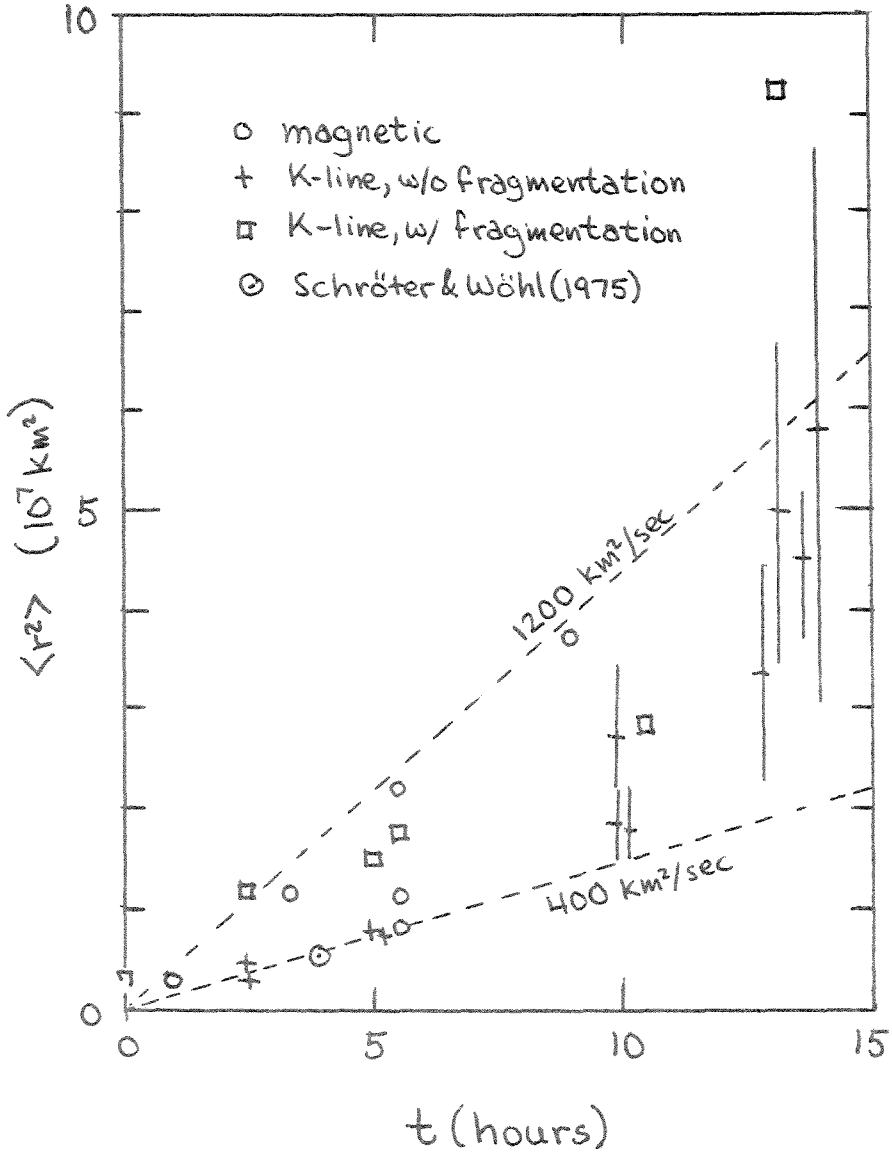


Figure 31: Collected results for visual impressions of mean-squared displacement versus time.

4.5 A Refinement of the Smithson Technique

4.5.1 Theory

In principle, it is possible to analyze the motions of any array of long-lived particles in a completely objective manner by marking, independently, their positions first at a time which we can call $t = 0$, and then at a time which we can call $t = t_1$, overlaying the patterns, and measuring (without bias) the distance from each point in the pattern at $t = t_1$ to its nearest partner at $t = 0$. If we can show that this distance, when pictures at a particular interval of time are compared, is greater than it would have been if the interval had been smaller, yet less than the value which it approaches at very long times; then it seems obvious that there must be a definite statistical procedure for extracting the most probable amount of actual motion, and a procedure which could be employed (even though the particles are indistinguishable) without invoking any vague (and highly questionable) notions concerning what constitute "reasonable" or "proper" motions.

In the largest sense of things, the effort to determine this "degree of correlation" by hand is rather pointless, since a far more elegant statistical procedure exists (§4.6); but it will be explained anyway, both because it has been used, and because it provides an intuitive framework within which the physical significance of cross-correlation functions

can be better understood.

Now, at short times (i.e., when the motions are small compared to the distance between features) the manner in which the displacements should be analyzed is fairly clear. To the extent that the sampling is complete, nearly all the points marked at $t = t_1$ will have an unmistakable partner at $t = 0$; and the observed distribution of displacements can be interpreted simply as reflecting the combined effect of real motions and plotting errors. In addition to these "normal" points, there would be a few "odd" points with unusually large separations from their nearest partners; but it would be clear that they were caused by those occasional instances in which a particular point had been inadvertently left out of one pattern or the other, and that they should be ignored in computing the mean-squared displacement. This intuitive approach is essentially the method used by Smithson in comparing the form of decaying plages over 24 hours. As the field weakens, however, the number of fragments which must be considered increases, and the identification of individual points becomes increasingly uncertain. We will, therefore, attempt to describe a more correct mathematical procedure.

Basically, we can say that we understand how to interpret the data if we can understand what distribution of measured displacements would result from any given distribution

of true displacements; and, in order to do this, there are two important effects which must be considered. First, there is the possibility that a particular point recorded at $t = t_1$ was accidentally omitted (or perhaps was not sufficiently prominent to be recorded) at $t = 0$. In that case it is likely that the apparent "displacement" will be much larger than the true motion. Secondly, there is the possibility that a feature at $t = t_1$, even though it was properly recorded on both frames, has moved so much that it will (by accident) be identified with the wrong point at $t = 0$.

It is worth noting that, at least in principle, there is no error introduced by the possibility of fragmentation when the 'objective' method is used, provided the sample is "complete", by which we mean that there is one dot for each independently moving feature. Imagine, for example, that a clump consisting of three independent flux tubes breaks up into its component parts, which proceed during the course of the observation to move, in different directions, by some definite amount. Since at $t = 0$ the features 'lie on top of each other', they would be represented by a single dot, and at $t = t_1$ by three, but the ambiguities come out in the wash: when the nearest neighbor distances are measured, each of the three points at $t = t_1$ is (properly) identified with the same initial point, and the correct set of displacements is recorded.

Unfortunately, complete sampling is never possible when data are being processed by hand; and even if it were, it would be extremely difficult to explain mathematically what should happen in a situation, such as that which exists on the sun, where the objects of interest are organized into a few relatively isolated clumps, composed of pieces which may or may not be capable of independent motion. In view of these difficulties, and since a better method is available, we will settle for describing the evolution of the nearest-neighbor distribution in the artificial, but illustrative, example of a fixed number of long-lived "points" subject to purely random motions, and placed at random as well.

The problem, then, is: given a two-dimensional random array with n_o points per unit area, and each point subject to independent displacements described by a probability density of the form:

$$P(\delta)\,d\delta = \frac{2\delta}{\delta_o^2}\, e^{-\delta^2/\delta_o^2}\, d\delta \tag{4.10}$$

(where δ_o is the true rms displacement); what will be the measured nearest neighbor distribution when the pattern at time t is laid over the pattern at $t = 0$, assuming that some fraction f of the features recorded at time t were inadvertently omitted from the pattern at $t = 0$?

This distribution is fairly easy to obtain, provided we know (see Appendix V) that the distribution of nearest neigh-

bor distances in a random array is given by:

$$n(r)dr = 2\pi n_o r\, e^{-n_o \pi r^2} \tag{4.11}$$

The key to the problem is to determine the probability, $P(r_o)$, that a point at time t will have a partner within distance r_o, or less, at time t; since then the probability of having a measured displacement in the range $\ell \to \ell + d\ell$ is given by:

$$P(\ell)\, d\ell = \left.\frac{d\,P(r_o)}{d\,r_o}\right|_{r_o = \ell} \tag{4.12}$$

Now the probability $P(r_o)$ can be determined if we can divide up the possibilities into manageable pieces. First, there is the chance that the true displacement was less than r_o, and the chance that it was greater. These are given by:

$$\int_o^{r_o} \frac{2\delta}{\delta_o^2} e^{-\delta^2/\delta_o^2} d\delta = 1 - e^{-r_o^2/\delta_o^2} \tag{4.13a}$$

and

$$\int_{r_o}^{\infty} \frac{2\delta}{\delta_o^2} e^{-\delta^2/\delta_o^2} d\delta = e^{-r_o^2/\delta_o^2} \tag{4.13b}$$

respectively.

If the true displacement is <u>greater</u> than r_o, then there is still the random chance:

$$\int_o^{r_o} n(r)\, dr = 1 - e^{-n_o \pi r_o^2} \tag{4.14}$$

of having a partner sufficiently close to satisfy the condition.

If the true displacement is <u>less</u> than r_o there is a probability $(1 - f)$ that the condition will be satisfied (i.e., if the point was properly marked at $t = 0$) and an <u>additional</u> probability:

$$f \cdot \left(1 - e^{-n_o \pi r_o^2}\right) \tag{4.14}$$

that even though the point was not recorded the condition will still be satisfied by accident.

By adding all these possibilities together we obtain:

$$P(r_o) = e^{-r_o^2/\delta_o^2}\left[1 - e^{-n_o \pi r_o^2}\right] + \left(1 - e^{-r_o^2/\delta_o^2}\right)\cdot\left[(1-f) + f\left(1 - e^{-n_o \pi r_o^2}\right)\right] \tag{4.15}$$

from which, by applying equation (4.12) we obtain, at last, the measured distribution of displacements:

$$P(l)\,dl = (1-f)\cdot\left(2\pi n_o + \frac{2}{\delta_o^2}\right)\cdot l\, e^{-\left(\pi n_o + \frac{1}{\delta_o^2}\right) l^2} dl + f\, 2\pi n_o l\, e^{-n_o \pi l^2} dl \tag{4.16}$$

In spite of its somewhat forbidding appearance, one will see that, on reflection, equation (4.16) has in fact a very simple and elegant form. It consists of two of the now-familiar modified Gaussian distributions (characteristic of random

displacements in two dimensions), where the true distribution (given by eqn. 4.10) has been perturbed by "noise". The noise serves both to broaden the distribution of "normal" displacements (the first term of eqn. 4.16), through the possibility of misidentification; and to contribute an independent background distribution (the second term of eqn. 4.16) through the possibility of omission.

Under conditions where the real motions are small compared to the spacing between points (i.e., $\delta_0 << \frac{1}{\sqrt{\pi n_0}}$) the second distribution will be much broader than the first, and this provides the mathematical justification for ignoring the high-displacement "tail", as was done in Smithson's work. There is still however a slight "saturation" effect which must be considered.

This effect can be explained as follows. If we take equation (4.16) and compute the mean-squared displacement which would be observed we obtain:

$$\langle l^2 \rangle_{obs} = \frac{(1-f)\,\delta_0^2}{1 + \pi n_0 \delta_0^2} + \frac{f}{n_0 \pi} \quad . \qquad (4.17)$$

Even at very long times ($\delta_0 \to \infty$) the observed displacements cannot rise above the background level set by the random distribution of nearest neighbors.

Equation (4.17) can be inverted to express the true mean-squared displacement in terms of the observed one:

$$\langle \ell^2 \rangle_{true} = \frac{\langle \ell^2 \rangle_{obs} - \frac{f}{m_0 \pi}}{1 - \pi m_0 \langle \ell^2 \rangle_{obs}} \qquad (4.18)$$

and this can be rewritten in the form:

$$\langle \ell^2 \rangle_{true} = \frac{\langle \ell^2 \rangle_{obs} - \langle \ell^2 \rangle_{\Delta t = 0}}{1 - \langle \ell^2 \rangle_{obs} / \langle \ell^2 \rangle_{\Delta t = \infty}} \qquad (4.19)$$

Equation (4.19) applies even when only the first term of equation (4.16) is considered, as is done when the high displacement tail in the histograms is ignored. In general, the omission of the high displacement points is the smartest thing to do, both because the estimate of f is uncertain (it may not even be completely constant), and because the background points would exert an undue leverage on the measured mean-squared displacements, which could only on the average be suppressed by a correction such as equation (4.13).

As an example of the application of our new formalism, we can consider the work of Smithson (1972, 1973), from which we surmise:

$$m_0 \simeq 1 \times 10^{-9} \text{ km}^{-2}, \quad f \simeq 0.3$$

$$\left.\begin{aligned} \langle \ell^2 \rangle_{\Delta t = 0} &\simeq 8 \times 10^6 \text{ km}^2 \\ \langle \ell^2 \rangle_{24 \text{ hr}} &\simeq 7.3 \times 10^7 \text{ km}^2 \end{aligned}\right\} \text{ ignoring high displacement tail}$$

$$\langle \ell^2 \rangle_{\Delta t = \infty} \simeq \frac{1}{m_0 \pi} \simeq 3.2 \times 10^8 \text{ km}^2$$

Inserting these values into equation (4.19), we deduce:

$$\langle r^2 \rangle = \langle \ell^2 \rangle_{true} = 8.7 \times 10^7 \text{ km}^2 \quad \text{in 24 hours.}$$

That is, the measured mean-squared displacement is (at least in theory) expected to underestimate the true mean-squared displacement by only about 20%.

There is serious question, however, about applying our method in this case. If the field is concentrated into poorly resolved clumps, and our dots represent only the center-of-weight of a number of independently moving field lines; then the motion of the dots might severely underestimate the true motion of the individual lines. Thus, in the following sub-section we will consider what happens when, at higher resolution, the sampling density is increased from ~1 point per supergranule to 5 or 10. We will find (surprisingly) that this increased thoroughness of sampling brings out little or no evidence for significant additional fine-scale motion; but the result is uncertain, because with the higher density of points comes an increased probability of misidentification, and at 5 - 10 points per supergranule it is only possible to follow the motions, reliably, for ~3 - 10 hours. As a result we will still be left in some confusion as to how to extrapolate the short-term motions to longer times: is the continuous, fine-scale, internal rearrangement of field within the clumps important to the long range

evolution of the pattern or not? Only in § 4.6 will we be able to provide sound statistical evidence to the effect that the fine-scale motions do not contribute significantly to the long-term large-scale rearrangement of field.

4.5.2 Application to the July 3, 1972 Data

The hand-sampling technique was applied to composite magnetograms (which brings out weak background detail) for July 2, 1972; to ordinary magnetograms for July 3, 1972; and to the 0.6 Å bandpass K-line filtergrams taken on July 3. The sampling was done by marking the strongest features first, and then moving down to less and less prominent ones, until at last a reasonable threshold seemed to have been reached. As the sensitivity was increased, an effort was made to assign additional dots to the original, strong features, so that, to some approximation, each dot could be thought of as representing a "quanta" of flux. This could not be done perfectly, however, and there is almost certainly some bias towards the weaker features.

In the area of the enhanced plage, the natural-seeming density corresponded to about 4.5 points per supergranule for the normal magnetograms (7-3-72), 5.5 points per supergranule for the composite magnetograms (7-2-72), and 12 points per supergranule for the K-line filtergrams ("supergranule" here means an area of 10^9 km^2). The high density required to describe the K-line features is probably due more to the

Figure 32: Histograms of observed displacements for July 3, 1972 magnetic data. The displacements are measured in millimeters (with a 1/2 mm increment). 1 mm = 1270 km.

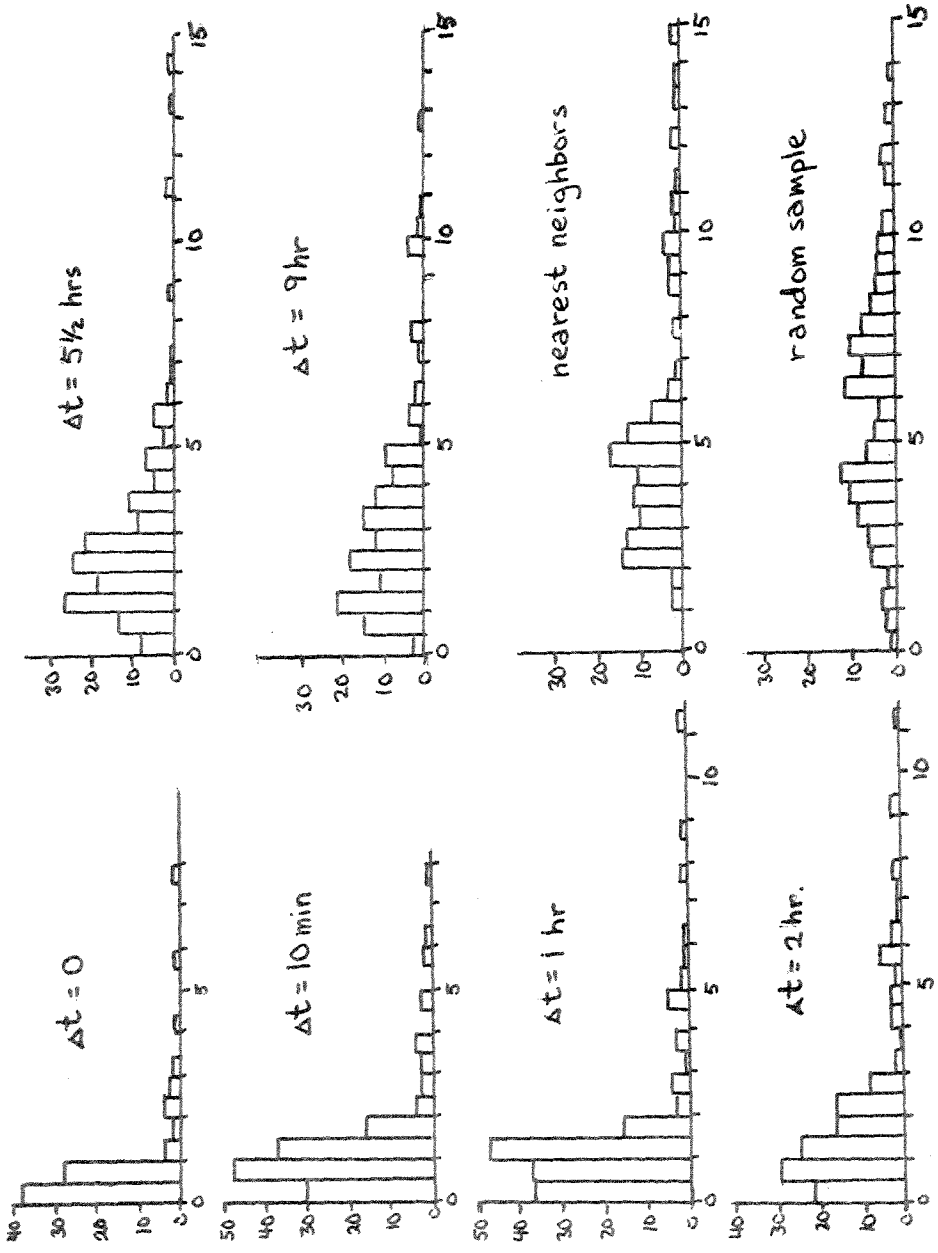

Δt = 0
Δt = 10 min
Δt = 1 hr
Δt = 2 hr.
Δt = 5½ hrs
Δt = 9 hr
nearest neighbors
random sample

Figure 32

confusion caused by short-term heating fluctuations, than by a genuinely lower sensitivity threshold; as some of the weakest isolated features on the magnetic composite appear to correspond to a single, unusually bright K-line "granule". A typical sample consisted of about 100 points.

For each series of dot patterns, the following quantities were determined:

(a) The density of points $\equiv n_0$.

(b) The actual distribution of nearest-neighbor distances on a single frame $\equiv \langle \ell^2 \rangle_{nn}$.

(c) The distribution of apparent displacements (nearest-neighbor distances) when dot patterns recorded at different times were superimposed in register $\equiv \langle \ell^2 \rangle_{obs}$

In each case, $\langle \ell^2 \rangle$ was determined by making a histogram of the measured displacements and fitting the standard modified-Gaussian to it, ignoring the extended random tail, as described in Appendix III.

The idea was to plug these numbers into equation (4.19) so as to obtain the true mean-squared displacement, but a problem immediately arises: the dots are not randomly distributed, they come in clumps. Most of the points (the ones in the clumps) have nearest neighbors closer than would be expected at random, but a few (the ones that happen to be in the middle of a cell) have no close neighbors at all. The assumption we used in deriving equation (4.15) -- that the

probability of having a neighbor at a certain distance is independent of where you happen to be -- is no longer valid. At short times the measured mean-squared displacements are limited to rather small values by the main nearest neighbor distribution, but for long times the displacements can be as great as $\langle \ell^2 \rangle_{random}$: the mean-squared distance to a nearest neighbor when the pattern is sampled from a random point. $\langle \ell^2 \rangle_{random}$ can be determined, empirically, by laying the dot pattern on a regular rectangular grid, and measuring the distance from each point on the grid to its nearest point in the pattern. At short times it is appropriate to use $\langle \ell^2 \rangle_{nn}$ as $\langle \ell^2 \rangle_{\Delta t=\infty}$ in equation (4.19); whereas at long times, $\langle \ell^2 \rangle_{random}$ is correct. In general, the problem of how the saturation level should be varied between the two extremes is a very complicated one and we will not attempt to solve it here, as it seems to be sufficient simply to assume that the proper value $\langle \ell^2 \rangle_{\Delta t=\infty}$, at any time, is greater than $\langle \ell^2 \rangle_{nn}$ but less than $\langle \ell^2 \rangle_{random}$. By the way, it is easy to show that when clumping is present:

$$\langle \ell^2 \rangle_{nn} < \frac{1}{n_o \pi} < \langle \ell^2 \rangle_{random} \quad . \qquad (4.20)$$

We present now the results:

Figure 32 shows the histograms of observed displacements for the July 3, 1972 magnetic data, with the characteristic progression from near-perfect overlap to randomness.

The observed and corrected mean-squared displacements were computed and are given in the following tables, where the corrected value has been computed once using $\langle \ell^2 \rangle_{random}$ and once using $\langle \ell^2 \rangle_{nn}$ as the limiting value in equation (4.19). To the extent that the method works, the true mean-squared displacement should lie between these two numbers. Most intervals were examined twice using different pairs of pictures.

The mean-squared displacements are plotted in Figure 33.

The numbers given in the tables are in units of 10^6 km^2.

Table 4.1 : July 2, 1972 Magnetic Data
Mean-squared Displacements

($n_o = 5.5 \times 10^{-9}$ km^{-2})

Time Interval	Observed	Corrected	Corrected
(nearest neighbors)	28.2 ± 5.9		
(random sampling)	63.7 ± 13.4		
5½ hours	6.96 ± 1.46	7.81	9.24

Table 4.2 : July 3, 1972 Magnetic Data
Mean-squared Displacements

($n_o = 4.5 \times 10^{-9}$ km^{-2})

Time Interval	Observed	Corrected	Corrected
(nearest neighbors)	45.1 ± 8.0		
(random sampling)	73.9 ± 12.5		
0	0.23 ± .05		
6 minutes	2.32 ± .24	2.16	2.20
13 minutes	2.10 ± .47	1.92	1.96
1 hour	2.52 ± .40	2.37	2.42
2 hours	4.39 ± .75	4.42	4.61
5¼ hours	11.7 ± 1.9	13.6	15.4
9 hours	15.0 ± 2.5	18.5	22.1

Table 4.3 : July3, 1972 K-line Filtergrams
Mean-squared Displacements

($n_o = 12.1 \times 10^{-9}$ km^{-2})

Time Interval	Observed	Corrected	Corrected
(nearest neighbors)	12.5 ± 2.2		
(random sampling)	29.6 ± 4.8		
0	0.34 ± .06		
4 seconds	0.34 ± .06	0	0
9 minutes	1.85 ± .29	1.61	1.77
30 minutes	1.37 ± .24	1.07	1.15
1 hour	2.93 ± .52	2.87	3.39
1½ hours	3.87 ± .76	4.06	5.12
3 hours	5.98 ± 1.09	7.07	10.8
5 hours	9.18 ± 1.82	12.8	33.7

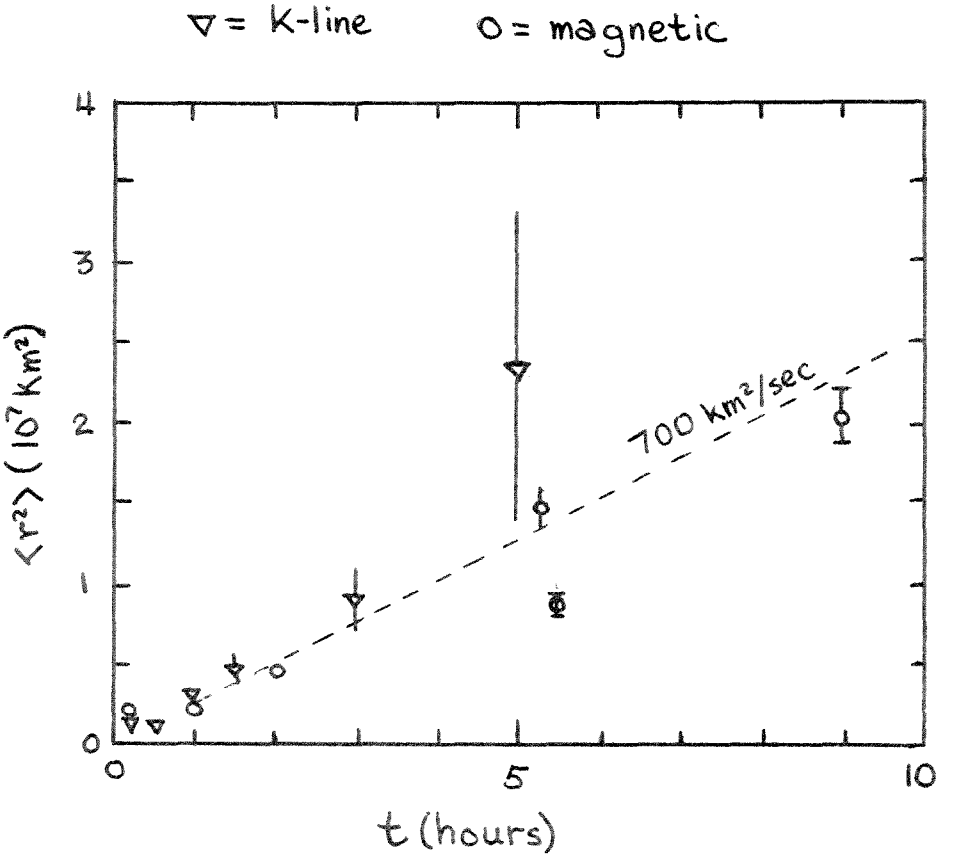


Figure 33: Objective measurements of mean-squared displacement based on nearest-neighbor method. Data in Tables 4.1-4.3 .

As stated in the introduction to this section, one seems to find, over those intervals where a comparison is possible, essentially the same result independent of the sampling density that is used. The K-line data may show a slightly more rapid increase of $\langle r^2 \rangle$ with time, but the deviations come mainly towards the end of the interval when the saturation correction is most uncertain. Both the K-line and the magnetic data show $\langle r^2 \rangle$ increasing rapidly (over 5 - 10 minutes) to a value of about 2×10^7 km^2, remaining steady at that value for about an hour, and then developing a steady rise. Overall, the growth seems to be at about 700 km^2/sec -- corresponding (cf. eqn. 4.8) to $D \approx 175$ km^2/sec -- and there is little evidence (aside from one highly uncertain K-line point at 5 hours) of any extended upward curvature of the kind which we might expect if the dominant random walk mechanism involved steady motions over periods of more than a few hours (cf. Figure 25). If Smithson's (corrected) point at 24 hours is included, a slight upward curvature <u>is</u> indicated giving a final slope of perhaps as much as

$$\frac{d\langle r^2 \rangle}{dt} \approx 1200 \text{ km}^2/\text{sec}, \quad \text{i.e.,} \quad D \approx 300 \text{ km}^2/\text{sec}.$$

The interpretation of the plateau between $t = 0$ and $t \simeq 1$ hour is something of a puzzle. The most obvious interpretation would be that it is just a slight residual "noise" caused by plotting errors and seeing fluctuations, but, at least for the K-line data, this does not seem to be

the correct explanation. Seeing changes totally in less than a second, yet Table 4.3 shows clearly that a K-line frame taken at $t = 4$ seconds is more similar to the one taken at $t = 0$ than is one taken at $t = 9$ minutes.

The second most obvious explanation would be that it is a physical effect of the kind anticipated in § 4.2 . One could say: aha! this is the effect of the granulation which can push the field around for a few minutes, but which is restrained, by the supergranulation, from ever pushing it very far. Unfortunately, as we will see in Section 5, if the granulation interacts with the field in at all the same way as the supergranulation, then one would expect it to displace the field, in a granule lifetime, by at most a tiny fraction of a granule diameter; yet, according to the tables, the rms fluctuation in position of the dots is about 1400 km in 6 - 9 minutes.

We are left with one remaining plausible possibility: that we are seeing fluctuations in the localized heating effects which determine the visibility of the field at any particular instant. Over periods of 5 - 10 minutes, there is evidently a tendency for one portion of a field patch to dim and another to brighten, so that one assigns the position of the peak field in a slightly different way, even though there may actually have been no motion of flux involved. The presence of such temperature fluctuations is certainly

not in contradiction of one's impression on viewing the K-line filtergram movies. On magnetic movies, the features seem quite steady in intensity, but they change in shape and area.

As a final note, we should mention that an effort was made to apply the hand-sampling technique to photographic prints of Kitt Peak full-disk magnetograms showing very quiet unipolar regions near disk center on successive days. A point density of $\sim 3 \times 10^{-9}$ points/km^2, seemed sufficient to represent most of the flux; and at such a density, if the motions are as small as we claim, it should be feasible to identify most of the features after 24 hours. In spite of the slight correlation which was visible to the eye, the observed nearest-neighbor distribution at Δt = 24 hours, was much closer to the ultimate, completely random distribution of distances than expected. This is disturbing, but it is very likely due in large part to the small scale (1mm $\simeq$ 8000 km) of the prints which were used. The cross-correlation functions, which are to be discussed in the next sub-section, will provide a much more objective indication of the amount of rearrangement which takes place in weak-field areas over 24 hours.

4.6 Extraction of the Diffusion Constant by Cross-correlation

4.6.1 Introduction

The simplest intuitive procedure for showing that there is correlation between two photographs is not to draw dots showing the positions of a few of the most prominent features, and to try to show that those dots lie closer together than they would have at random; but rather, it is to take transparencies of the photographs themselves, and to lay those on top of each other. If there is a correlation, then there will be a definite position in which the overlap of features, though perhaps not perfect, is at least distinctly "non-random".

This intuitive method of measuring the correlation can be made completely quantitative and objective by measuring the transmission of light through the sandwiched pair. The transmission will be highest when the overlap is most perfect, and, to the extent that the sample is statistically homogeneous (that is, to the extent that the surroundings have the same general appearance as the area being studied), it will fall to a definite background level when the "plates" are slid far out of register. The height of the resulting "correlation curve" is a measure of the degree of non-random overlap, and its width is a measure of the "correlation distance". When the pictures are the same, the tracing is called an "auto-correlation" curve, and the correlation distance is

interpreted as a measure of the size of the features. When they are different, it is called a "cross-correlation".

If a time sequence of photographs is being considered, it is customary to express the residual correlation found when a picture at time t is compared to one at $t = 0$ as a fraction of the "perfect correlation" found when two pictures with $\Delta t = 0$ are compared. Generally this number is found by dividing the height of the cross-correlation curve by the height of the auto-correlation curve (which <u>is</u> the cross-correlation at $\Delta t = 0$). Sometimes the areas under the curves are also used.

For a more precise mathematical description of the photographic correlation method, and for a consideration of the effects introduced by film contrast, non-linearities, background fog, etc., the reader is referred to Noyes (1963) and Simon (1963). Numerous other references exist.

Leighton's (1957) measurement of the lifetime of granulation was one of the first applications of this technique to the study of evolving solar features. We will describe that work briefly now; both because it provides a useful illustration of the way in which the method works, and because there may have been some tendency on the part of others to carry ideas which work very well under one set of circumstances, over into a quite different application where a new interpretation is more appropriate.

Empirically, one finds that the cross-correlation curves obtained between granulation plates taken over different intervals of time are all about the same width, but that their height falls off rapidly as the interval is increased. If we think of the granules as ephemeral features which come and go primarily by changes in brightness (as opposed to changes in position), then the interpretation of this behavior is straightforward. The background level to which the correlation heights are referred is found either by comparing two identical plates when they are placed out of register, or by comparing two plates taken with a long interval between, using any relative orientation. Mathematically it is determined by the condition that a given bright (dark) feature on the first plate will have a purely random chance of being overlain by a bright or dark feature on the second. At places where a granule "died", the average transmission drops automatically to the background level, so that the residual transmission above the background that is found when the plates are registered is simply a measure of the number of surviving granules (or, more precisely, of the fraction of the original number which survive).

The width of the curve is a measure of the size of the granules (or at least their brightness pattern) and it is not expected to change very much (as long as the granules don't shuffle around during their lifetime) because the size

of the ones surviving after 5 minutes would not be expected, on the average, to be either much smaller or much larger than the ones at $t = 0$ (note that the width is due solely to those surviving features which correlate, and is not affected by the random background).

Now, the photographic correlation technique can obviously be applied equally well to pictures showing magnetic features changing with time, but in making that application we must reconsider what we are doing. The field lines, or flux tubes, of which the magnetic pattern is composed cannot, like granules, be thought of as temporary features that can disappear and never be heard from again. The average field line, though it may occasionally fluctuate in visibility, will persist, in some form or another, for times very long compared to those over which the "visual pattern" changes (see §6.1). The loss in correlation found when magnetograms taken some number of hours apart is, then, caused not, primarily, by the disappearance of the original objects, or by the emergence of new ones (in the sense of granules being born and dying); but rather, it is caused, primarily, by the <u>spatial rearrangement</u> of a fixed number of long-lived features.

While this may, in one sense, complicate our interpretation of the cross-correlation curves; it will, at the same time, provide us with exactly what we have been trying to find. If we imagine the flux tubes as being subject to random

displacements, then (provided we consider only times ≥ 10 minutes, over which the short-term variations in visibility have had a chance to smooth themselves out) we should be able to predict in a precise quantitative manner: given an initial pattern, how will the correlation decline as that pattern fuzzes out into a new and unrelated one?

In general this might seem to be a very complicated problem, considering all the possible initial patterns which might be encountered; but, as we shall see, there is, for all practical purposes, only one property of the initial pattern which affects the rate of decline of correlation, and that property is its "width" or "correlation distance". Thus, if our theory works, it should be possible to extract the diffusion constant in a precise and unambiguous manner, since there will be only one diffusion constant which, in conjunction with the measured correlation width is capable of reproducing the observed decay of correlation with time.

The photographic procedure is superior to the hand-sampling technique in inumerable ways. Not only is the sampling complete and unbiased; but, to the extent that the "signals" from the individual features add together linearly, and to the extent that the correlation can be detected, the errors due to accidental overlap, clumping, and finite resolution, can all be properly accounted for and corrected simply by: (1) comparing the transmission in register to

the transmission out of register; and (2) using the empirically determined correlation width in the formulas. The ability to properly incorporate these corrections would permit even the extraction of a statistically significant diffusion constant for areas, such as an active region plage, in which the packing of lines is so dense that the identification of individual features is uncertain after any finite time.

The only real disadvantage of the correlation data is that the actual distribution of displacements (as opposed to the overall mean-squared value) is hidden in the shape of the correlation curve, along with a number of other parameters, such as the size and shape of the individual features, and the extent of large-scale ordering. Even this, however, may not long be a disadvantage, since the ability to produce correlation curves with high-speed scanning digital microphotometers permits the analysis of those shapes at a level of sophistication that was never before feasible.

Unfortunately, to the author's best knowledge, the photographic cross-correlation technique has never been applied to any of the vast number of high-resolution photographic images of solar magnetic fields obtained in the last ~15 years; perhaps because the potential significance of the results has not been fully appreciated.

In the following paragraphs we shall attempt to clarify that significance, and to show how the results, at least for quiet regions, can already be anticipated on the basis of the "network lifetime" studies of Simon (1963) and Rogers (1970). The ability of the derived formulas to reproduce the broad qualitative features of their observations, and to account quantitatively for the numerical differences between them, must be regarded as impressive evidence for the validity of thinking of the cross-correlation data in terms of a diffusion model.

4.6.2 Theory

4.6.2a The Movable Disk

The key to understanding the auto- and cross-correlation functions is to understand how they work in a specific example, and one of the easiest examples to think about is that in which the photograph being studied consists of isolated circular patches, or disks (these could be the flux tubes, smeared out by heating effects and seeing to a finite size).

Consider what happens when two such images are placed on top of each other. Intuitively, it is easy to see that if the disks are opaque, and if the rest of the plate is clear, then twice as much area will be covered up when the images are out of register as when they are in register. It is this dip in transmission which generates the correlation curve. It is also clear that the distance over which

the plates must be moved in order to go from one situation to the other is on the order of the radius of the individual disks. There may seem to be some question about the exact background level, because, depending on the size of the disks and the number per unit area, there will be some probability, even when out of register, that a disk on the first plate will accidentally overlap one on the second; but this probability of random overlap is the same for every pair of plates in the series, and it will be subtracted in computing the correlation heights. There may also seem to be some question about the overall normalization of the curve -- which, surely, depends on the contrast of the film and the background density -- but since we are concerned here only with fractional correlations, that, too, will cancel out.

The only thing we do have to worry about is the possibility that the disks are organized into a larger-scale pattern, with some degree of non-random ordering. For example, if they are scrunched together into a few isolated clumps, then the distance which has to be moved to go from "in register" to "out of register" is the size of the clumps, rather than the size of the individual disks, and that can change the way the correlation develops with time. This is a point which is crucial to understanding the interpretation of the solar data, and we will come back to it shortly, but for the moment let us suppose that the disks are placed at

random.

For the reasons given above, the only non-random effect capable of causing the transmission to rise above the background level (when the plates are placed in register) is the probability that a disk overlaps with the original image of itself (we are assuming that the intensities add linearly, so that the overall solution can be found by a superposition of the individual solutions). In other words, to compute the shape of the auto-correlation function, all we have to do is to calculate $A(x)$, the area over which a circle of radius r_o overlaps with itself when the centers are displaced by some distance x. This is easily done, and the result is:

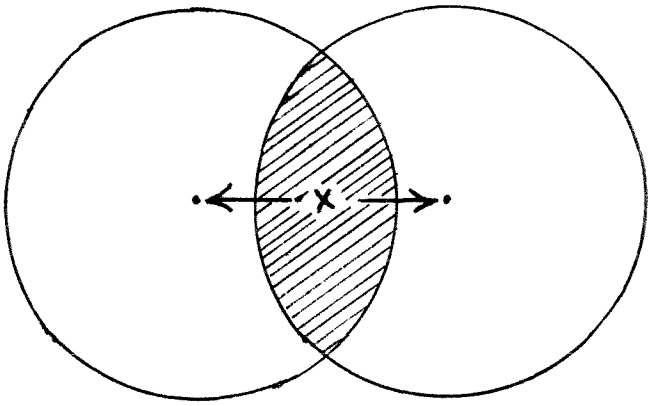
$$A(x) = 2r_o^2 \cos^{-1}\left(\frac{x}{2r_o}\right) - x\sqrt{r_o^2 - \frac{x^2}{4}} \qquad (4.21)$$

This function is plotted in Figure 34. It is very nearly a triangle, and can be approximated by:

$$A(x) \doteq \pi r_o^2 \left(1 - \frac{x}{1.7r_o}\right) \quad \text{for } 0 \le x \le 1.7r_o$$
$$\doteq 0 \quad \text{for } x \ge 1.7r_o \qquad (4.22)$$

The photographic technique will pick out this function <u>independent</u> of the density of patches, and <u>independent</u> of any larger-scale (random) splotchiness which the eye may see. The full width at half maximum of eqn. (4.21) reflects <u>only</u> the size of the component disks:

(a)



(b)

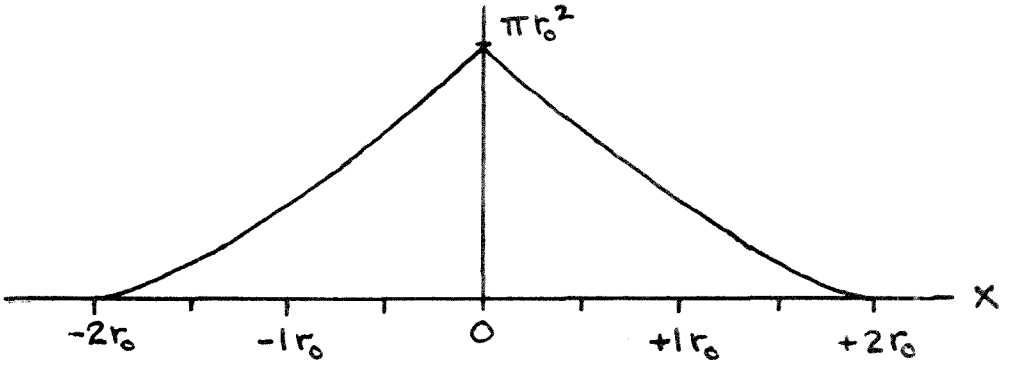


Figure 34: Overlap between displaced circles

$$(FWHM)_{AC} = 1.6\, r_o \tag{4.23}$$

So that if the size of the disks were not known beforehand, it could be obtained from the width of the auto-correlation curve. The result in eqn. (4.23), by the way, is considerably more general than it might seem. Simon (1963) has shown that the factor of 1.6 relates the width of the correlation curve to the "size" of the features for a wide variety of reasonable geometric shapes.

By the same token as we were able to predict the <u>shape</u> of the auto-correlation curve by considering the overlap of a single disk with itself, we can predict the <u>height</u> of the cross-correlation curve (in the presence of random motions) by calculating the average overlap which a disk will have with itself if, in the second image, it is subject to a random displacement.

For simple diffusion (cf. eqn. 3.7), the distribution of displacements, δ, after time t goes like:

$$P(\delta)\,d\delta = \frac{\delta}{2Dt}\, e^{-\frac{\delta^2}{4Dt}}\, d\delta \tag{4.24}$$

If the disk in the second image is subject to such displacements, then we can use eqn. (4.22) to calculate, approximately, the average overlap which it will have with the original image (we are assuming they are placed in register):

$$\langle A \rangle = \int_0^{1.7 r_0} \pi r_0^2 \left(1 - \frac{x}{1.7 r_0}\right) \frac{x}{4Dt} e^{-\frac{x^2}{4Dt}} \, dx . \tag{4.25}$$

The integral is straightforward, and if we divide by πr_0^2, then we will have expressed the average overlap at time t as a fraction of the overlap at $t = 0$:

$$CC = \frac{\langle A \rangle}{\pi r_0^2} = 1 - \frac{\sqrt{\pi D t}}{1.7 r_0} \, \mathrm{erf}\left(\frac{1.7 r_0}{\sqrt{4Dt}}\right) . \tag{4.26}$$

This, then, is our fundamental result for the fractional correlation of a collection of randomly-moving circular disks with their image at $t = 0$. The correlation starts at unity at time zero (meaning that the overlap is perfect), and drops to zero at long times (meaning that there is nothing left but the random overlap, and hence that the cross-correlation curve has zero height). At short times the correlation can be expressed as:

$$CC(t) \sim 1 - \frac{\sqrt{\pi D t}}{1.7 r_0} \qquad \left(\sqrt{4Dt} \ll r_0\right) \tag{4.27}$$

At long times, it goes like $1/t$:

$$CC(t) \sim \left(\frac{1.7^2}{12}\right) \frac{r_0^2}{Dt} \tag{4.28}$$

Roughly speaking, the time it takes for the correlation to fall off significantly is on the order of the time it

takes for the particles to move by their own diameter. More precisely (for eqn. 4.26);

$$CC = 0.5 \quad \text{when} \quad \sqrt{4Dt'} = 0.98\, r_0 \tag{4.29}$$

where $\sqrt{4Dt'}$ is the rms displacement of the disks.

Unlike in the case of granulation (§ 4.6.1) where the correlation curve simply fell in height without growing in width, the curve here <u>does</u> increase in width, indicating that the pattern is changing by (on the average) "fuzzing out", rather than by the physical disappearance of features. To explore this quantitatively one would want to calculate the <u>shape</u> of the cross-correlation function, which is to say the fractional overlap when the disk in the second image is subject to random displacements about a point <u>off-set</u> from the original center by an amount x. This is a very messy problem, and we will give only the asymptotic result at long times:

$$CC(x) = \frac{\langle A \rangle}{\pi r_0^2} = \frac{(1.7)^2 r_0^2}{12Dt}\, e^{-\frac{x^2}{4Dt}} \tag{4.30}$$

that is, the width of the cross-correlation curve at long times is proportional to the rms displacement of the particles. More precisely:

$$\text{FWHM} = 1.665\, (\Delta r)_{rms} \quad \text{for } \sqrt{4Dt'} \gg r_0 \tag{4.31}$$

Comparing this to eqn. (4.23), and considering that random things usually add up by the sum of the squares, one would not be surprised to find that in general the width of the cross-correlation went like:

$$\text{FWHH} \approx 1.65 \sqrt{r_o^2 + 4Dt} \qquad (4.32)$$

but we have not proved this.

In summary, we have found that we can deal in a quite precise mathematical way with the correlation of moving disks, provided that they are placed, initially, at random. Unfortunately, the objects on the sun in which we are interested (the flux tubes) are not placed at random, but rather are very definitely bunched together in non-random clumps. Can the formalism which works for random initial conditions be modified to account for non-random ones, and if so, how? (The reader may be reminded of the frustrations of trying to answer this question for "hand-sampling").

While there may in general be no simple perfectly-correct solution to the problem, there is, at least, considerably more reason for optimism in the present case. The fundamental flaw in the hand-sampling method was that in the presence of clumping, the probability of "accidental overlap" with a neighboring feature depended on where you were in the picture, which in turn depended on the pattern and on how much diffusion had taken place, so that it became unclear what to use,

at any particular time, for the limiting or "random" nearest neighbor distribution. In the present case, the probability of random overlap is explored empirically, by placing the plates out of register, and the result is automatically subtracted in determining the height of the curve. Thus we completely avoid the basic problem which caused our downfall before (giving us, perhaps, a clue as to how we could "correct" the hand-sampling method if it ever seemed worthwhile to resurrect it). In addition, the correlation curves give us a quantitative estimate of the importance of the non-random effects, for the width of the auto-correlation curve reflects both the size of the "features" and the size of the "clumps".

Let us, then, consider what effect the presence of a non-random initial clumping would have on the decay of the correlation height, in the case where pattern is evolving by simple diffusion. There are two obvious considerations. First, if there is a significant amount of clumping, then the crucial distance over which objects have to move in order to affect the correlation is the size of the clumps rather than the size of the individual features; and if the clumping were only partial, then it would not be surprising to find that the mixture of clump sizes and particle sizes represented by the width of the auto-correlation curve is the proper mixture to use as the "critical" distance (cf. eqn. 4.29). The second consideration is that if the clumps

are made up of a number of independent parts (whose displacements have the diffusion-like form) is it the distance traveled by the individual parts which has to be comparable to the size of the clumps? or is it the center of weight of the clump as a whole (which would move much more slowly) that has to go that far?

We propose the following solution: if we are considering a pattern generated by the superposition of any number of independently moving particles, whose displacements are described by an effective diffusion constant D, then even though those particles be initially organized in a non-random fashion, the decline in correlation will for all practical purposes, still be described by eqn. (4.26), with only one modification -- that for r_o we use not the true size of the independent features, but instead, the effective correlation distance as defined by eqn. (4.23). That is:

$$CC(t) \approx 1 - \frac{\sqrt{\pi D t'}}{1.7 r_o} \operatorname{erf}\left(\frac{1.7 r_o}{\sqrt{4 D t'}}\right) \qquad (4.33)$$

where

$$r_o \equiv \frac{(FWHM)_{AC}}{1.6} \qquad (4.34)$$

In other words it is only the motion of the independent particles that counts, and when that motion is comparable to

the size of the features, as defined by the width of the auto-correlation curve, it will destroy the correlation.

Our "proof" will consist of two examples in which the initial distribution is non-random, and in which fragmentation of the constituent particles is permitted.

4.6.2b The Gaussian Swarm

As our first example we will consider what has to be about as opposite an extreme as there can be from randomly placed particles: namely a situation in which all the particles are bunched together into a few discrete sources from which they are released at time zero, being permitted to swarm out in accordance with the diffusion equation. The density about each injection point (cf. eqn. 3.6) can then be written in the form:

$$n(x,y,t) = \frac{N_o}{4\pi D(t+t_o)} e^{-\frac{x^2+y^2}{4D(t+t_o)}} \qquad (4.35)$$

where the parameter t_o has been introduced to give the distribution a finite size (which we might visualize as being comparable to the size of the observed network patches) at $t = 0$.

In order to determine the cross-correlation we define the overlap function:

$$g(\Delta,t) \equiv \iint_{x,y=-\infty}^{+\infty} n(x+\Delta, y, t)\, n(x,y,t=0)\, dx\, dy \tag{4.36}$$

where we have introduced both a spatial offset, Δ , and a time interval t . From our previous discussions, it is clear that the normalized cross-correlation curve is related to the overlap function by:

$$C(\Delta,t) = \frac{g(\Delta,t) - g(\Delta\to\infty,t=0)}{g(\Delta=0,t=0) - g(\Delta\to\infty,t=0)} \tag{4.37}$$

$C(\Delta, t)$ being the fractional correlation between a plate at time t and the one at time zero, when placed with an offset Δ from the point of perfect registration. Again, the integrals are straightforward, so we give only the result:

$$C(\Delta,t) = \frac{2t_o}{t+2t_o}\, e^{-\frac{\Delta^2}{4D(t+2t_o)}} \tag{4.38}$$

which exhibits the characteristic (and expected) features of a curve decaying and broadening with time.

The correlation coefficient is the value of equation (4.38) at the point of zero offset:

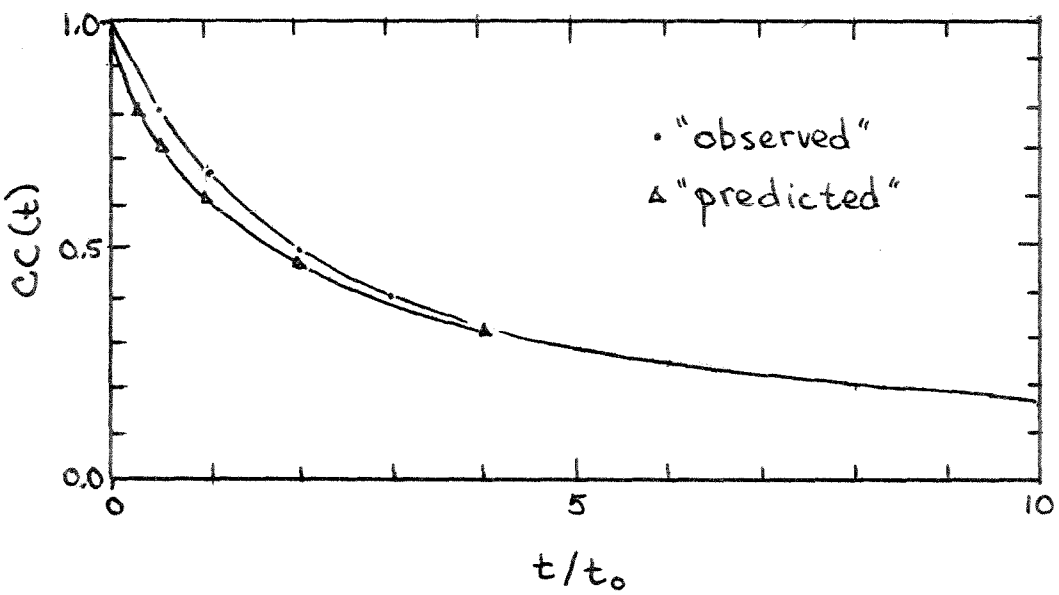
$$CC(t) = \frac{2t_o}{t+2t_o} \tag{4.39}$$

Functionally, this doesn't look very much like equation (4.33), but to make the comparison objectively we have to follow through with the second part of our program. That is, we claim that the present "observed" correlation (equation 4.39) will be described by equation (4.33) provided we examine the width of the "observed" correlation curve (equation 4.38) at $t = 0$, and base our value of r_o (in equation 4.33) upon it (cf. eqn. 4.34). Now the full width at half maximum of equation (4.38) at $t = 0$ is: $2.35\sqrt{4Dt_o}$. So according to equation (4.34) we should use $r_o = 1.39 \times \sqrt{4Dt_o}$. Thus the "predicted" cross-correlation for the "Gaussian swarm", based on our proposed rule, is:

$$CC(t) = 1 - 0.375\sqrt{\frac{t}{t_o}}\,erf\left(2.362\sqrt{\frac{t_o}{t}}\right). \tag{4.40}$$

The numerical values of "observed" (eqn. 4.39) and "predicted" (eqn. 4.40) correlation curves are compared in Figure 35. The two functions are remarkably similar. Equation (4.40) consistently underestimates equation (4.39), but never by more than 10%.

Thus, as stated, we find that for two extremely different situations -- one in which a few isolated features of a definite size are placed at random, and another in which the particles are released in great unresolved globs -- the predicted decay of correlation with time is described by essentially the same mathematical curve, provided only that the



Fugure 35: The Correlation Curve for a Gaussian Swarm

constituent particles move independently and at random, and that we use for the characteristic distance the auto-correlation width. In one case the auto-correlation width is the size of the moving features, and in the other it has no relation to them.

4.6.2c The Fragmented Disk

To demonstrate that the preceding result is more than some strange pecularity of the Gaussian distribution we consider one additional example: that in which a number of circular disks, of the kind discussed in §4.6.2a, are bundled together, at $t = 0$, so as to form a larger disk of radius R_o. We shall attempt to show that if it is the smaller particles which are moving, then it is the time for them, and not the center-of-mass, to move by R_o which is important.

Let us suppose, then, that each small disk is subjected to diffusion-like motions about its original position in the bundle. After any time t, some fraction of the particles, having executed their random walk, will continue to overlie the original position of the bundle, whereas the rest will have moved outside its boundaries. Clearly, the fractional correlation between the pattern at time t and that at $t = 0$ is simply the fraction of the original particles which remain within the original boundary. In principle, this is a quantity which is easy to calculate, since each particle simply

has an expanding Gaussian-like probability distribution (eqn. 4.24) about its point of origin, but in practice, trying to integrate a Gaussian over an offset circle is hardly trivial. It is, however, possible to investigate the asymptotic cases of very short and very long times.

Without going into details, the short time problem is essentially that of calculating the probability of "evaporation" from a surface. Only particles close to the surface have a chance of escaping, and the probability for them is easily calculated because when you're close the wall looks long and flat. The result is:

$$CC(t) \sim 1 - 1.128 \frac{\sqrt{Dt'}}{R_o} \qquad \text{for } \sqrt{4Dt'} << R_o \tag{4.41}$$

At long times, the Gaussian probability functions are so fuzzed out that the probability density near the origin can be regarded as a constant, and the integration becomes simply a matter of multiplying by the area of the "acceptance circle": πR_o^2. The result is:

$$CC(t) \sim 0.25 \frac{R_o^2}{Dt} \qquad \text{for } \sqrt{4Dt'} >> R_o \tag{4.42}$$

This is the "observed" correlation.

For the "predicted" correlation, we have to look at the shape of the $t = 0$ auto-correlation curve, whose width will tell us the value of r_o to use in eqn. (4.33). Obviously, the auto-correlation function is going to look like eqn.

(4.21) and the full width at half maximum will be 1.6 R_o. That is, the only thing the sliding plates show is the main clump --none of the fine structure. By equation (4.34) we find (not surprisingly) that we should use $r_o = R_o$ in eqn. (4.33). Thus the "predicted" correlation is:

$$CC(t) = 1 - \frac{\sqrt{\pi D t'}}{1.7 R_o} \operatorname{erf}\left(\frac{1.7 R_o}{\sqrt{4 D t'}}\right) \tag{4.43}$$

which at short times (cf. eqn 4.27) goes like:

$$CC(t) \sim 1 - 1.043 \frac{\sqrt{D t'}}{R_o} \quad \text{for } \sqrt{4 D t'} \ll R_o \tag{4.44}$$

and at long times (cf. eqn. 4.28) like:

$$CC(t) \sim 0.241 \frac{R_o^2}{D t} \quad \text{for } \sqrt{4 D t'} \gg R_o \tag{4.45}$$

Comparing equations (4.44) and (4.45) with equations (4.41) and (4.42) one sees that the asymptotic forms of the "predicted" and "observed" correlation curves (for the fragmenting disk) have precisely the same mathematical form, and very nearly the same coefficients. One is slightly higher at short times, the other is slightly higher at long times. At most the deviations are a few percent. Thus once again, we have found that the loss in correlation with time is caused by the motion of the independent fragments (of which the pattern is composed) over distances comparable to the size of the clumps (as seen by the auto-correlation device).

4.6.2d Generalized Initial Conditions and Conclusions

It is possible to generalize the results of § 4.6.3c to include situations in which, at time zero, the fragments are organized not in circular bundles, but rather in the form of figures of arbitrary shape with perimeter p and area A. In that case, the asymptotic forms are given by:

$$CC(t) \sim 1 - 0.564 \frac{p}{A} \sqrt{Dt} \qquad \text{for } Dt \ll A \tag{4.46}$$

and

$$CC(t) \sim 0.0796 \frac{A}{Dt} \qquad \text{for } Dt \gg A \tag{4.47}$$

Clearly, it would not, in general, be possible to pick a single value of r_o (in equation 4.33) which will precisely imitate both of these asymptotic conditions; but it is also clear that the value of r_o dictated by equation (4.34) will represent a reasonable compromise. The presence of an intricate fine structure (such as we would probably expect to find only in very high-resolution magnetograms of young active region plages) causes mainly a slightly more-rapid-than-average decline in correlation at short times, and a somewhat less-rapid-than-average decline at long times.

If a higher level of sophistication is required, then it would, in principle, be possible to compute (numerically) the average effects of diffusion on the actually observed initial configuration, and to deduce the result by comparing

the observed correlation curve to those generated by various values of the diffusion constant.

Such a degree of sophistication would be superfluous here, however. For our present purpose, it is quite sufficient to use equations (4.33) and (4.44)

4.6.3 Results

The principal published data to which the results of §4.6.2 can be applied are the "network lifetime" studies of Simon (1963; Simon and Leighton, 1964) and Rogers (1970). Simon obtained cross-correlations between K-line spectroheliograms of selected quiet regions near disk center "at various wavelengths within $\Delta\lambda \simeq 0.2$Å of line center" over intervals of up to 50 hours. Rogers used a 60 hour sequence of high-resolution off-band H-alpha filtergrams ($H_\alpha + 0.65$ Å) to obtain cross-correlations over intervals of up to 30 hours (see Figure 36).

Rogers indicates that the full width at half maximum of the correlation curves is 3780 km at $\Delta t = 3$ sec, and that it increases monotonically, reaching 7670 km at $\Delta t = 4$ hours. Simon does not specifically give the half-width of any of his cross-correlation curves, but states elsewhere that the full-width at half maximum for auto-correlation curves of comparable spectroheliograms taken in K_{2v} is 6000 - 7000 km, and 9000 - 10,000 km in K_3; expressing the belief that the latter figure is closer to the "true" network size (the

former being somewhat influenced by the prominent granulation-like fine structure that is seen in the wings). It is possible that some of the plates used in the cross-correlation experiment were of poorer quality.

Numerically, the results obtained by Simon and Rogers are quite different. Rogers shows the correlation dropping rapidly to about 25% over ~4 hours, and then declining more slowly. Simon shows a gradual decrease over ~20 hours, and then an even more gradual decrease for the rest of the observation period.

Now there are obviously other effects present, but to first order it certainly seems plausible that the principal agent causing the decline in correlation (at least for times $\gtrsim$10 minutes), is the general, continuous, rearrangement of field lines; since in one case it is the bright "flocculi" which are being correlated, and in the other, the dark spicules -- both of which phenomena are well-known to be intimately associated with the location of magnetic flux (Leighton, Noyes, and Simon, 1962; Simon and Leighton, 1964).

Surprisingly, it is possible to simply take the data at face value, using $r_o = 2400$ km for Rogers and $r_o = 5000$ km for Simon (as dictated by eqn. 4.34), and obtain a not at all unreasonable fit to the data. Simon requires $D \simeq 200$ km^2/sec. Rogers takes $D \simeq 300$ km^2/sec (if one is permitted to over-

Figure 36: Cross-correlation data for K-line and H-alpha network features.

(a) Simon's K-line data, fit by $r_o = 5000$ km and $D = 200$ km^2/sec (dashed), and by $r_o = 6000$ km, $D = 200$ km^2/sec (solid -- normalized to 0.8 at $t = 0$).

(b) Rogers' H-alpha data, fit by $r_o = 2400$ km and $D = 300$ km^2/sec (dashed, and by $r_o = 3500$ km, $D = 200$km^2/sec (solid -- normalized to 0.5 at $t = 0$).

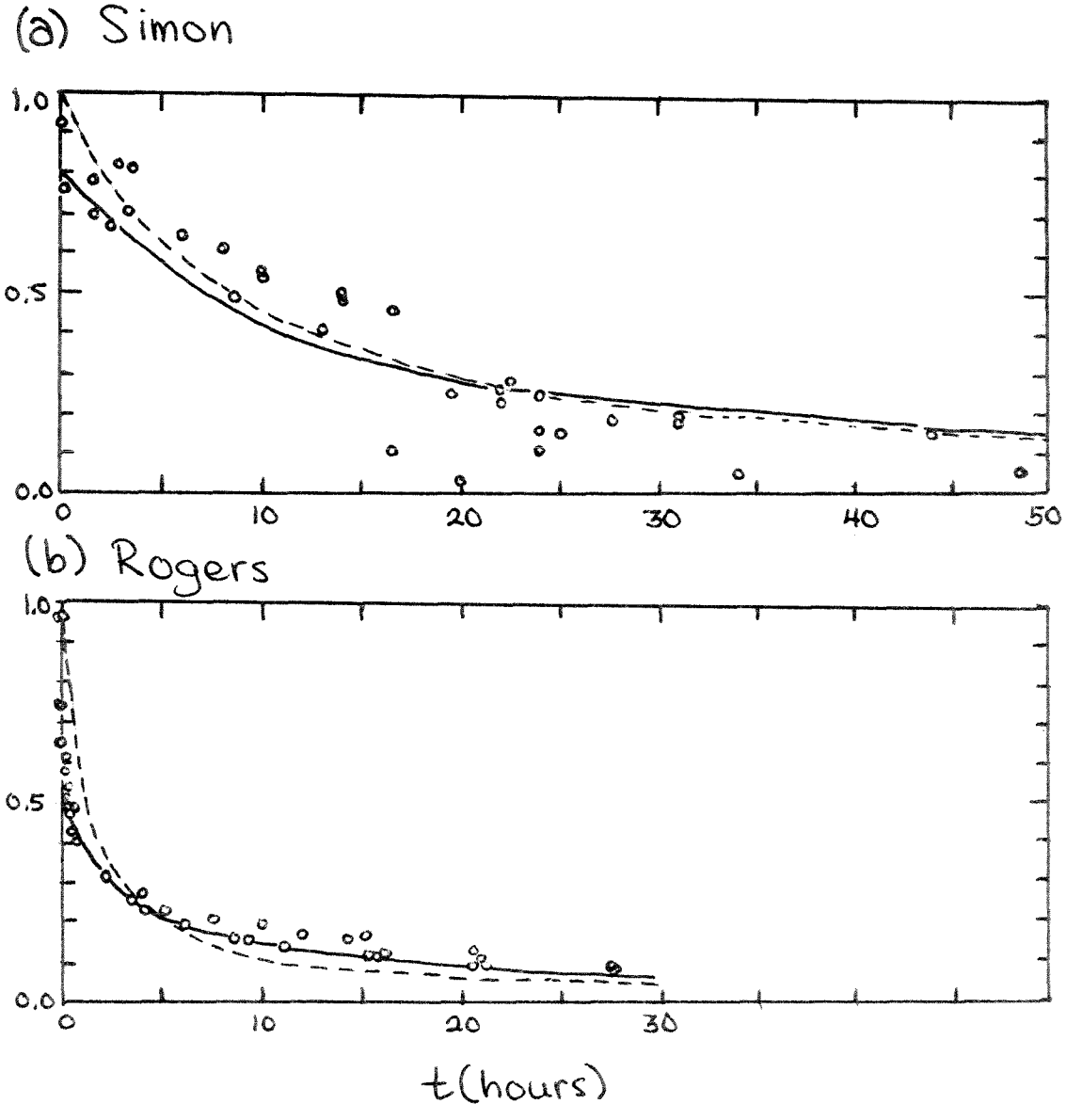


Figure 36

estimate the correlation a little at short times, and underestimate it at long times).

Such success is too much to ask for, however, because it is well known that the flocculi fluctuate in brightness and that the spicules come and go over times $\sim$10 minutes; and surely most of the initial very rapid drop in correlation (to $\sim$80-85% for Simon, and 50-60% for Rogers) is caused by these "visibility fluctuations" rather than by an actual motion of field lines. It is only the data at $t \gtrsim 10$ minutes, for which the fluctuations have had a chance to smooth themselves out, that we would expect to accurately reflect the more gradual motions of the sites of floccula and spicule production (i.e., the magnetic plages).

Our discussion of the expected correlation curves has not dealt explicitly with the possibility of visibility fluctuations (which could be largely avoided by using magnetic cancellations), but it is fairly clear that the only significant modification that would be necessary would be that in evaluating r_o (eqn. 4.34) one should use not the width of the auto-correlation function (which is influenced by the brightness fine-structure), but rather the width of the cross-correlation at $t \simeq 10$ minutes (which shows mainly the size of the non-fluctuating, average pattern).

Unfortunately, neither author tells us what this size is. Based on the limited information available to us, we

would tend to guess that the correlation width is ~9600 km for Simon (i.e., close to the K_3 size) and ~5600 km for Rogers (i.e., somewhere between the 3 second and the 4 hour widths) giving $r_o \approx 6000$ km and 3500 km, respectively. Using these values in equation (4.33), and normalizing the curves to 0.8 and 0.5 at $t = 0$, one obtains very reasonable fits using $D \approx 200$ km^2/sec (Figure 36).

Actually, because of the form of equation (4.33), it is only the value of D/r_o^2 which is being determined. Thus, if our assumptions about r_o are incorrect, a somewhat different value of D might be indicated.

There is one important effect which we have not so far included, and that is (cf. § 4.2), that if the motions are being generated by a random walk, the effective diffusion "constant" would be expected to increase slightly over times short compared to the 'flight time'. This might, in part, explain the "shoulder" in Simon's data, which causes the early points to lie a little above the predicted curve. The somewhat lower-than-expected correlation coefficients at $t \gtrsim 20$ hours, could also, in part, be due, as he suggests, to problems with geometric distortion; although the pattern should, by then, be sufficiently fuzzy that precise registration is less important than it was initially.

A partially independent estimate of the diffusion constant can be made by comparing Rogers' width data at 3

seconds and 4 hours. If we assume that the width grows as indicated by equation (4.32), then the total effective mean-squared displacement, 4Dt, including both real motions and fluctuations in the spicules, could not have been more than $\sim 1.64 \times 10^7$ km^2; corresponding to a diffusion constant $D \lesssim 284$ km^2/sec over 4 hours.

The ability of the diffusion model to reproduce the main qualitative features of the data of Simon and Rogers, and to account, in a quantitative fashion, for the differences between them is very encouraging; and it would be most interesting to see what the cross-correlation technique says when applied to magnetic photographs of regions in various states of decay.

4.7 Conclusions Regarding the Short-Term Motions of Magnetic Features

In summary, we have investigated the dispersal of magnetic features over periods of up to $\sim$ 20-50 hours, using three different techniques:

(1) Direct (subjective) estimates.

(2) Objective sampling of prominent features.

(3) Cross correlation.

The results of all three methods are consistent with a diffusion constant somewhere around $D \simeq 200$ km^2/sec.

The subjective estimates indicate a diffusion most probably in the range $D \simeq 139 - 300$ km^2/sec, depending upon

how much fine-scale fragmentation one is willing to recognize. The "objective" sampling tends to indicate values closer to $D \simeq 175$ km^2/sec, over 10 hours, but possibly as high as $D \simeq 300$ km^2/sec if we include Smithson's data at 24 hours. There is little or no evidence of steady motions (in quiet regions) lasting more than ~ 5 hours, so the short-time process can be properly characterized as a random walk, but in neither case is it clear that the long-range effects of fine-scale motion and fragmentation have been properly dealt with.

The cross-correlation technique provides an objective indication of the total motion of the independent field fragments. It has not yet been applied to the magnetic data, but preliminary results, based on published cross-correlation studies of H-alpha and K-line photographs, suggest that a total effective diffusion constant of $D \simeq 200$ km^2/sec is appropriate in "quiet" areas. In other words, it appears to be only the gross motions of the obvious field patches (which can be easily followed by hand) and not the fine-scale distortions and fragmentations, which lead to the long-range dispersal of field.

We reject the idea put forward by Smithson (1972, 1973) that the most important mechanism in field dispersal is a "squirting" process <u>distinct</u> from the "normal" random walk. Occasional motions over distances ~ 5000 km in a few hours

(say once every 6 hours) are not (cf. eqn. 4.4) incompatible with an overall diffusion constant of $D \approx 200 - 300$ km^2/sec; nor (cf. § 5) with the kinds of instabilities which might be generated by changes in the supergranulation (as Smithson himself points out).

The short-term diffusion constants found above are in excellent agreement with the active region spreading rates measured in Section 3. It is felt, that because of the widely different scales and sources of the data employed, and because of the numerous internal checks, this agreement must be regarded as something more than an observational bias. It appears, therefore, that a random walk of field lines, with "annihilation" in the places where an overlap of opposite polarities occurs, is the dominant process governing the geometry of a decaying active region, at least over periods of up to many months. Within the accuracy of the data, there is no reason for invoking the presence of any additional systematic mechanism (such as expansion or contraction) to explain the geometry; there may, however, (see Section 6.6) be reasons for suspecting significant non-random effects in connection with the eventual long-term disappearance of field lines.

5. The Dispersal of Field by Granulation and Supergranulation

5.1 Introduction

Having established that the 'random' daily motions of field lines in decaying active regions are (if extrapolated to greater times), in all probability sufficient to explain their long-range spreading and eventual "death", we are compelled to ask if that motion could be caused by the convective currents which are known to exist at the surface (i.e., the granulation and supergranulation) or if it necessarily reflects a connection of the lines to deeper layers of the sun.

5.2 The Bending of Flux Tubes

Before attempting to estimate the amount of diffusion that would be caused by the convective motions, we must first ask whether they are capable of causing motion at all. This may seem like a strange question, since it is well known (cf. §3.1) that the field lines are thoroughly "frozen" in a plasma whose conductivity is as high as that of the photosphere; but it is a necessary question, because the field lines have a tension, and if the tension is sufficient, it can (without violating the "frozen-in" condition) force the moving fluid to flow around the flux tube, leaving it virtually undisplaced.

To consider the problem of field line bending, we draw a picture (Figure 37) -- not unlike Figure 11 -- in which a section of length ℓ of a flux tube of radius r is subjected to the flow of an ionized fluid at density ρ and velocity v. We estimate the force acting on it (assuming it tries to remain stationary) in exactly the same way we might estimate the force of a wind blowing against a house, or of the water beating against a stick in a stream. The force is generated by the condition that the fluid must alter its momentum in order to avoid the obstruction, and since the amount of material (per second) which strikes the tube is proportional to the velocity, the pressure will be proportional both to the density and to the velocity squared. To compute the force, we multiply the pressure by the surface area which the tube presents to the flowing fluid. Hence:

$$F_{fluid} \simeq (\tfrac{1}{2}\rho v^2)(2r\ell) \tag{5.1}$$

There should, of course, be an additional coefficient of some sort (a "drag"), but since we know very little about either the true fine-structure of the tube or of its interaction with the surrounding fluid, we will simply assume that it is one of the order of one.

The only force which the tube can offer in opposition to the fluid flow is the tension. According to Cowling

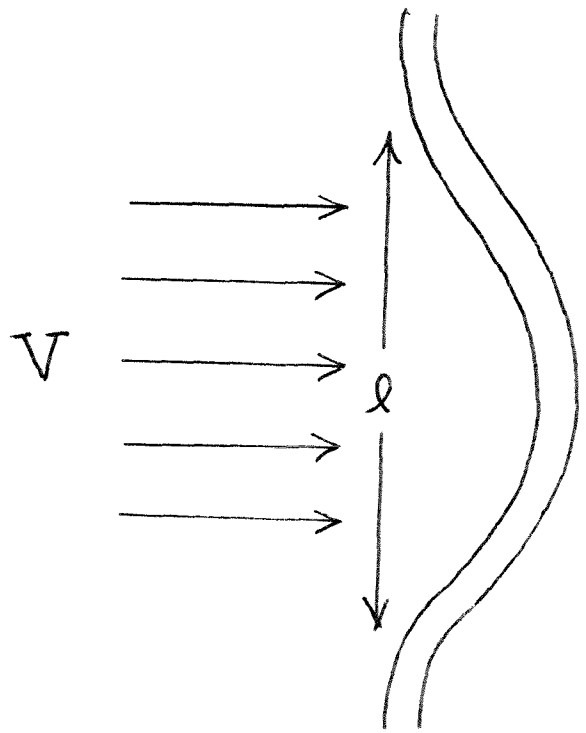


Figure 37: The bending of flux tubes.

(1953), the tension (per unit cross-sectional area) is $B^2/4\pi$; and since, at the moment when the bent portion of the tube is just about to be swept away, nearly the whole tension tries to resist the flow, we can write:

$$F_{\perp} \approx \left(\frac{B^2}{4\pi}\right)(\pi r^2) \tag{5.2}$$

Comparing equations (5.2) and (5.1) we conclude that the tube will be "bendable" -- that is, it will move with the fluid flow -- if:

$$4\rho v^2 l \geq B^2 r \tag{5.3}$$

(at least in order of maginitude).

We conclude that a flux tube of <u>any</u> size and strength can be bent by a flow at any speed, provided only that the flow acts over a sufficiently deep layer, l . On the other hand, for a flow of specified depth and strength, the tube will look increasingly "stiff" as either the field strength or the radius is increased.

In order to apply this result to the sun, we have to make assumptions about densities, velocities, field strengths, and dimensions. None of these is likely to be too accurate, but that may not be as important as it might seem, since the balance of equation (5.3) could easily go very strongly one way or the other; and in any event it is likely that we will be able to make a reasonable estimate as to the <u>relative</u>

effectiveness of granulation versus supergranulation.

For the depth, we will assume that the convective motions persist over a layer whose thickness is comparable to, or perhaps somewhat smaller than the radius of the cells: say, $\ell \approx 5{,}000$ - $10{,}000$ km for the supergranulation, and $\ell \approx 200$ - 400 km for the ordinary granulation. Using these values of ℓ, and consulting the tables in Allen (1973; §§76 & 77) one would guess that the appropriate densities to use in equation (5.3) are $\rho \approx 5\times10^{-5}$ gm/cm^3 for the supergranulation, and $\rho \approx 4\times10^{-7}$ gm/cm^3 for the ordinary granulation.

The typical horizontal velocities at the depths where these densities occur are obviously not well known, but one would assume that they have to be at least as great as the cell radius divided by the cell lifetime. That is:

$$v \simeq \frac{15{,}000 \text{ km}}{20 \text{ hr}} = 0.2 \text{ km/sec} \quad \text{for supergranules}$$

and

$$v \simeq \frac{500 \text{ km}}{8 \text{ min}} = 1.0 \text{ km/sec} \quad \text{for granules}$$

The true velocities could easily be higher by a factor of two or more.

Finally, as to the fine structure of the fields which are to be pushed about, the author is inclined (based primarily on instances in which a sunspot seems to be balanced by a plage of nearly equal area) to accept the view of Stenflo (1973), and others, that in most of the important network features, the flux has been concentrated into small

(second of arc) areas of high (thousand gauss) strength -- say $B \simeq 1000$ gauss and $r \simeq 500$ km (Stenflo, indeed, suggests $B \sim 2000$ gauss and $r \simeq 50 - 150$ km). These values apply at the surface. Below the surface, even higher field strengths and smaller dimensions might be found.

Putting these numbers ($B \simeq 1000$ gauss; $v \simeq 0.2$ and 1.0 km/sec) into equation (5.3) we obtain the condition that the fluid flows can freely bend (and, therefore, transport) flux tubes of radius r provided:

$r \lesssim \frac{1}{12} \ell \simeq 600$ km for the supergranulation

and $r \leq \frac{1}{60} \ell \simeq 5$ km for the ordinary granulation.

Although these numbers can (at best) be trusted only as to order of magnitude, two important conclusions emerge. First, the granules, in spite of their high velocities, are less effective than supergranules at pushing fields around (the supergranules win in part because they are assumed to operate upon the tube along a greater length, but more importantly because they are assumed to operate at a much higher average density). Second, because of the great difference in relative "bending power", it seems conceivable that a flux tube which appears pliable to a supergranule could appear stiff to an ordinary granule. Thus we would expect the supergranules to be the dominant force behind the movement of fields over substantial distances. There could

be a significant amount of interaction at the granule level, but it would be expected to be much more of a give-and-take situation, with the field affecting the velocities at least as much as the velocities affect the field. Indeed, the observation that the network field is (apparently) concentrated in the lanes between the granules (Dunn and Zirker, 1973) could arise more because the granule motions are forced to avoid the flux tubes, than because the field has been physically swept to the boundaries. If the field <u>does</u> move in response to the granulation, then, according to the present considerations, that would seem to imply either that the tubes are frayed to extremely small dimensions at the surface, or else that their shapes have been distorted in such a way as to present a much larger-than-expected surface area to the flows.

Substantially the same conclusion regarding the relative importance of granulation versus supergranulation was reached by Parker (1963) using very different words, and (what now seems) possibly out-dated information about the strength of the network field.

5.3 The Predicted Rate of Dispersal for Fields Embedded in a Convective Network

The prediction of the effective diffusion constant for particles injected into various kinds of one- two- and three-dimensional flows, is a branch of fluid mechanics which has

reached very high levels of sophistication (see, for example, Kraichnan, 1974). Such levels are hardly appropriate here, however, considering the crudeness of our present understanding of the properties of the solar flows.

All we really know is that the magnetic fields appear to be concentrated at the boundaries of the velocity supergranules (Simon and Leighton, 1964). It is thought that the time over which the cells come and go is on the order of 20 hours (Simon and Leighton, 1964; Janssens, 1970), but that cannot be regarded as being very certain, since (cf. § 4.6), the "lifetime" estimates are in fact only measurements of the time it takes for the boundaries of the cells to distort to where the magnetic features are hard to recognize as what they were before. It is conceivable that the velocity structure could persist for a much longer time, but it seems unlikely.

Virtually nothing is known about the dynamics of how the velocity cells develop or of how the field comes to be located at the boundary. It seems likely that there is a kind of equilibrium in which the field finds itself always squeezed between flows, in such a way that as new cells come and old ones go it can accomodate to the changing boundary without actually ever finding itself in the center of an established cell; but this has not been proved.

In addition it is not known if there is any tendency

either for the supergranules to repeat, or for them to come up preferentially in the spaces between the existing cells; nor if, when a new cell emerges, it must physically push aside the existing cells, or if it can superimpose itself upon them in a more subtle fashion. The answer to any one of these questions could dramatically alter the way in which one would have to attempt to estimate the rate of field dispersal.

In view of this uncertainty, we seem justified in adopting a far simpler approach to the basic question of whether the supergranulation is capable of producing the observed diffusion constant. Namely, we will ask: given a convective network with certain dimensions, and given that the fields are concentrated at the boundaries of the cells, how much motion is involved if the field has to adjust to a completely new pattern of cells in some characteristic time τ ? It would be very surprising if the numbers inferred in this way did not, at least roughly, agree with our previous results, but it will still be satisfying to see how the magnitude of the displacements is related to the size of the cells.

If we assume that the adjustment of the field to the new convective pattern requires displacements of size L , then according to equation (4.4) the resulting diffusion constant (regarding the process as a random walk) is:

$$D = \frac{\langle L^2 \rangle}{4\tau} \tag{5.4}$$

where $\langle L^2 \rangle$ is the mean-squared displacement.

In the simplest possible supergranular model we would assume that the field is concentrated uniformly around the borders of a pattern of closely-packed circular cells of radius $R_o \simeq 15{,}000$ km; and we can estimate how much the average field line has to move by imagining that the original field is suddenly placed in a new and unrelated convective pattern. Of the field fragments that find themselves within the midst of these new cells, a few will be at the center, and have to move by the full cell radius; others will already be near the edges, and not have to move very far to get to their new "stable" positions. On the average:

$$\langle L^2 \rangle = \int_0^{R_o} (R_o - r)^2 \frac{2\pi r\, dr}{\pi R_o^2} = \frac{1}{6} R_o^2 \tag{5.5}$$

where the integration variable, r, represents the distance of a fragment from the center of the cell.

Using $R_o = 15{,}000$ km and $\tau = 20$ hours, we deduce (cf. eqn. 5.4) $D \simeq 130$ km^2/sec. Essentially the same argument could be applied to the granulation (<u>if</u> we believed that it could bend the lines) which, with $R_o = 500$ km and $\tau = 8$ minutes would give $D \simeq 22$ km^2/sec.

These numbers are less than the diffusion constants calculated by Leighton (1964), both because of the factor of 6 in equation (5.5) and because of the factor of 4 (instead of 2) in equation (5.4). They could be made even smaller if we considered the fact that some fraction of the field lines will probably find themselves not in the midst of new convective cells, but rather in the "gaps" or "dead spaces" between. If we accept the estimates of Leighton, Noyes and Simon (1962) that there are ~5000 supergranules on the sun at any one time ($4.0\text{-}5.5 \times 10^3$ according to Simon, 1963), then there is evidently something like ~50% of the surface which is between the 30,000 km diameter cells. Since the particles which find themselves in that area do not have to move at all in order to adjust to the new pattern, our estimate of the diffusion constant could be lowered to as little as $D \approx 60\text{-}70$ km^2/sec. It seems unlikely, however, that such a large correction would be needed. In fact, we should probably be using in equation (5.5), not the apparent radius which the cells have at the surface, but rather the effective radius which they would have in the layers below the surface, where it is likely that they are packed densely together. That radius could easily be closer to 20,000 km than to 15,000 km (if the figure of 5000 cells for the whole sun is correct), raising the predicted diffusion constant to a little over 200 km^2/sec.

The predicted diffusion constant continues to rise if we include the effect described by Smithson (1972), and others, which is that the field appears to be concentrated at a few, isolated "stable points", rather than being spread either randomly or uniformly around the boundaries of the supergranules. (It is generally assumed that these stable points are at the major stagnation sites where several velocity cells come together). If this is actually the case (it is very difficult to establish observationally), then the distance which a field fragment would have to move in order to adjust to the new convective pattern is not simply the distance to the nearest border, but rather, the distance to the nearest stable point (which will tend to be farther away). We can estimate this distance by saying that on the average the locations of the new stable points will be distributed at random relative to the old ones, and that most of the time a field fragment will be able to move to its nearest neighbor in this new pattern. Thus, we can use the results of Appendix V to obtain:

$$\langle L^2 \rangle = \frac{1}{n_o \pi} \tag{5.6}$$

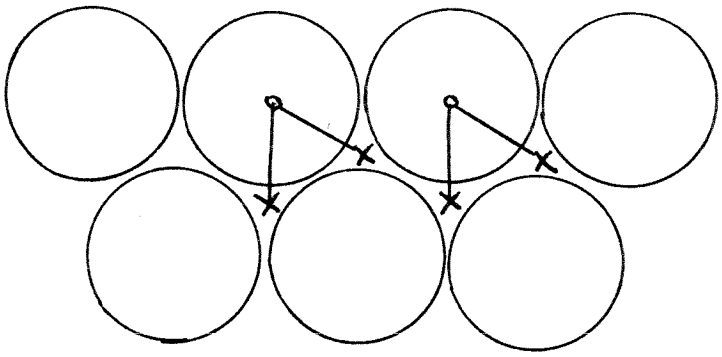
where n_o is the number of stable points per unit area.

The predicted value of n_o varies depending on exactly how the cells are packed together. For a network of close-packed circular cells, one would expect an average of about

two major stagnation points per cell; but for a loose-packed network, there might only be one (see Figure 38). Thus, if the total number of cells on the sun is ~5000, we would estimate $n_o \simeq (8\text{-}16)\times10^{-10}$ km^{-2}, giving $\langle L^2 \rangle \approx (2 - 4)\times10^{8}$ km^2. According to equation (5.4) this implies a diffusion constant of $D \approx 700$ km^2/sec - 1400 km^2/sec (using the standard $\tau = 20$ hours).

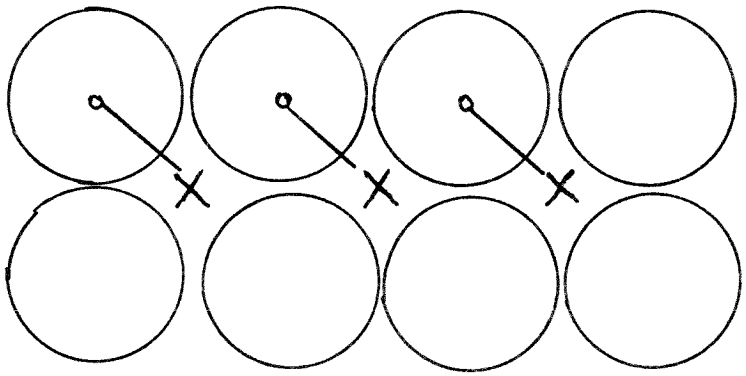
These numbers begin to strain the limits on the observed short-term (and long-term) diffusion constant of $D \approx 200$ km^2/sec - 400 km^2/sec. Thus if we wish to accept the "stable point method" of estimating the rate of field dispersal as being the more reasonable of the two alternatives, then we must seek some explanation for why the predicted D comes out so high. There are two obvious possibilities. One is that there could be more stable points than anticipated. To some extent this is undoubtedly true, since there are always a few places where nearly complete cells are outlined by field; but in general, counts of the number of important, distinct, isolated magnetic features in quiet regions tend to give numbers in the range $(1\text{-}3)\times10^{9}$ km^{-2}, which is not very different from what we have used above (the point densities of 5×10^{-9} - 12×10^{-9} km^2 reported in §4.5 included a considerable amount of "fine structure" within the clumps).

(a) Close-Packed Cells



2 stable points per cell

(b) Loose-Packed Cells



1 stable point per cell

Figure 38: Stable points in a convective network.

The other possibility is that we have underestimated the lifetime of the velocity cells, perhaps by as much as a factor of two or three. When one realizes that the standard 20 hour lifetime is based on the appearance of the magnetic network, this does not seem at all impossible. If the field patches are really concentrated at stable points, and if the stable points are determined by, say on the average, the intersection of three velocity cells, then (very roughly speaking) each point would be vulnerable to changes in any one of three cells; and so the pattern of field patches could easily change enough to become difficult to recognize in on the order of one-third the lifetime of the individual cells. (In this respect, it may be interesting to note that one frequently hears comments to the effect that in the enhanced network around active regions -- where the velocity pattern is presumably more clearly defined by the network -- it sometimes seems possible to follow gradual distortions in cells over periods of up to several days).

The true explanation of the observed diffusion constant is probably a combination of these two effects: that the field is not completely constrained to a few stable points, but only prefers to be there; and that the lifetime of the velocity cells is a little longer than generally assumed.

5.4 Conclusions

In summary, we have found that it is not at all difficult to account for the observed diffusion constant on the basis of motions in response to a convective network with the general size and lifetime of the supergranulation.

The observation that the field is concentrated at the boundaries of the velocity cells makes it virtually certain that the supergranulation plays some role in dispersing the field. The present considerations make it seem unlikely that any additional mechanism is necessary, at least over the periods of several months for which the observed diffusion constant of $D \approx 200$ km^2/sec seems to apply. The calculations are very sensitive, however, to conditions regarding which we have little information; and thus it is impossible, by this means, to definitely rule out the need for couplings to deeper layers, either to enhance or to restrain the long-range effects of the supergranular motions.

Finally, it does not appear likely that the ordinary granulation would be effective in transporting the field over any substantial distance.

6. Implications of the Random Walk Model of Active Region Decay

6.1 Introduction

In the preceding three sections we have been able to develop a reasonably clear and coherent picture of the decay of active regions, in which the dominant mechanism appears to be a 'random walk' of field lines. The broad outlines of the average pattern of development -- the growth in area, the increase in separation, the decline in flux, the lifetime -- can all be quantitatively reproduced by assuming only that the field lines move freely and independently in response to the supergranular motions. We have been able to show first that there is reason for believing that supergranular motions could displace field lines, second, that the predicted rate of dispersal agrees reasonably well with the motions actually observed over periods of 10-30 hours, and finally, that if those motions are extrapolated to longer periods ($\sim$ months), they would indeed generate the proper rate of diffusion to produce the observed long-term spreading and decay. At each stage in the argument, there is considerable uncertainty ($\sim$a factor of 2), and we cannot say with certainty that a refinement in the numbers would improve the agreement; but there does, at least, seem to be a certain "correctness" to this way of viewing active region decay.

In accepting the basic validity of the diffusion or random-walk model, we are, by implication, accepting certain consequences of the model with which we have not yet specifically dealt. It is the purpose of this section to enumerate a few of those implications, and to attempt to demonstrate that in those areas a comparison of the observations with the predictions of the model does indeed improve our understanding of what we have found, by isolating those aspects which are simply a reflection of the random motions, from those which are of other origins.

6.2 The "True" Lifetime of Network Features

6.2.1 The Lifetime of Magnetic Field Lines

In Section 4, we assumed explicitly that in studying the motion of magnetic flux in a decaying region, the field lines could be regarded as long-lived objects. Whether or not the random-walk model correctly predicts the rate of flux annihilation, it is clear that this is a reasonable assumption: for if a decaying plage can remain visible over several months, then surely the fraction of lines which it loses during a single day's observations cannot be very great.

Using the model, however, it is possible to make an even more definite prediction. For a doublet source (cf. eqn. 3.21), the surviving flux in each plage goes like:

$$N_+ \approx \frac{N_o a}{\sqrt{\pi D t}} \qquad (6.1)$$

where N_o is the initial strength. The fractional change, then, in any interval t, is simply:

$$\frac{\Delta N_+}{N_+} = -\frac{1}{2}\frac{\Delta t}{t} \qquad (6.2)$$

Thus, for a region which is, say, two rotations old ($t = 5 \times 10^6$ sec), the fractional loss in a typical 8 hour observing period would be less than 0.3%. While there might be some additional annihilation due to the interaction of the flux with pre-existing fields, and also from the extra mixing caused by differential rotation, it is clear that if the diffusion model is at all correct, the loss of flux in a decaying plage would be expected to be a very slow and subtle thing. Only for the smallest and most short-lived regions could the loss in a single day be important.

6.2.2 The Lifetime of Magnetic "Features"

In view of the preceding remarks, it may seem difficult to understand how there could be measurements indicating that magnetic (K-line) points last only 3 or 4 days (Smithson, 1973), or perhaps even only 10 hours (Macris, 1962). The dilemma is resolved when one realizes that these measurements refer not to field lines, but rather to the larger "network

features" which are formed by the association of many lines. These features can "die" either by the dispersal of the field lines, or, simply by their moving relative to one another until identification becomes difficult. In neither case is it necessary that the constituent field lines actually 'disappear' or be annihilated. The measured "lifetimes", then, are presumably an indication both of the characteristic time over which the associations 'disband', and of the time over which they move, as units. The distinction between these two aspects of the decay is seldom very clear, and the usual procedure is to pick some arbitrary "rejection criterion", r_o, such that a point will be declared as having "disappeared" if nothing can be found within that distance of the original point. The diffusion model does not tell us anything about how long the clumpings should hang together (that is a complicated matter involving exactly how the field adjusts to the continual changes in the locations of the stagnation points provided by the supergranulation); but it does provide us with a more objective means for isolating the points which escape detection due to motion from those which actually undergo changes in appearance.

Using the reasoning in Section 4, and the nearest-neighbor distribution for a random array (Appendix V), we would estimate that the probability, P, of finding a point, after time t, within a distance r_o of one of the orig-

inal points is:

$$P = \left(\int_0^{r_0} \frac{r}{2Dt} e^{-\frac{r^2}{4Dt}} dr\right) + \left(\int_{r_0}^{\infty} \frac{r}{2Dt} e^{-\frac{r^2}{4Dt}} dr\right)\left(\int_0^{r_0} 2\pi n_0 r e^{-n_0 \pi r^2} dr\right) \tag{6.3}$$

where n_0 is the density of points, and D is the diffusion constant. The first term represents the probability that the real motion is less than r_0. The second, indicates the probability of an accidental identification even if the true motion exceeds the "rejection criterion" (r_0). This expression can be rewritten in the form:

$$P = 1 - e^{-\left(n_0\pi + \frac{1}{4Dt}\right) r_0^2} \tag{6.4}$$

Since even infinitely-long-lived clumps would escape identification at this rate, a lifetime measurement should be regarded as being significant only if it shows that the fraction of points surviving after a time t is less than the fraction specified by equation (6.4).

Thus, for example, for Smithson, (1973) with $n_0 \simeq 1 \times 10^{-9}$ km^{-2}, and $r_0 \simeq 15{,}000$ km, we would expect $P = 0.98$ after 24 hours; that is, only $\sim 2\%$ would escape detection due to motion if $D = 200$ km^2/sec. Since Smithson in fact finds only 65-75% of the original points to be recognizable after 24 hours, one would be tempted to think that the bulk of the "missing" features had undergone changes in form. This interpretation is not totally obvious, however, since

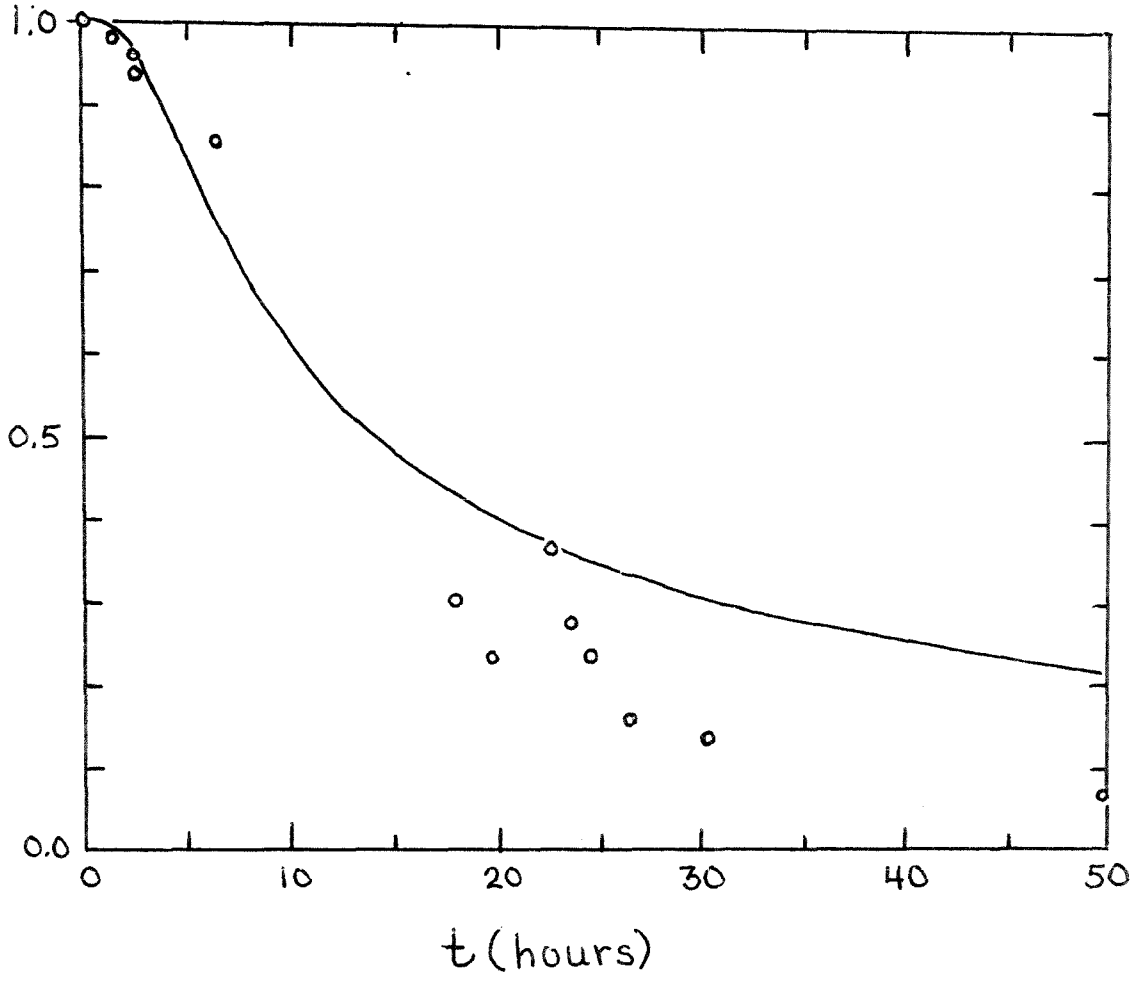


Figure 39: The lifetime data of Macris, and the predicted correlation for infinitely long-lived clumps.

he also reports that there is, by the same criterion, only about 66% reproducibility when different observers attempt to identify features on the same picture.

The results of Macris (1962) are even more difficult to interpret because we don't know exactly how he decided which points had survived. Figure 39 shows how, at $D = 200$ km^2/sec, the general form of his results could be produced by infinitely long-lived clumpings assuming a rejection criterion of $r_o \simeq 5{,}000$ km and a point density of $n_o \simeq 1 \times 10^{-9}$ km^{-2} (~ 1 point per supergranule). This is meant mainly to illustrate the shape of the predicted curve. It is not known if these are even close to the parameters used by Macris.

In summary, we must conclude that the question of how long the field line associations stick together has not yet been satisfactorily answered, because of the possibility of confusion between changes and notions. The diffusion model does, however, provide a quantitative framework, within which these two aspects of the problem can, at least in principle, be isolated.

6.3 Flare Production, and the Rate of Magnetic Energy Dissipation in a Decaying Region

In a similar vein, it is of interest to estimate the amount of magnetic energy which is, potentially, available from the random collisions (and subsequent 'annihilations') between opposite-polarity features in a decaying region.

It would not be surprising to find this available energy is related to the rates of surging, flaring, X-ray production, and other manifestations of active region 'activity'.

The exact rate of flux annihilation in a diffusion model with discrete sources injected at $t = 0$ can be found by differentiating equation (3.14):

$$\frac{dN_t}{dt} = -\frac{N_0}{\sqrt{\pi}\, t}\sqrt{\frac{a^2}{4Dt}}\, e^{-\frac{a^2}{4Dt}} \qquad (6.5)$$

The form of this function is shown in Figure 40. It reaches a peak rate of:

$$\left(\frac{dN_t}{dt}\right)_{max} = -\frac{0.325\, N_0 D}{a^2} \qquad (6.6)$$

at $t = a^2/6D$. For our standard region with $a = 30{,}000$ km, this condition would occur at an age of 4 - 9 days, for $D = 200\text{-}400$ km^2/sec. Intuitively, the maximum rate of flux annihilation occurs when the region is in its maximum state of development. In the model it happens at this time because it is when the polarities from the two point sources first start to strongly interdiffuse. The curve in Figure 40 is very reminiscent of the graph in Waldmeier (1955; his Figure 46) showing the rate of flare production in a 'typical' region rising to a maximum at about the time of maximum sunspot development, then falling.

In order to convert the rate of flux annihilation into an energy we have to consider the dimensions involved, and

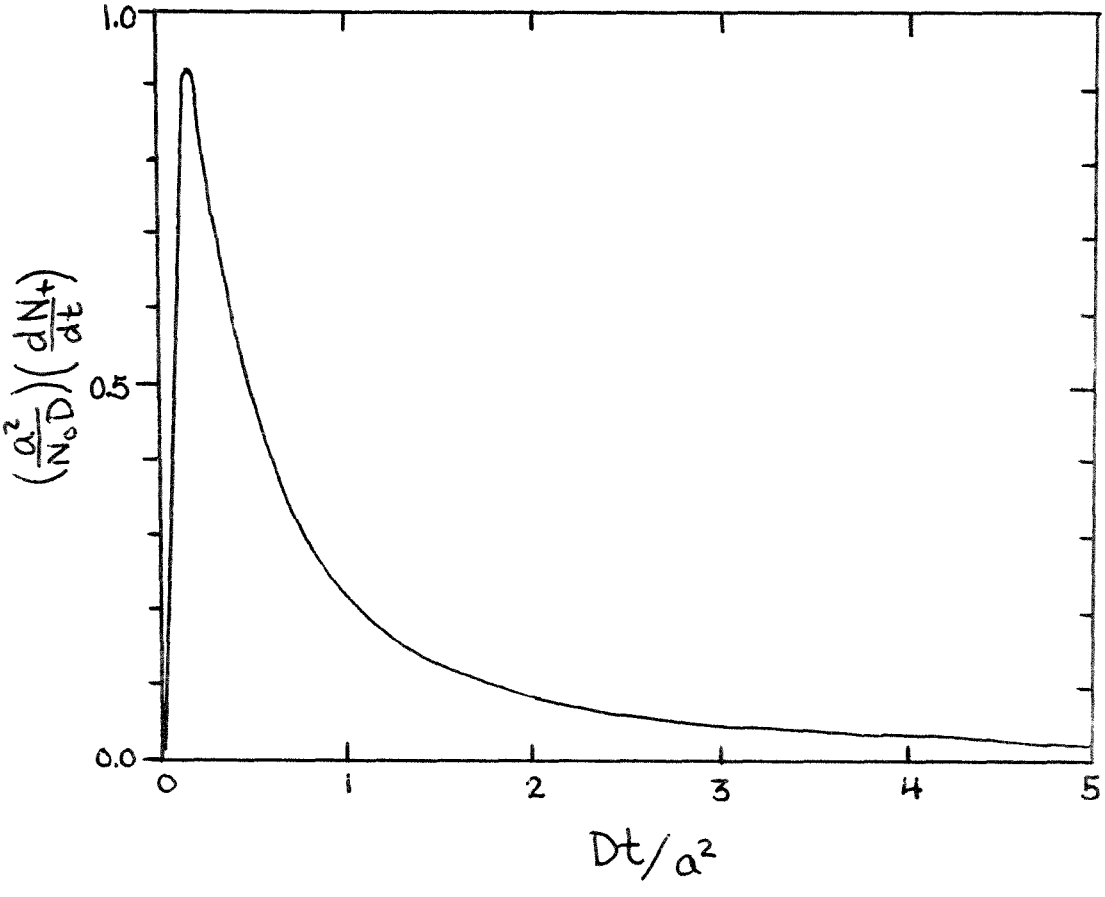


Figure 40: The predicted rate of flux annihilation for discrete bipolar sources.

at this point our arguments become very rough. If the flux is at a strength B, and spread uniformly over an area A, then the energy in connecting loops would be on the order of $(B^2/8\pi)AL$, where L is the characteristic "size" of the region (see Figure 41). Rewriting this in terms of the flux, and setting $A \approx L^2$, we obtain the following expression for the magnetic energy:

$$N_+ \equiv BA \Rightarrow \quad \mathcal{U} = \frac{N_+^2}{8\pi L} \tag{6.7}$$

(the energy could be considerably higher if the coronal field is assumed to be non-uniform).

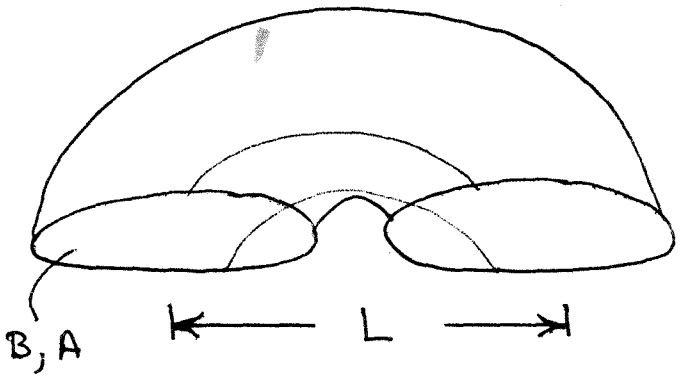
For reasonably short times we might be justified in assuming that L is reasonably constant and $\approx 2a$. Then:

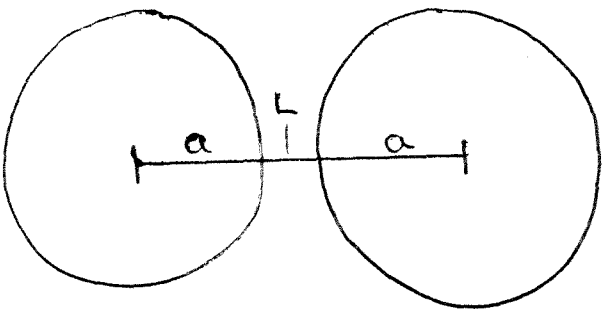
$$\frac{d\mathcal{U}}{dt} \approx \frac{N_+}{4\pi L}\frac{dN_+}{dt} \approx \frac{N_+}{8\pi a}\frac{dN_+}{dt} \tag{6.8}$$

At the point of maximum development, using equation (6.6), and taking $N_+ \approx N_o$ we obtain:

$$\left(\frac{d\mathcal{U}}{dt}\right)_{max} \approx \frac{N_o^2 D}{8\pi a^2} \tag{6.9}$$

For our standard region with $N_o = 10^{22}$ Mx and $a = 3\times10^4$ km, this gives a dissipation rate of $\left(\frac{d\mathcal{U}}{dt}\right)_{max} \approx (4 - 8)\times10^{26}$ erg/sec. If all this energy were to go into producing flares, it would be sufficient to produce a class 3 flare every ~ 4 days -- accepting Bruzek's (1967) estimate of 2×10^{32} ergs for U_{total} -- or perhaps $\sim$10 class 2 flares a day -- taking Tandberg-





$$L \approx 2a$$

$$A = \pi a^2 \approx L^2$$

Figure 41: Calculation of magnetic energy dissipation

Hansen's (1967) estimate of 10^{30} - 10^{31} ergs.

It should be noted, perhaps, that flare production represents only a very small part of the overall energy balance of an active region. If, for example, the associated faculae were only 1% brighter than the surrounding photosphere over an area as small as 500 millionths of a hemisphere, then the steady excess energy radiated by them would be nearly a hundred times greater than the rates which we are considering here. Thus the decision as to whether some small fraction of the total energy should be released in one way or another, or at exactly what moment that energy should be released, may well be determined by relatively subtle conditions -- such as the exact configuration of field along the neutral line -- beyond the scope of the simple diffusion model. What the diffusion model does indicate is that there is likely to be a more than ample pool of magnetic energy available simply as a result of the normal decay of flux; so that flares could be produced should the conditions be favorable for tapping it.

When the observed rate of flare production (on the basis of the reports in the monthly NOAA bulletins) is evaluated for some of the simpler regions listed in Appendix II, it is found that equation (6.9) offers at best only a very rough guide as to the maximum level of activity. For example, McMath 13790 and 13278 have approximately the same value of

N_o^2/a^3 , and similar, simple bipolar configurations, yet the former one has 22 class 1 flares and 143 subflares during its first disk passage; whereas the latter has only ~45 subflares.

As far as the amount of activity for a particular region on successive rotations, we can give one example. The The number of flare reports for Region 13790 on its first three disk passages (lumping the class ones and the subflares together) goes like: 165 : 26 : 0 . This can be compared to the 'loss of magnetic energy', in the respective two week periods, as computed using equation (6.7) with the observed fluxes, and $L \approx 150{,}000$ km. These are (in ergs): 2.3×10^{33} : 2.8×10^{32} : 1.22×10^{32} . Thus there is some difficulty in understanding why there should be flares on the second rotation but not on the third; and the confusion would be even worse if we had used the predicted rate of flux decline (see Figure 22). In any case, there seems to be considerably more energy available per event than is necessary, so perhaps the important question is not where the energy comes from, but why it is sometimes released as flares and sometimes not.

6.4 Differential Rotation

There are two long-standing and well-known observational problems with respect to the sun's differential rotation. One is that the sunspots seem at times to rotate faster than

the photospheric plasma by as much as almost a degree per day (see, for example, Howard and Harvey, 1970). The other is that many of the manifestations of weak magnetic fields, such as filaments (see, for example, Glackin, 1974) -- and indeed direct cross-correlations of the magnetic signals themselves (Wilcox and Howard, 1970) -- seem to indicate that the weak fields rotate more rigidly than their sources, the spots.

As to the problem of the discrepancy between the spot and photospheric rates, we can have little to say. If the diffusion model is right, and the field lines move freely in response to the supergranulation, it is difficult to see how they could ignore a general streaming of the photosphere past them. In other words, we would expect the decaying fields to become very strung out in longitude, something as if we had dipped a pen in a moving stream of water and were watching the ink flow out. While it is possible that this effect could indeed happen to some extent on the sun, it is a little hard to believe that it could happen to the extent of one degree per day. Perhaps there is some unsuspected subtlety in the interpretation of the Doppler observations.

The diffusion model is more helpful in explaining why the weak fields should rotate more rigidly than their sources: the answer to that must to some extent be simply that the source fields are, on the average, aligned with the plus and

minus polarities separated in the east-west direction. Each source, as it decays and spreads is trying to build a north-south neutral line, and hence will influence the measured rotation rate at both higher and lower latitudes. The net effect is to broaden the differential rotation curve, and that seems to be pretty much the effect that is observed.

The exact calculation of how much diffusion broadens the differential rotation curve is rather complicated, and we will not attempt it here. It depends to some extent on the latitude distribution of the source fields. For example, the weak fields at 30° latitude would be strongly dominated by the remnants of lower-latitude injections, and hence tend to reflect the relatively high rotation rate of their sources. Weak fields at, say, 10° latitude, on the other hand, would be influenced by both higher-and lower-latitude sources in more or less equal numbers; and hence might, on the average, rotate at about the same rate as the spots.

Without explicitly making the calculation it is very difficult to estimate what fraction of the filament-spot "error" is a simple consequence of diffusion and what fraction is necessarily of a different origin. We note only that the amount of neutral line shearing found by Leighton (1964), in calculations of diffusion in the presence of differential rotation, is much closer to the filament rotation law of d'Azambuja and d'Azambuja than to the spot rotation law of

Newton and Nunn, even though the photospheric plasma was assumed to rotate in accordance with the latter law. We attribute this to the tendency of the east-west sources to fight to establish 'rigidly rotating' north-south neutral lines.

6.5 The Rate of Magnetic Recombination; and the Mixed-Polarity Background

We have already noted (§ 3.3.1) that the absence of any obvious fine-scale mixing of polarities along the neutral line separating the two halves of a decaying active region seems to imply that in the areas where opposite polarity lines diffuse together they are able to recombine and disappear in a very rapid and efficient manner. It is somewhat surprising, then, to find that when one examines full-disk magnetograms with both high spatial resolution and high sensitivity (such as those from Kitt Peak), there are indeed areas of weak mixed polarity. These are most prominent in the gaps between the extended 'unipolar' patches that are the obvious remnants of substantial active regions, but that could be simply because there is no dominant field there to obscure the underlying background. While it is possible that the process of recombination is less efficient between regions than it is within regions, so that greater mixing could occur, the evidence does not seem to point in that direction, and it is not felt that these areas of mixed polarity result

from the random association of lines which once emerged in major active regions. Rather, it seems more likely that they reflect a general background maintained by a continuous series of 'micro-eruptions' (ephemeral regions and smaller) which appear and dissipate more or less in place. One will note that the diffusion model predicts a very short lifetime for the residue of an eruption with a small initial separation (cf. §3.2.3c).

6.6 The Development of Large Scale Fields

It is of considerable interest to try to follow the progress of the diffusion model over periods longer than the $\lesssim 6$ months for which individual regions can be observed; and there are two approaches which suggest themselves.

One is to attempt to model the observed fields for the whole sun be inserting sources at each place where fields are seen to erupt, and then computing the total superimposed field that would result from all of these, assuming each field line to move independently of the others, and then comparing the predicted background pattern to that which is actually observed, hopefully improving the agreement by adjusting the diffusion constant, and perhaps the differential rotation law.

One might worry that all the tiny eruptions -- many of which would be missed because they happened on the back side -- would have to be included to get a reliable result, and hence

the very magnitude of the task would make it quite impossible. This is not true, however, for even if, as Harvey and Martin (1973) suggest, the total flux injected onto the surface by the small, spotless eruptions is as great or even greater than that injected by the major active regions, their net effect, at least as far as the large scale fields are concerned, would be much less. This is for two reasons. First because of their small initial separations, the injected flux does not linger very long on the surface: most of it will be gone after a day or two, whereas a significant amount of the flux injected by a major active center can remain for many weeks (the characteristic time goes like a^2/D where a is half the initial separation). Second, they emerge with nearly random orientations (Harvey, Harvey, and Martin, 1975), and will probably tend to cancel out each other's effects more thoroughly than would the larger regions, with their systematic $+6^{\circ}$ tilt. Thus it is only the larger spot-strength regions which will contribute significantly to the formation of large-scale fields, and <u>their</u> source strengths (flux times separation), can, on the average, be pretty well specified just by giving the size of the spots. As a result, the program of physically modelling the observed pattern of eruptions is not as unfeasible as it might at first seem, but it would still certainly be messy.

A far more elegant approach, if one is only interested in the long-term diffusion constant, is that suggested by Leighton (1964): which is to look only at the 'component' of field averaged over longitude, thereby reducing the problem to the one-dimensional one of how the sources spread in latitude. This spatially-averaged field, as it turns out, is expected to evolve in a way which is completely independent of any assumptions regarding the differential rotation (or, more generally, regarding systematic currents along lines of constant latitude) (Leighton, 1964; eqn. 15):

$$\frac{\partial \langle n \rangle}{\partial t} = \frac{D}{R_\odot^2 \cos\lambda} \frac{\partial}{\partial \lambda} \left(\cos\lambda \frac{\partial \langle n \rangle}{\partial \lambda} \right) \qquad (6.10)$$

where $R_\odot$ is the radius of the sun and λ is the latitude. In general, this equation can be solved in terms of spherical harmonics (see Leighton, 1964); but, we can adopt a simpler approach, since, considering the small diffusion constants which we have found in Sections 3-5, we will be interested mainly in times over which the spreading is small compared to the radius; and furthermore, in sources which are placed close to the equator ($\lambda = 0$).

Under those circumstances, and knowing that we don't have to worry about differential rotation, we can approximate the solutions of equation (6.10), by considering the now-familiar doublet source (eqn. 3.20), in flat, non-shearing coordinates; evaluating the average field by adding up

the flux which it contributes to various strips of "latitude", and dividing by the area of the strips -- assuming that the source is inclined at some angle α relative to the equator. The result is:

$$\langle B \rangle = -\frac{2N_o a \sin\alpha}{R_\odot (4\pi Dt)^{3/2}} y\, e^{-y^2/4Dt} \qquad (6.11)$$

where y is the linear distance (km) from the point of origin in the north-south direction. If $\alpha = 0$ -- that is, if the source is aligned parallel to the equator -- the average field is identically zero at every latitude (since the region contributes equal amounts of preceding and following polarity to each strip). On the other hand, if it has a positive inclination (p-end closer to equator) then there will be an excess of following (negative) polarity at $y > 0$ (higher latitudes), and an excess of preceding (positive) polarity at $y < 0$ (lower latitudes). In other words, a tilted region contributes a tiny doublet moment in the north-south direction, and as the region decays, that moment fuzzes out in latitude. This behavior, for the standard region with $N_o = 10^{22}$ Mx, $a = 30000$ km, and $\alpha = +6^\circ$, is shown in Figure 42, assuming $D = 200$ km^2/sec.

Because of the large dimensions involved, the spreading in latitude is quite gradual. The peaks of equation (6.11) occur at:

$$y_{max} = \pm\sqrt{2Dt} \qquad (6.12)$$

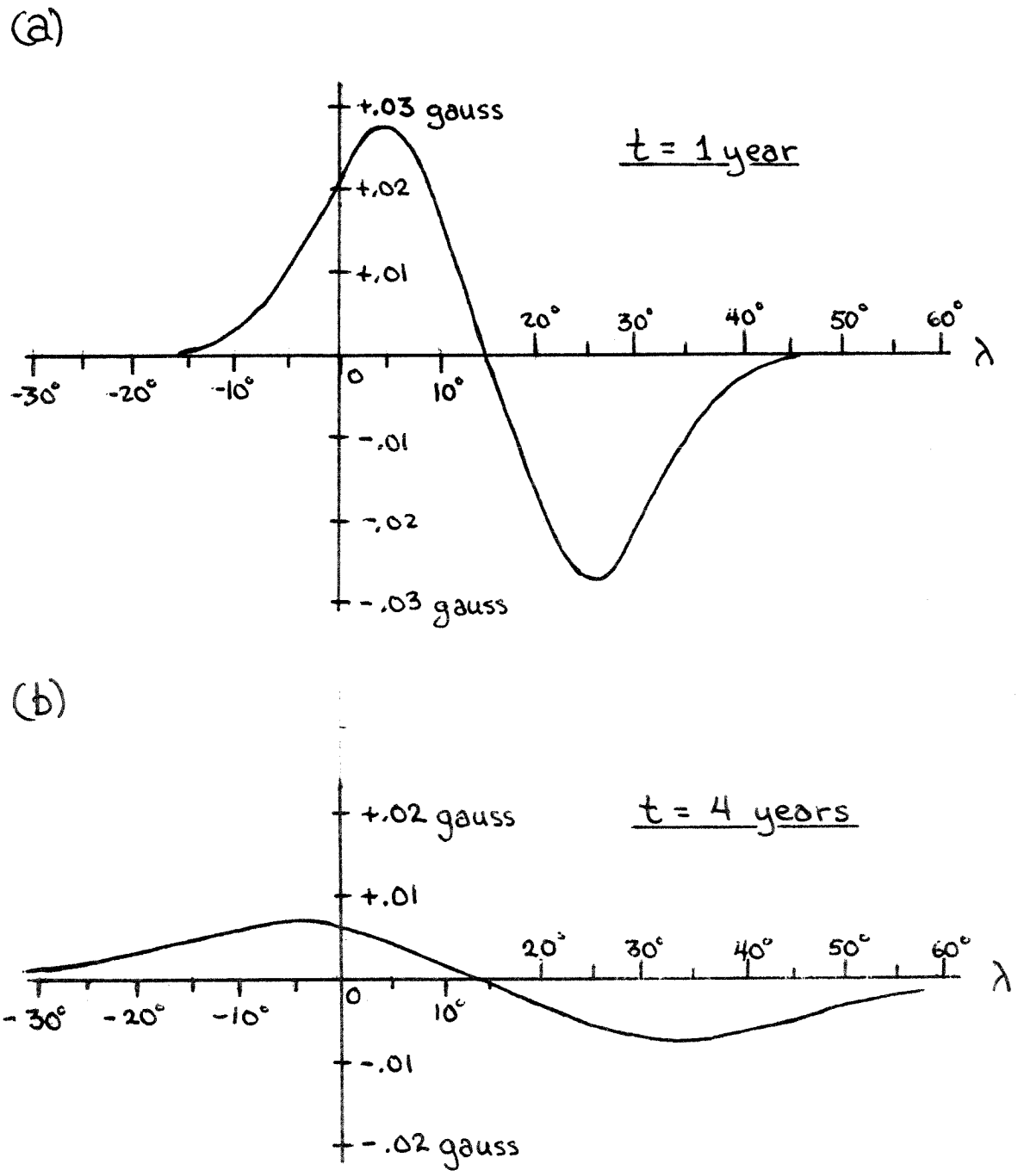


Figure 42: Contributions to the longitudinally-averaged field for a single region injected at $\lambda = 15^{\circ}$

and the peak field is given by:

$$\langle B \rangle_{max} = \frac{N_0 a \sin\alpha}{(2\pi)^{3/2} e^{1/2} R_\odot D t} \quad . \tag{6.13}$$

Thus, for example, if $D = 200$ km^2/sec, then after one year ($t \simeq 3.16 \times 10^7$ sec), the primary influence of the region has spread by only $\sim 100{,}000$ km or $\sim \pm 10^\circ$ in latitude.

As a result, we expect the accumulated effects of the many individual regions that erupt over the cycle to produce a "striping" or "banding" when the observed field is averaged over latitude. That is, the poleward side of the active zone should be fringed by following polarity, and the equatorward side by preceding polarity. It is these accumulated remnants, by the way, which are supposed to fuzz out to form the polar fields.

Stenflo (1972) has actually added up the average field seen in various latitude zones on Mount Wilson magnetograms from 1960-1969. If we accept his numbers (and they seem to be confirmed by Howard, 1974a), then there is some slight evidence for the predicted active zone "striping", but it is only at most the ~ 1 gauss level (even after correction for line weakening) in the yearly averages; and the polar field, in his 'butterfly diagram', seems to form more by a concentration of weak remnants than by the diffusion of strong ones (see Figure 43).

Figure 43: "Butterfly diagrams" of longitudinally-averaged solar magnetic fields.

(a) The results of Stenflo (1972) based on Mount Wilson daily data. The contours correspond to ½, 1, 2, 4, 8, and 16 gauss (after correction for line weakening). The strengths near the poles have been corrected for foreshortening (assuming a radial field) and are highly uncertain. The low-latitude peaks in 1966-1971 coincide almost exactly with the observed distribution, and are not offset towards lower or higher latitudes, as might have been expected. They exhibit the same polarity as the preceding spots in each hemisphere.

Stenflo's data end in 1970. The contours beyond that date have been added by rough visual estimates based on the daily contour plots, and should not be trusted as to strength.

(b) Predicted large scale field for the same period based on Leighton's (1964) calculations with $T_o = 20$ years ($D = 767$). The contours represent linear increments in field strength, but there is no overall calibration. Note that the low-latitude peaks are offset from the zone of miximum spot activity.

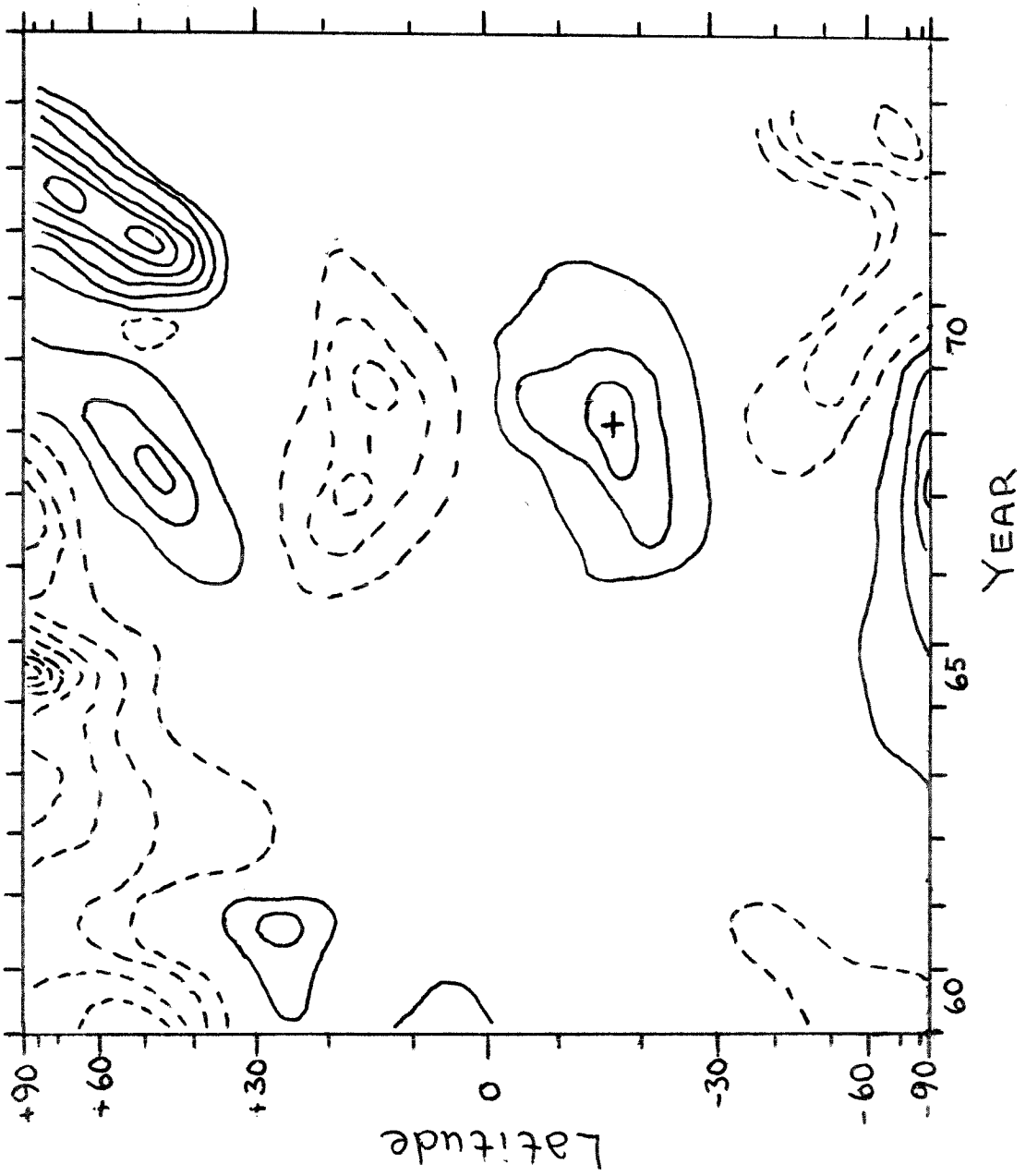


Figure 43a

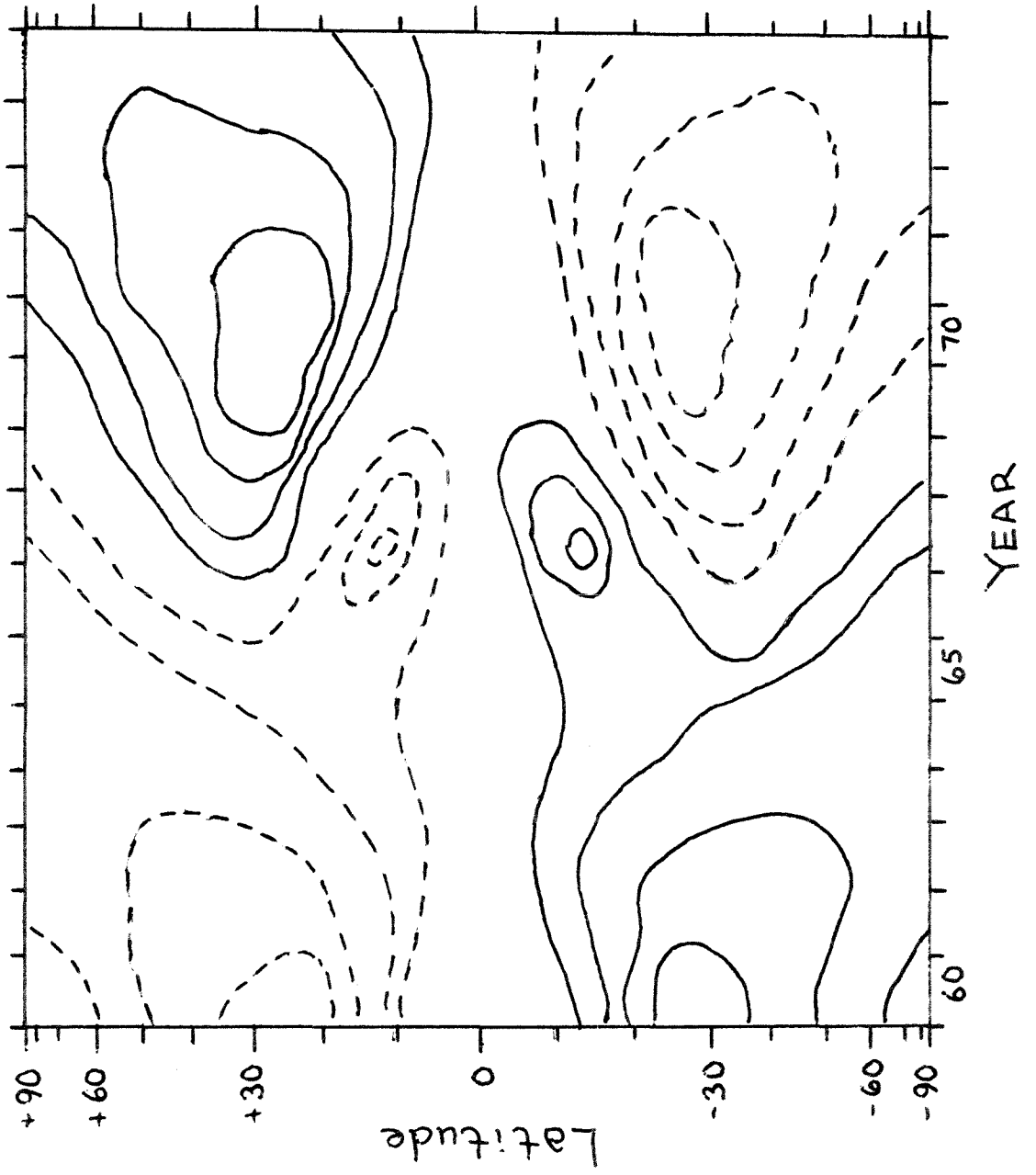


Figure 43b

We can estimate the predicted strength of the remnant bands, for any given diffusion constant (they are expected to be most prominent when the diffusion constant is lowest), by adding up the contributions from a reasonable number of sources (much as was done numerically by Leighton, 1964, though he assumed a much higher diffusion constant than now seems appropriate).

In order to determine what number of sources is 'reasonable' we counted the number of sunspot groups of various sizes actually reported in different latitude zones for a six month period (May-October) in 1972. Since that was a year of average activity, one might expect the results to be typical of the yearly eruption rate for a single hemisphere. The results are given in Table 6.1, where each group is listed according to its maximum total (i.e., umbra plus penumbra) sunspot area in millionths of a hemisphere. No effort was made to correct either for 'returns' or for the cases in which a group was probably not seen at its point of maximum development.

Table 6.1 Typical Number of Sunspot Groups in One Hemisphere over a 12-Month Period

Latitude	Peak Sunspot Area (millionths of $2\pi R_\odot^2$)							
	10	10-50	50-100	100-200	200-500	500-1000	1000-1500	1500-2000
0-5°	0	2	1	0	2	0	0	0
6-10°	12	8	4	8	7	2	0	0
11-15°	10	6	7	7	5	2	1	1
16-20°	5	5	2	4	3	0	0	0
21-25°	3	1	0	1	0	0	0	0
Total	30	22	14	20	17	4	1	1

While the smaller groups are more numerous, it will be noted that the average area is ~160 millionths, and that half of the (1.7×10^4 millionths) total area is contributed by regions with peak areas greater than ~400 millionths.

Now since we know that the source strength is roughly proportional to the spot area, and since we know that our nominal region with $N_o = 10^{22}$ Mx and $a = 30{,}000$ km has an area of about 250 millionths; the level of activity indicated by the table is equivalent to something like 70 of the nominal groups per year. This could be raised, perhaps, by a factor of 2 if we wished to correct for the flux which was missed when groups reached their peak on the back, and perhaps by another factor of 2 if we wished to consider a year of higher activity; so we would conclude that around the

peak of the cycle, the number of eruptions in each hemisphere is probably equivalent to about 250-300 of the nominal regions per year.

Since the spreading is so slow, the residual field at any time can be thought of as representing the sum of contributions from all the regions erupting over the previous 5-10 years, but it will be dominated by the most recent ones. We could estimate a typical residual field by adding the contributions over, say, ~ 4 years around the peak of the cycle. From Table 6.1 we deduced that there would be ~1000 (nominal) regions in that period; and according to Figure 42, the contribution from each would be $\sim .007 - .028$ gauss (or even more depending on how close the eruption was to the moment of observation). Thus we conclude that if $D = 200$ km^2/sec, the residual background "banding" that results from the systematic tilt of the individual regions, should build up to $\sim 10 - 30$ gauss in the declining years of the cycle. This is clearly very much greater than what is observed.

There are several possibilities for why the anticipated long-term residual fields might fail to materialize in the observations. One is that the magnetic regions may not be tilted in the way that has long been inferred from sunspot drawings; but we have checked that ($\S$ 2.4) and the inclinations seem to be as advertised. Another is that we may have grossly underestimated the diffusion constant -- but that too

seems unlikely, since we have made many checks; considering not only the areas of the decaying plages (which are admittedly hard to measure) but also the separation, and the width of the filament channel, and the short-term motions -- and in any event, the polar field doesn't look (observationally) as if it is formed simply by the fuzzing out of an active region residue at any diffusion constant. Finally, there is the possibility suggested in § 3.2.3d, that even though each field line may seem for all practical purposes to be moving at random, and even though the area develops in that way, it is still possible that the two ends may never entirely lose track of each other, so that there could be a very slight systematic "attraction", which in the end dominates the annihilation of flux, and causes the two halves of the region to destroy each other much more thoroughly than would have been expected at random.

It will perhaps be recalled that this is not the first indication we have had that there could be something wrong with the rate of flux annihilation. The direct measurements from the Mount Wilson magnetograms (§ 3.2.3d) seemed also to show that the flux was disappearing too rapidly -- although we could write that off as perhaps being due to a lack of knowledge about the amount of flux present below the '5-gauss' level. And the lifetime measurements (§ 3.2.3c), too, seemed to come out shorter than expected; but at the time

we were mainly interested in the slope of the line (Figure 21) and managed to patch up the normalization by using a threshold field somewhat higher than may have seemed reasonable. Even so, "large" diffusion constants seemed to be needed to get the proper lifetime ($B_{thresh} \simeq 6$ gauss to get $D \approx 375$ km^2/sec). When these difficulties are combined with the absence of the expected long-term residual fields, the weight of circumstantial evidence seems clearly to be leaning in the direction of accepting the notion that the long-term annihilation of field is dominated by non-random forces.

In the final section (§ 7) we will endeavor to show that even the slight residual field which <u>is</u> observed could be sufficient to account for the observed variations in the polar fields, provided the diffusional motions are assisted by a systematic poleward meridional flow.

7. Polar Fields, Long-term Mixing, and Meridional Flows

7.1 The Poleward Migration of Filaments

The poleward migration of filaments that is observed during the ascending part of each cycle has long been associated with the encroachment of new-cycle polarities on the existing polar fields (see, for example, Waldmeier, 1960; and Howard, 1974a). Leighton (1964) suggested that the phenomenon could be quantitatively understood on the basis of diffusion alone. By subjecting the observed pattern of sunspot eruptions to a diffusion-like spreading, he was able to show that the zone of zero longitudinally-averaged field would be expected to move towards the pole at approximately the observed rate, provided one was willing to assume that $D \approx 1000$ km^2/sec (To = 15 years; cf. eqn. 3.28).

We have two reasons, now, for believing that this is not an entirely complete explanation. First, the observed diffusion constant appears to be much lower -- closer to 200 km^2/sec . And second, there does not seem to be a sufficient pool of residual following polarity along the poleward sides of the active zones to account for the formation of the polar fields purely by diffusion. Observationally, the average 'mid-latitude' field appears to be less than a gauss (even after correction of line weakening), whereas the polar fields are perhaps several gauss (they are too close to the limb to measure accurately). In any purely diffusional model the

numbers would have to be turned the other way around -- that is, the mid-latitude 'source' has to be stronger than the tail which reaches to the pole.

Figure 44 presents observations of the poleward migration of filaments as derived from the data of Bocchino (1933) and Waldmeier (1946-1975). Each author gives yearly averages of the area of prominences in 5° wide zones of latitude. When one examines the numbers, it is customary (during the ascending part of the cycle) to find, in each hemisphere, a major peak somewhere near the equator, and a secondary peak at higher latitudes. The curves in Figure 44, represent the position of this higher latitude peak as a function of the number of years since sunspot minimum. In a few cases the equatorward peak has also been indicated.

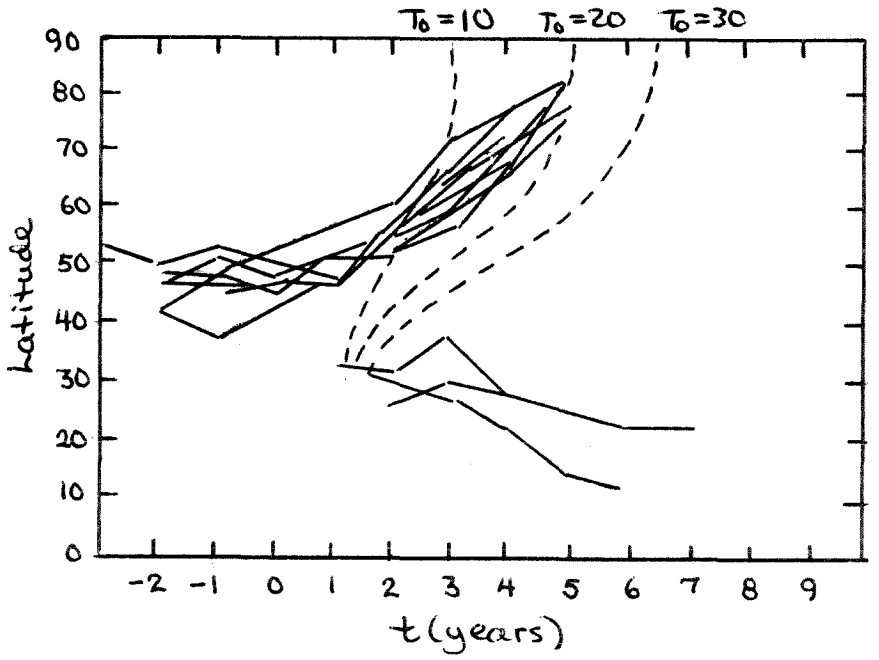
There seems to be a slight systematic difference in the rates of migration for the years reported by the two authors, but in both cases the basic poleward motion is clearly defined. The migration, which appears to begin at $\sim 40\text{-}50^{\circ}$ latitude very near the time of sunspot minimum, advances steadily in latitude 'reaching' the pole (by extrapolation) in $5\frac{1}{2}$ - 6 years (the actual migration dissipates at $\sim 75\text{-}80^{\circ}$). Bocchino's data suggest a speed of $\sim 6.8^{\circ}$/year, with a slight hint of acceleration. Waldmeier shows an almost perfectly linear motion at $\sim 9.2^{\circ}$/year. The average is 8°/year, or 3 meters/sec.

Figure 44: The filament migration observations of Bocchino (1880-1931) and Waldmeier (1945-1973).

Data from the northern and southern hemispheres have been combined. The horizontal scale is calibrated in years since sunspot minimum (for the whole sun).

In the upper figure, the dashed lines indicate the poleward motion of the zone of zero longitudinally-averaged field as predicted by Leighton (1964) on the basis of the observed average pattern of sunspot eruptions at T_o = 10, 20, and 30 years. (D = 1534, 767, 511 km^2/sec, respectively).

(a) Bocchino 1880-1931



(b) Waldmeier 1945-1973

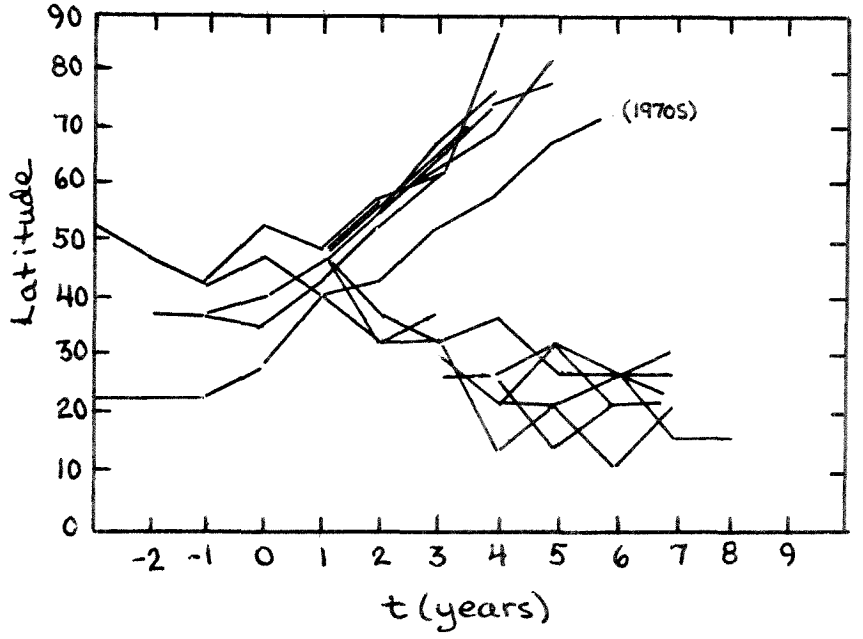


Figure 44

While there may, as Howard (1974a) points out, be some difficulty in identifying the prominence zone exactly with the latitude of zero longitudinally-averaged field, and that could conceivably account for part of the phase difference, it seems clear that the observed migration is both much earlier and more rapid than would be expected on the basis of a pure diffusion model with $D = 200$ km^2/sec ($T_o = 80$ years). The next most obvious explanation, then, would be that the fragments of insurgent field are simply being carried toward the pole by a meridional flow, operating at least between $\sim 40 - 70^\circ$ and moving at ~ 3 m/sec (cf. Bocchino, 1933).

7.2 White Light Faculae

Further evidence for the probable existence of such a systematic flow was uncovered during the course of a study of the latitudinal distribution of white light faculae, as represented on the daily plates in the Mount Wilson archives. Sheeley (1964) had already shown that the number of isolated faculae shown on these plates, at the pole, varies in a simple manner (and again in a manner which could easily be understood on the basis of diffusion at $T_o = 15$-20 years, but not at $T_o = 80$ years) but had not indicated very clearly what happened in the mid-latitudes.

In the present study, tracings were produced by marking the positions of the most prominent facular granules on a transparent overlay. Since the white light faculae are only

visible within about 20° of the limb, it is easy to entirely miss an equatorial region if it does not happen to be favorably located on the chosen day. For that reason, the tracing shown in Figure 45 is actually a composite of tracings for 10 separate days during the rotation of August, 1972; which gives a reasonably accurate visual impression of the latitude distribution of the faculae for that year. One will note that the polar fields (particularly in the north, which happens to be tipped towards the observer by 7° in August) are clearly detached from the fields of the lower latitude active regions, with very little flux between ~40° and 65°. This distribution was found to be characteristic of the declining years of the cycle, and of the sunspot minimum.

The interpretation of such an observation is not very obvious in terms of any purely diffusional model of active region decay. One could understand how, perhaps, during the ascending years of the cycle, when the following polarity being injected by the new cycle spots is opposite to the polarity of the established polar field, that there could be some annihilation and that might lead to a dip in the number of faculae; but this is the descending part of the cycle, and the polar flux is of the same sign as the field being injected by the spots, and there is no obvious reason to expect a minimum in the distribution. If anything, diffusion should be working to smooth out the existing distribution,

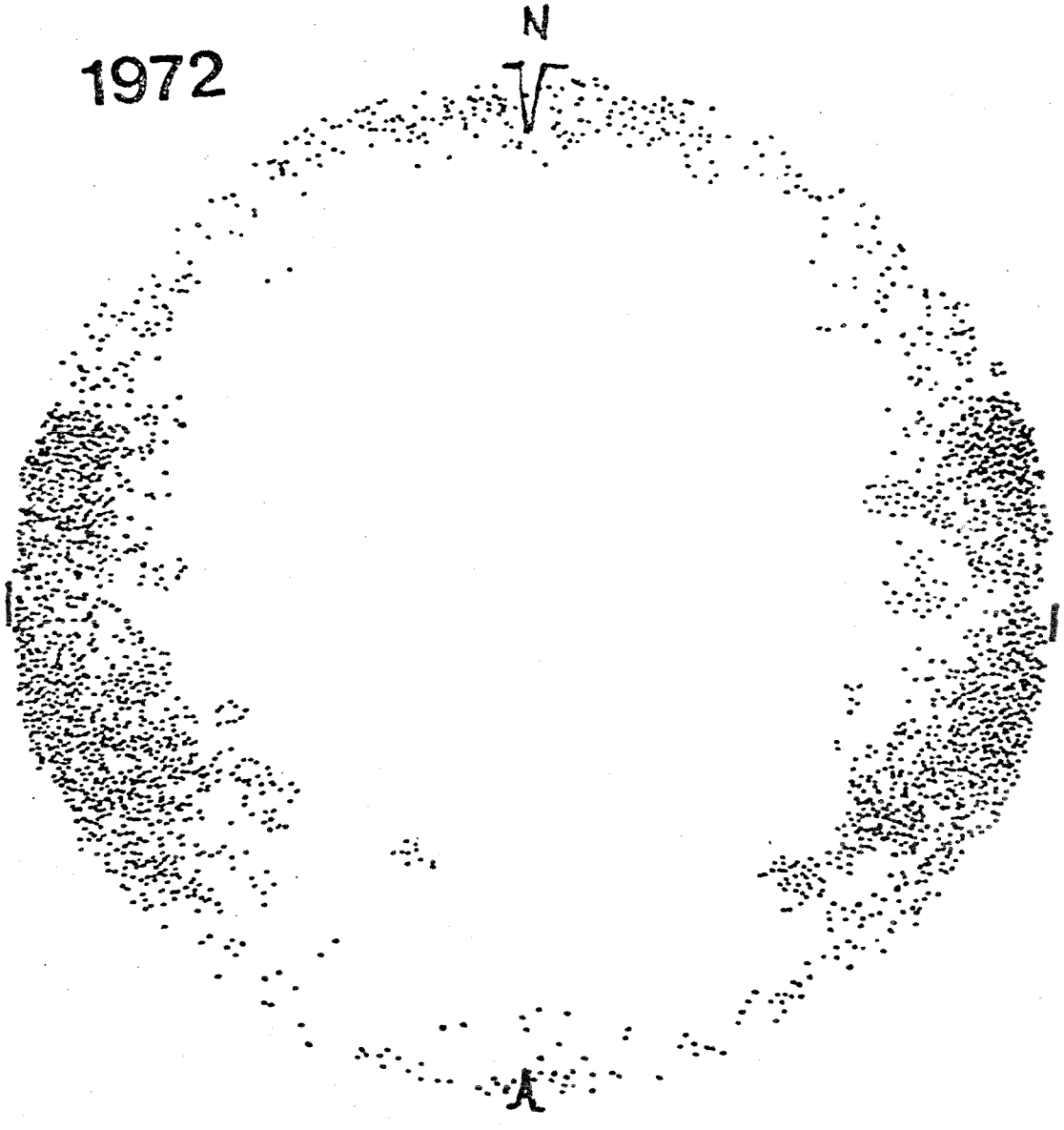


Figure 45: Pattern of white light faculae for August, 1972

and even to disperse the polar field that had already accumulated. In fact, the cap-like concentration of field seems to be becoming continually stronger and more prominent.

The minimum in flux density could, however, be easily understood if the field fragments were being physically swept across the mid-latitudes, from a source near the equator to an accumulation near the pole.

In all, we have three pieces of evidence suggesting the existence of the current:

(1) The appearance in the butterfly diagram of longitudinally-averaged field that the polar field forms by a concentration of weaker fragments

(2) The fact that the poleward migration of filaments is earlier and more rapid than expected in the diffusion model

(3) The minimum in the distribution of white light faculae (and, by inference, a minimum in the density of magnetic flux) at mid-latitudes during the declining years of the cycle.

7.3 The Relative Strengths of Mid-latitude and Polar Fields in the Presence of a Meridional Flow

Diffusion is not a very efficient way of getting field remnants to the pole. Much of the flux is cancelled on the way, and the part that does get there doesn't have to stay. On the other hand, the polar fields are quite weak compared

to the total amount of flux made available by the major active regions which erupt during the course of the cycle. If we visualize the polar field (at sunspot minimum) as consisting of a cap of more-or-less uniform intensity, $B_{pole} \simeq$ 1 - 2 gauss, above some latitude $\lambda_p \simeq 60^\circ$, then the total polar flux would only be:

$$F_{polar} = (1 - \sin\lambda_p) 2\pi R_\odot^2 B_{pole} = (4-8) \times 10^{21} \text{ Mx} \qquad (7.1)$$

or about half the strength of a single plage in our nominal 10^{22} Mx 'average-sized' region. Since there are well over a thousand such 'average' regions in the course of the cycle, less than 0.1% of the original following polarity field would ever have to make it to the pole.

Indeed, one might worry, then, that a <u>systematic</u> flow, which could catch each available fragment, sweep it to the pole, and hold it there for the rest of the cycle would produce not too weak, but <u>too strong</u> a polar field. It is difficult to answer this problem directly, because, as we have seen in § 6.6 , there is some confusion about just how much residual field is actually left by the typical decaying active region -- it seems to be much less than would have been expected on the basis of random diffusion alone. About the best we can do, then, is to take the <u>observed</u> mid-latitude field (i.e., that in an area a little above the sunspot zones) and ask: if that flux were being systematically transported

towards the pole, would it produce a reasonable final field?

If we assume that the average field at some middle latitude, λ_M has a strength B_M, then the current of flux moving towards the pole will be:

$$\left(\frac{dF}{dt}\right)_{\lambda_M} \simeq 2\pi R_\odot \cos\lambda_M B_M v_M \tag{7.2}$$

where v_M is the speed of the flow. In order to make the polar fields work in the way that is observed, this current, integrated over the cycle, has to be sufficient to supply about <u>twice</u> the peak polar flux (once to cancel the existing field, and once again to establish the new one). That is:

$$\left(\frac{dF}{dt}\right)_{\lambda_M} \cdot T_{cycle} \simeq 2 F_{polar} \tag{7.3}$$

Combining the three equations we obtain:

$$B_M = \frac{2R_\odot(1-\sin\lambda_p)B_{pole}}{v_M \cos\lambda_M T_{cycle}} \tag{7.4}$$

Using $\lambda_p \simeq 60^\circ$ and $T_{cycle} = 11$ years, one finds that for a 3 m/sec flow, the average field at 45° would only have to be about one-quarter the strength of the peak polar field. This is not at all unlike what actually seems to be observed.

The same result could be obtained in a more intuitive way by noting that the total area of the mid-latitude zone (say, 30° to 60°) is approximately three times as great as the area above 60°. Thus, if the average flux density in

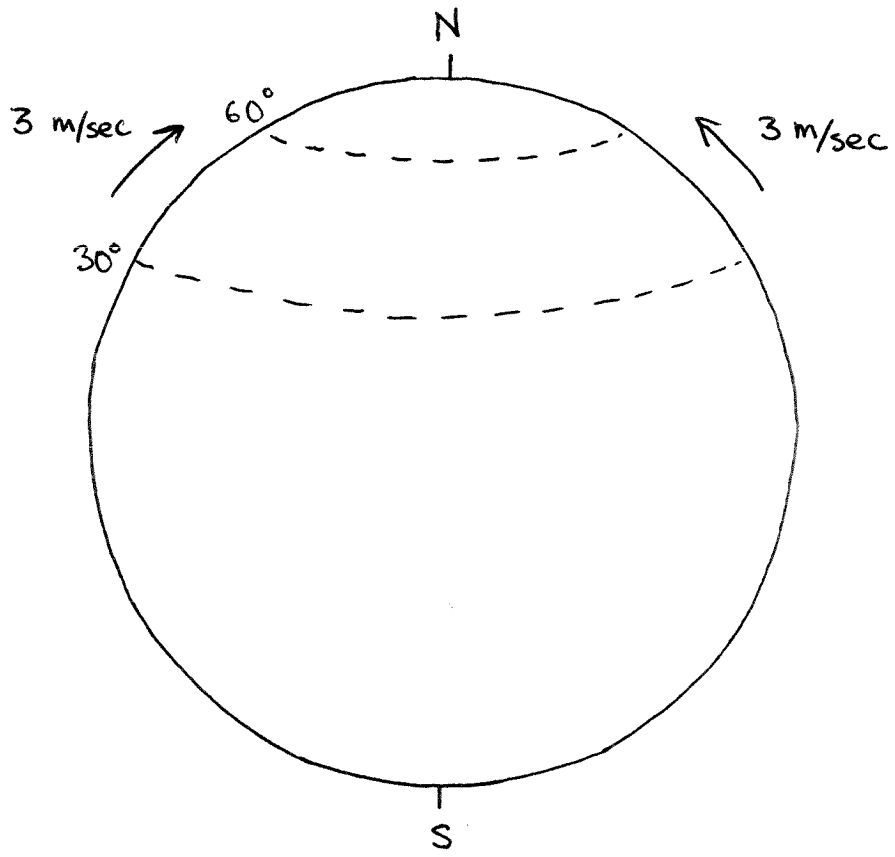


Figure 46: The Presumed Meridional Flow

the mid-latitudes was one-third that in the polar cap, then the mid-latitudes would have to be swept out about twice during the cycle to produce the polar field (once to cancel the old one, and once to build the new one). A 3 m/sec flow is just about sufficient to do that.

Whether or not the formation of the polar fields can be properly regarded in such simplistic terms is far from obvious. It is not even clear that the cycles which show the largest injection of sunspot sources produce the most intense polar fields (as one would expect in nearly any model where the polar field is formed from active region remnants). The 1957 sunspot maximum, for example, the largest on record ($R_{max} = 190$), produced only relatively weak and erratic polar fields at the 1964 minimum; whereas the 1969 spot maximum, which was only of average strength ($R_{max} = 106$), appears to have produced strong and easily detectable fields at the current sunspot minimum. Of course, this could be because the 1957 cycle was already working against a strongly established polar field (as seen at the 1958 minimum), whereas the 1969 cycle seems to have had relatively little established polar flux to cancel.

In any event, these are not problems which are unique to the meridional flow interpretation; and it is clear that the inclusion of that flow improves on the purely diffusional model not only by explaining how the polar fields could be con-

structed out of the very weak field remnants that seem to be available, but also by providing a rough quantitative 'justification' for the relative strengths of the polar versus mid-latitude fields that are actually seen.

7.4 Summary

The relatively low diffusion constant ($D \approx 200$ km^2/sec or $T_o \approx 80$ years) suggested by our measurements of the rate of active region spreading would, if confirmed, destroy much of the beauty of Leighton's (1964) calculations of the development of large scale field patterns over the course of the cycle. The active region remnants would, if only under the influence of diffusion at $D = 200$ km^2/sec, arrive at the pole much too late and with too little strength. It would be impossible, then, to explain, on the basis of the observed pattern of sunspot eruptions, either why the filaments migrate towards the pole as early in the cycle as they do, or why the polar field reaches its maximum intensity around the time of sunspot minimum.

We find, however, that the basic agreement can in large part be restored if one is willing to assume that in the middle latitudes the diffusional processes are assisted by a systematic poleward-moving meridional flow with a velocity ~ 3 m/sec. The migration of filaments is, then, automatically accounted for. The timing of the polar reversals and maxima can be thought of simply in terms of the available

source fields. The pole will intensify as long as there is a fresh supply of the appropriate polarity, but weaken as soon as the following parts of new cycle regions begin to inject the opposite sign. Moreover, the proposed current appears to reproduce the approximate ratio of field strengths between the poles and the mid-latitudes that is actually observed, and to explain the general impression one gets from the magnetic butterfly diagram that the polar field is formed more by concentration than by diffusion. The mid-latitude minimum in the observed distribution of white light faculae (during the declining years of the cycle) provides additional evidence for the existence of such a current.

8. Conclusions

8.1 The Need for Better Observations

We have attempted to use recent solar observations to follow the history of selected magnetic regions; and in so doing to develop questions through whose answers we might hope to gain a better understanding of the fundamental mechanisms underlying the process of active region decay. While a more careful analysis of the same data would undoubtedly result in a considerably more refined comprehension of the answers, a number of interesting points emerge even at the present relatively low level of sophistication; and a number of cautions and suggestions for those who might wish to attempt to improve upon the results.

Most importantly, we find that the random walk model of active region decay provides a remarkably useful key to interpreting and understanding the observations. The rate of increase in area and separation is well within the range that would be anticipated on the basis of the observed daily motions (provided one includes the effects of differential rotation), and those, in turn, are within the range that would be expected for motion in response to the changing supergranular pattern (although the velocity cells may last a little longer than generally assumed). The only obvious deviations from the pure random walk picture that have been uncovered, are that the annihilation of flux seems to be

more rapid and more thorough than anticipated; and that in order to form a proper polar field it would be necessary to assume that the diffusional decay is assisted, in the middle-latitudes, by a systematic poleward-moving meridional flow. It is important, then, to attempt to isolate the question of the geometric pattern of the decaying flux, from the question of its quantity.

Our best guess is that the diffusion constant (governing the geometric arrangement of field lines) is close to $D \approx 200$ km^2/sec (corresponding to $T_o \approx 80$ years in the models of Leighton, 1964, 1969); but, because of the crudeness of the present analysis, this could be off by perhaps as much as a factor of two. A diffusion constant as high as $D = 1000$ km^2/sec ($T_o = 15$ years) seems, however, definitely to be excluded.

This conclusion is based mainly on measurements regarding the separation between the opposite-polarity peaks in the decaying field distribution; and is somewhat sensitive to assumptions regarding the law of differential rotation. Thus, it is important, if the present method of analysis is considered interesting, that these conclusions be verified using a more precise modelling of the predicted pattern of evolution for the initial flux configurations that were actually observed, rather than relying completely on averages.

Similarly, if the measurements of area are repeated, it is important that it be clearly defined whether the measured areas correspond to the surface within which 2/3 or 90% or some other fraction of the surviving flux is confined; and also that proper allowance be made for the possibiility that even though the area seems to be growing at random, the flux may be being annihilated more rapidly than expected (we have avoided this problem here by using a "subjective" measure of the area, but at the expense of being uncertain as to how to interpret it). The most useful technique in this respect might well be to smear out the high-resolution magnetic observations in such a way that the width of the resulting Gaussian-like distribution could be taken as an objective measure of the area.

As far as the short-term motions of the individual features are concerned, we have found that two-dimensional cross-correlations offer, potentially, the most efficient and unambiguous way of determining the diffusion constant over periods of up to ~20 - 30 hours. The older measurements, based on photographs of the chromospheric network, should be repeated using magnetic cancellations. By carefully analyzing the short-term data one might be able to isolate the part of the diffusion which is due to granulation (probably very little) from that which is due to the more gradual supergranular motions. In any event it should be possible to

state quite precisely whether the observed short-term motions are exactly, or only roughly, sufficient to account for the long-term spreading.

Finally, the whole question of computing theoretical dispersal rates for field lines embedded in the supergranulation is one which we have barely touched; and which begs for a better definition of the observational parameters.

It will be noted, however, that the primary need here, as in so many areas, is not so much for new observations, as for a better analysis of those that already exist.

8.2 The Results

We come, then, to the summing up.

This work does not, obviously, represent the last or final word in the problem of active region decay. Many refinements are needed, some major, some minor; and in time, many of our assertions will probably be totally overturned. We do, however, feel sufficiently confident in the broad outlines to feel that it is a step in the direction of progress, and that the picture which we are about to paint lies somewhere between that first, tentative effort of Dowling (1946), and the final, ultimate, infinitely-precise description whose subtleties we cannot yet probably even guess.

Appreciating, then, that nearly every number in this list could be changed by a factor of ~2, and with due acknowledgement to those whose previous efforts have contributed

toward the development of this model, what we seem to have found is as follows:

1. The active region first appears as a localized bipolar eruption of magnetic flux, and grows, over a period of 4-5 days at a rate of $\sim 10^{16}$ Mx/sec , by a series of discrete sub-eruptions which coalesce to form a final configuration whose strength is roughly proportional to the total sunspot area.

2. Typical values are 10^{22} Mx for the plage of each sign, at a separation of $\sim$60,000 km, for a group with a sunspot area of 250 millionths (of a solar hemisphere).

3. The magnetic axes of the regions are systematically inclined, at an average of $+6^{\circ}$, with the preceding end closest to the equator.

4. The growth phase may be terminated by the development of a situation in which the rate of accidental recombination of opposite-polarity magnetic features becomes comparable to the eruption rate, but that is only speculation at this point.

5. Surging, flare and X-ray production are probably also related to the rate of magnetic recombination, which is expected to be greatest when the region is at its point of maximum development.

6. At the point of maximum development, about half the flux appears to be in spots, and half in plage. The transi-

tion from spot-strength to plage-strength field appears to be accomplished without effort, although the presence of major long-lived spots can delay the subsequent spreading in area.

7. Once the spots have disappeared, the development of the region seems to be dictated almost entirely by the condition that each field line move in response to the changing pattern of supergranular convection, with cancellation in those areas where opposite polarities overlap.

8. The absence of any obvious fine-scale mixing of polarities around the borders of decaying plages, and more particularly, along the neutral line separating the two halves of the decaying region, suggests that where overlap occurs, the cancellation is rapid and efficient.

9. Quantitatively, the observed increase in separation (between the peaks of the decaying opposite-polarity concentrations) at a rate of ~ 0.015 km/sec is consistent with a diffusion constant of $D \simeq 200$ km^2/sec, assuming that the sources are also, to some extent, pulled apart by differential rotation. The observed rate of increase in the area dominated by each polarity ($\sim 5 \times 10^3$ km^2/sec over 4 months), is also consistent with a diffusion constant $D \simeq 200$ km^2/sec (in the presence of differential rotation), assuming the measured areas correspond to that within which $\sim 90\%$ of the surviving flux is located.

10. This diffusion constant agrees closely with that obtained by direct measurement of the motions of individual magnetic features; and is sufficient to explain why, in 'quiet' areas, the network should be difficult to recognize after ~24 hours.

11. The quantity of surviving flux appears to be noticeably less than would be expected for purely random annihilation. The decay looks exponential on the Mount Wilson contour plots, but one does not know how to correct for the amount of flux that might be lost below the '5-gauss' contour. We estimate, however, that less than 0.1% of the injected following polarity ever reaches the poles.

12. The visibility lifetime of active region residues, over at least three orders of magnitude (in source strength), is also explicable in terms of simple diffusion ($D = 200$ km^2/sec), but only by assuming a threshold field of ~15 gauss (7½ 'Mount Wilson gauss') -- which probably corrects for the overly-rapid annihilation of flux.

13. The influence of a decaying region spreads in latitude as well as in longitude. Since the sources are disposed basically along lines of latitude, diffusion will attempt to establish a north-south neutral line, while differential rotation will attempt to draw it out east and west. The result is a compromise in which the neutral line is more nearly north-south than might have been expected, and as a

result, the weak field remnants appear to rotate more rigidly than their sources.

14. The scattered mixed-polarity component of the background field is probably due to small short-lived regions, which appear and dissipate in place. The larger patterns seem entirely to be the remnants of major active regions.

15. The proposed diffusion constant is, by itself, insufficient to account for either the poleward migration of filaments or the formation of the polar caps. We suggest that in the middle latitudes the basic diffusional process is assisted by a systematic meridional flow of ~ 3 m/sec, which sweeps up the long-term active region residues (which are much smaller than expected but still preferentially of the following sign), and transports them bodily to the poles.

It should be emphasized that each of these fifteen points refers only to an average pattern of evolution, which may not, in detail be followed by any single region. The actual observations suggest a great deal of scatter. We do not know what portion of that scatter is due to the crudeness of reduction, and what portion to a true diversity in the patterns of development available to individual regions; but it is clear that only by attempting to tackle such problems will we begin to be able to say that we truly understand the <u>non</u>-random processes which play a role in active region decay.

The questions, then, are there, and they are solvable, but much work remains to be done.

Appendix I Flux Measurements

The measurements of flux presented in this paper are rough estimates based on the area inside the magnetic contours as they appear on the 85 mm diameter full disk Mount Wilson magnetograms published in the NOAA Prompt Reports. The flux is computed according to the following formula:

$$F_{apparent} = (A_{10} - A_5)x7\tfrac{1}{2} + (A_{20} - A_{10})x15 + (A_{40} - A_{20})x30 + (A_{80} - A_{40})x60 + A_{80}x120 \qquad (A1.1)$$

or, equivalently,

$$F_{apparent} = (A_5 + A_{10})x7\tfrac{1}{2} + A_{20}x15 + A_{40}x30 + A_{80}x60 \qquad (A1.2)$$

where A_n indicates the total area inside the 'n-gauss' contour (5, 10, 20, 40, and 80 gauss levels are normally plotted; 1 mm = 16,380 km). Since the regions studied are generally near disk-center, the measured field strengths will be low due to line weakening. In accordance with the results of Howard and Stenflo (1972), the nominal flux is multiplied by a factor of 2 to get the true flux.

In addition, some correction has to be made for the sunspot flux, which is not properly recorded on the normal contour plots. This correction is made using the sunspot areas and umbral field strengths listed in the 'Regions of Solar Activity' section of the NOAA Prompt Reports.

If we use the formula given by Kiepenheuer (1953), that the distribution of field strength in a spot goes like:

$$B(r) = B_m \left(1 - \frac{r^2}{b^2}\right) \tag{A1.3}$$

where b is the penumbral radius; then the total flux of a spot should be:

$$F_{spot} = \int_0^b B(r)\, 2\pi r\, dr = \frac{1}{2} B_m \cdot \pi b^2 \tag{A1.4}$$

that is, half the peak field times the total area.

The NOAA reports do not generally specify in any quantitative fashion how the spots are divided between p- and f-polarity, so, unless it is obvious from the sunspot drawing that the spots are all at one end, the correction is divided equally between the two plages.

We have neglected certain considerations involving field geometry in computing equation (A1.4); and no effort has been made to correct the plage fluxes for fields not indicated by the 5 gauss and higher contours.

Appendix II List of Regions

The following table lists, by McMath plage number, all the regions used in the present study of active region development and decay. The numbers are intended to represent the source strength, that is, the flux and separation at the point of maximum development. For simplicity, however, the measurements were usually made when the region was close to the central meridian, and hence may not correspond precisely to the point of maximum development that would be disclosed in more careful day by day measurements.

The column labelled 'Age' gives the approximate time (in days) from the birth of the region to the time to which the measurements of flux, separation, and sunspot area refer. The symbol CM is used when the birth of the region was not witnessed. Such a measurement, made simply at the first central meridian passage, could refer to an age anywhere between 7 and 21 days. When the age is marked by an asterisk (*), this means that it was evident from the appearance of the region on the next rotation that it had reached its maximum development (either by continued growth or by resurgence) on the back side. In such cases, the flux is estimated by extrapolating backwards the subsequent curve of flux with time, and is somewhat higher than either of the first two measurements; the separation and tilt are averages of those seen on the first two rotations; and the sunspot area is the largest

actually reported.

The 'Tilt' of the magnetic axis is given in degrees, with positive meaning that the preceding polarity is closest to the equator. No tilt angle is given for two regions which develop complicated reversed polarity magnetic configurations.

'Fluxes' are in maxwells, and have been corrected for line weakening and sunspot fields as explained in Appendix I. They are the average of observations for the two plages.

The 'Separation' is the total distance between the apparent peaks of the positive and negative polarity concentrations, and wherever possible, the numbers given are an average of estimates made from the Kitt Peak and the Mount Wilson magnetograms.

The 'Sunspot Areas' are the largest values for the particular region, as reported in the NOAA bulletins. In some cases, obviously, the true maximum spot area may have occurred when the region was behind the limb. The units are millionths of a solar hemisphere (1 millionth = 3.04×10^6 km^2).

The 'Lifetimes' (in seconds) are rough estimates of the total time over which the remnants of the region remained distinguishable from the normal 'undisturbed' magnetic background. Generally, they were obtained by noticing that a region was visible on one rotation and not on the next. In many cases, however, the observations had to be terminated due to confusion with nearby activity. In such cases only

a lower limit could be placed on the lifetime. Note that 1 rotation = 2.35×10^6 sec.

The regions are ranked in order of decreasing source flux. McMath 11958 is a complex of activity that developed over several rotations.

To complete the table, we might note that, according to Harvey and Martin (1973), the spotless ephemeral regions, which never attain McMath plage status, have typical source strengths of $N \sim 10^{20}$ Mx and $2a \sim 15{,}000$ km (Harvey and Martin's definition of 'typical dimension' corresponds to roughly twice our definition of 'separation'). Their lifetimes are 1-2 days or less, and their inclinations are nearly random. This is, of course, only one in a continuous spectrum of region sizes to which there is no clear lower limit.

None of the measured values, incidentally, is likely to be accurate to more than one significant figure, but they are reported as they were recorded.

Table A2-1 List of Regions by Source Flux

McMath #	Age (days)	Lat.	Tilt	Flux (N_o)	Separation (2a: km)	Spot Area	Lifetime (sec)
11957	22	13° N	--	4.2×10^{22}	1.1×10^{5}	1120	$\sim 2\times10^{7}$
13225	CM*	7° N	--	4×10^{22}	5×10^{4}	1000	$\sim 3\times10^{7}$
13790	CM	9° N	-2	3.0×10^{22}	1.2×10^{5}	1140	2.2×10^{7}
13722	3*	7° N	+10°	2×10^{22}	8×10^{4}	460	$\sim 2.5\times10^{7}$
13736	3*	8° S	+8°	2×10^{22}	5×10^{4}	270	$>7\times10^{6}$
11958	--	11° S	+19°	1.7×10^{22}	1.6×10^{5}	500	$\geq 9.6\times10^{6}$
13338	CM	14° N	+3°	1.3×10^{22}	6.6×10^{4}	700	$\sim 1.5\times10^{7}$
13373	5*	12° N	+3°	1×10^{22}	5×10^{4}	60	$\sim 1.1\times10^{7}$
13278	10	19° S	+6°	9.5×10^{21}	5.7×10^{4}	150	$\sim 1.3\times10^{7}$
13343	CM	9° N	-11°	9.1×10^{21}	9×10^{4}	25	$\sim 1.8\times10^{7}$
11972	4	7° N	+18°	9×10^{21}	7.5×10^{4}	210	$>5\times10^{6}$
13890	5	5° N	0	8.2×10^{21}	6.5×10^{4}	220	1.4×10^{6}
13738	CM	7° N	+10°	6.6×10^{21}	8.2×10^{4}	290	1.2×10^{6}
13818	2*	13° S	+12°	6×10^{21}	8×10^{4}	200	$\sim 1.4\times10^{7}$
13783	4	9° N	+12°	5.6×10^{21}	5.7×10^{4}	140	$>2.6\times10^{6}$
11993	6	25° S	+26°	4.9×10^{21}	7.4×10^{4}	120	$\geq 5\times10^{6}$
13875	CM	33° N	-8°	4.3×10^{21}	5.7×10^{4}	140	$\sim 6\times10^{6}$
13341	4½	3° N	-17°	3.4×10^{21}	5×10^{4}	30	$\sim 3.9\times10^{6}$
12050	2*	16° N	0°	3×10^{21}	5×10^{4}	30	$\geq 3\times10^{6}$
13532	6	13° S	+9°	3×10^{21}	1.1×10^{5}	80	7×10^{6}
14120	4	34° N	+18°	2.1×10^{21}	3.8×10^{4}	50	--

Table A2-1

(continued)

McMath #	Age (days)	Lat.	Tilt	Flux (N_o)	Separation (2a: km)	Spot Area	Lifetime (sec)
13517	3*	13° N	+11°	2×10^{21}	7×10^{4}	50	6×10^{6}
13992	4	4° S	-12°	2.0×10^{21}	4.1×10^{4}	50	$>4.3\times10^{6}$
12008	2½	18° N	+9°	1.7×10^{21}	6.6×10^{4}	20	$<2\times10^{6}$
13611	CM	11° S	+17°	1.2×10^{21}	8.2×10^{4}	0	6.3×10^{6}
11989	CM	14° N	-6°	1.0×10^{21}	4.7×10^{4}	20	$\lesssim 2\times10^{6}$
13811	1*	28° N	+3°	1×10^{21}	6×10^{4}	30	$\sim 6\times10^{6}$
13859	4	7° N	+3°	7.6×10^{20}	4.1×10^{4}	20	1.2×10^{6}
13927	5	8° S	+3°	6.4×10^{20}	4.1×10^{4}	10	$\sim 3.10^{6}$
13340	4	4° N	0	3×10^{20}	2.4×10^{4}	10	$\sim 1.10^{6}$
13924	1	4° N	+4°	2.8×10^{20}	2.4×10^{4}	10	3.9×10^{5}
13499	3	10° N	+29°	2×10^{20}	3×10^{4}	10	5.6×10^{5}
13763	2	10° N	+16°	8×10^{19}	2.4×10^{4}	0	3.9×10^{5}

With the exception of Region 11957 (the complex of activity), all regions in the text are referred to by the McMath plage number which they were assigned on their first disk passage. In many cases, however, the regions maintained sufficient strength to be assigned new plage numbers on their second, or even later, passages.

Table A2-1 gives, in order of increasing number, each of the 33 regions used in this study, and, where appropriate, the numbers which were assigned to them on successive rotations (counting the first appearance as 'Rotation 1'). Asterisks are used to indicate cases where the identification of the plage number with the magnetic remnants of the original region is partial or imperfect (usually due to resurgence).

This is meant only to clarify the identification of the regions, and not (except roughly) to indicate their lifetimes. The remnants of many regions remain visible magnetically long after the plage numbers have ceased to be assigned.

Table A2-2 Identification of Regions on Successive Rotations
(McMath Calcium Plage Number)

Rot. 1	Rot. 2	Rot. 3	Rot. 4	Rot. 5	Rot. 6
11957	11976	12007	12045	-	
11958	11986	12025	-		
11972	11994*	12029*	-		
11989	-				
11993	12031	-			
12008	-				
12050	12077	-			
13225	13280	13324	-		
13278	13318	13361	-		
13338	13380	13423	13471	13526	-
13340	-				
13341	13384	-			
13343	13382	13426	-		
13373	13411	13463	-		
13499	-				
13517	13547	-			
13532	13568	13629*	-		
13611	-				
13722	13750	13786	13820	13865*	-
13736	13766	13796	13832*	-	
13738	13777	13808*	-		

Table A2-2

(continued)

Rot. 1	Rot. 2	Rot. 3	Rot. 4	Rot. 5	Rot. 6
13763	-				
13783	-				
13790	13826	13867	13906	13945	-
13811	13831	-			
13818	13849	13892	13933*	13970*	14013*
13859	-				
13875	13914	-			
13890	13926	13965*	14004*	-	
13922	13960	13992	-		
13924	-				
13927	-				
14120	14152*	no data			

Appendix III Diffusion Paper, Curve Fitting and Errors

When random motions occur in two dimensions, it is common to find particles with displacements, r, distributed according to a probability function of the form:

$$P(r)dr = \frac{2r}{r_o^2} e^{-r^2/r_o^2} dr \qquad (A3.1)$$

where r_o is the rms displacement.

The measured displacements, when presented in the form of a histogram, showing the number of results in different intervals, will generally show roughly this form, but the mean-squared displacement can vary considerably, depending on the chance occurrence of a few large-displacement events. One wishes, then, to develop a method for fitting a curve of the form (A3.1) to the data which does not rely unduly on the high-displacement points.

One simple method which suggests itself is to consider the integrated distribution $f(r<r')$, that is, the total number of events with displacements less than some threshold value r'. According to equation (A3.1):

$$f(r<r') = \int_{r=0}^{r'} P(r)dr = 1 - e^{-r'^2/r_o^2} \qquad (A3.2)$$

This can be converted into a linear relationship by means of a little algebra:

$$r' = r_o \sqrt{-\ln(1-f)} \qquad (A3.3)$$

In other words we expect a linear relationship if we plot $\sqrt{-\ln(1-f)}$ against r'. This can be done most easily by constructing a special non-linear graph paper on which the horizontal scale is calibrated in terms of the indicated quantity. The process is analogous to the construction of ordinary "probability paper" for studying a normal Gaussian distribution. Figure A3-1 shows an example of the paper, on which the data for the K-line displacements shown in Figure 29 of the text, at intervals of $2\frac{1}{2}$, 5, and 10 hours, have been plotted. If the observed distribution differs from the form of equation (A3.1), by having too few or too many high displacement points, the data will veer off the line at the larger values of f. The points for $\Delta t = 2\frac{1}{2}$ hours show an example of this. The effect of these points can be minimized by ignoring them in drawing the best-fit straight line.

For those who may be interested, the positions of the horizontal ticks (in arbitrary units) are:

Table A3 - 1 Diffusion Paper

f	$\sqrt{-\ln(1-f)}$	f	$\sqrt{-\ln(1-f)}$	f	$\sqrt{-\ln(1-f)}$
0	0	.30	.5972	.70	1.0973
.01	.1002	.40	.7147	.80	1.2686
.05	.2265	.50	.8326	.90	1.517
.10	.3245	.60	.9572	.95	1.7308
.20	.4724	.632	1.0000	.98	1.9779

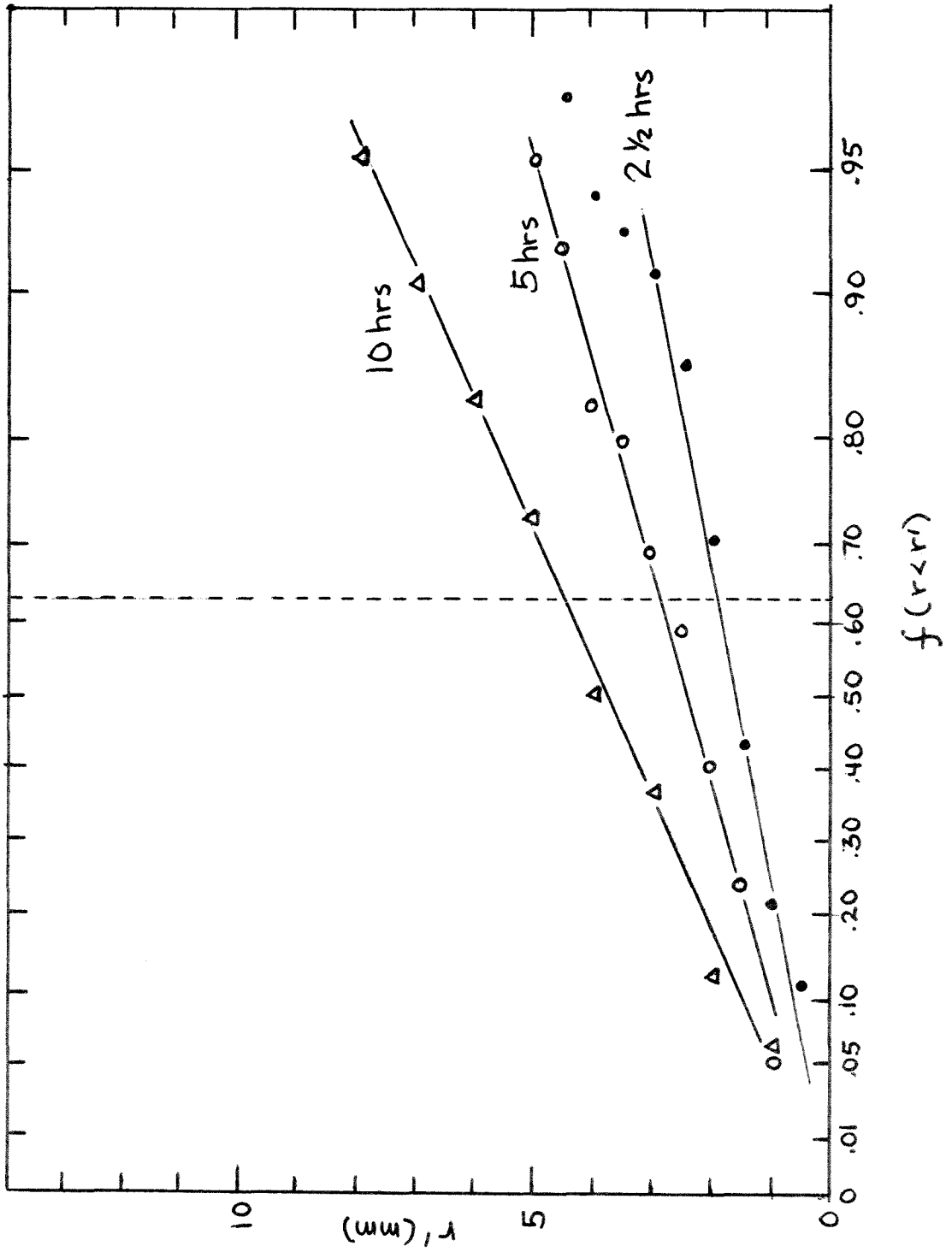


Figure A3-1: Diffusion Paper

According to equation (A3.1) the best fit curve is determined by the single parameter r_o (plus the overall normalization to the number of points); and equation (A3.3) implies that this quantity can be read directly off the "diffusion paper graph" in any one of a number of ways. For example:

$$\begin{aligned} r_o &= 1.414...r' \quad \text{when} \quad f = .3935... \\ r_o &= 1.2011..r' \quad \text{when} \quad f = .50 \\ r_o &= r' \quad \text{when} \quad f = .6321... \end{aligned} \tag{A3.4}$$

which correspond to fitting the curve at (a) the peak in the histogram; (b) the median; and (c) the point at which $r' = r_o$. We can call these three 'natural' fitting points r_{39}, r_{50}, and r_{63}, respectively. Figure A3-1 illustrates the use of r_{63} to determine r_o (note that one does not actually have to use diffusion paper to determine the median or either of the other two fitting points, but to see the straight line is heartening).

In practice, each of the fitting points was read off the graph and the average of the three inferred values of r_o was used in all subsequent calculations.

To estimate the reliability of this method of determining r_o from the median of the observed distribution, when the histogram is based on a limited amount of data, we have to ask ourselves: if we were to draw, at random, a sampling of N points from a population distributed <u>exactly</u> accord-

ing to equation (A3.1), how close would we expect the median of those N points to be to the nominal value ($r_{50} = .8326..r_o$)?

Without going into details, this question is answered by saying that if on the average $N/2$ of the points lie to the left of the median, and $N/2$ to the right; then in any particular sample, those numbers could (with 2/3 confidence) fluctuate by $\pm\sqrt{N}/2$ (the difference between them fluctuates by $\pm\sqrt{N}$). Looking at the probability density you can estimate how much that excess (or deficiency) is likely to move the measured median, and in turn how much that could change the inferred value of r_o . The result is:

$$\Delta r_o = 0.7231.. \frac{r_o}{\sqrt{N}} \qquad \text{(A3.5)}$$

where Δr_o is the accuracy of inferred rms displacement based on the median of a sample of N points distributed according to equation (A3.1).

In principle, the accuracy of the present method is a little higher, because by drawing a straight line on the diffusion paper we are smoothing out some of the statistical fluctuations. But there is also some fudging involved in 'disregarding' the high displacement tail. For that reason, we have rounded off the 0.72... to 1 . The errors quoted in the text of the paper, then, are obtained simply by dividing the inferred rms displacement by the square root of the

number of measurements. The mean-squared displacements have twice that fractional error.

N.B.: If the same reasoning were applied to a normal gaussian distribution, one would decide that (at the 2/3 confidence level) the median based on a sample of N points should be within:

$$\Delta\mu = \pm\sqrt{\frac{\pi}{2}}\,\frac{\sigma}{\sqrt{N}} = \pm 1.253\ldots\,\frac{\sigma}{\sqrt{N}} \qquad \text{(A3.6)}$$

of the true central value, where σ is the standard deviation.

Appendix IV: Separations Between Randomly Moving Particles

The motion of magnetic features can be studied either by measuring directly the displacement of individual points (as in Figure 29), or by studying variations in the separation between selected pairs.

As stated in the text, there are certain advantages in using the separations; but also, certain subtleties in their interpretation. In Figure A4-1(a) we visualize two points, separated initially by a vector $\vec{s}$, subject to random displacements $\vec{\Delta r}_1$ and $\vec{\Delta r}_2$. The length of the displacement vectors is taken to be ℓ, the "true" size of the individual point displacements. The net effect of $\vec{\Delta r}_1$ and $\vec{\Delta r}_2$ will be to add an amount $\vec{\Delta s}$ to the original separation, where $\vec{\Delta s}$ is a random vector of length $\sqrt{2}\,\ell$.

Thus, by the "law of cosines", the connection between the original separation, $|\vec{s}|$, and the final separation, $|\vec{s}'| = |\vec{s} + \vec{\Delta s}|$, is:

$$|\vec{s}'| = \sqrt{|\vec{s}|^2 + |\vec{\Delta s}|^2 + 2|\vec{s}||\vec{\Delta s}|\cos\theta} \qquad \text{(A4.1)}$$

where θ is the angle between $\vec{s}$ and $\vec{\Delta s}$ (which is equally likely to assume any value between 0 and 2π).

In searching for spatial coherencies in the pattern of motion, the natural thing is to try to show that the mean-squared change in separation for closely-spaced pairs is different from the mean-squared change in separation for widely-spaced pairs.

According to equation (A4.1) this quantity is given by:

$$\langle(|\vec{s}'|-|\vec{s}|)^2\rangle = 2\epsilon^2 + 2\ell^2 - \frac{4s}{\pi}(s+\sqrt{2}\ell)\int_0^{\pi/2}\sqrt{1-\frac{4\sqrt{2}\ell s\sin^2\gamma}{(s+\sqrt{2}\ell)^2}}\,d\gamma \qquad \text{(A4.2)}$$

which can be re-written in the form:

$$\langle(|\vec{s}'|-|\vec{s}|)^2\rangle = 2\epsilon^2 + 2\ell^2 - \frac{4s}{\pi}(s+\sqrt{2}\ell)\,E\left(\frac{2(\sqrt{2}\ell s)^{1/2}}{s+\sqrt{2}\ell}\right) \qquad \text{(A4.3)}$$

Where E is the complete elliptic integral of the second kind (Abramowitz & Stegun, 1965; § 17.3.4).
This function (eqn. A4.3) is sketched in Figure A4-1(b).

It will be noted that even for completely random motion there is a difference in the expected mean-squared change in separation of close- and wide-spaced pairs. If the points are extremely close together, the mean-squared change in separation will be twice the mean-squared displacement. However, in practice we will not be able to explore this regime because we cannot distinguish the motions of (unlabelled) features whose separation is closer than the typical displacement.

Thus, in our solar work, the rule will be that the mean-squared change in separation is equal to the mean-squared displacement, as shown in the latter part of Figure A4-1(b). The reason that it is equal, and not twice, as might have been guessed, is that only that component of displacement which acts parallel to the original direction of separation is effective at increasing the separation. The perpendicular

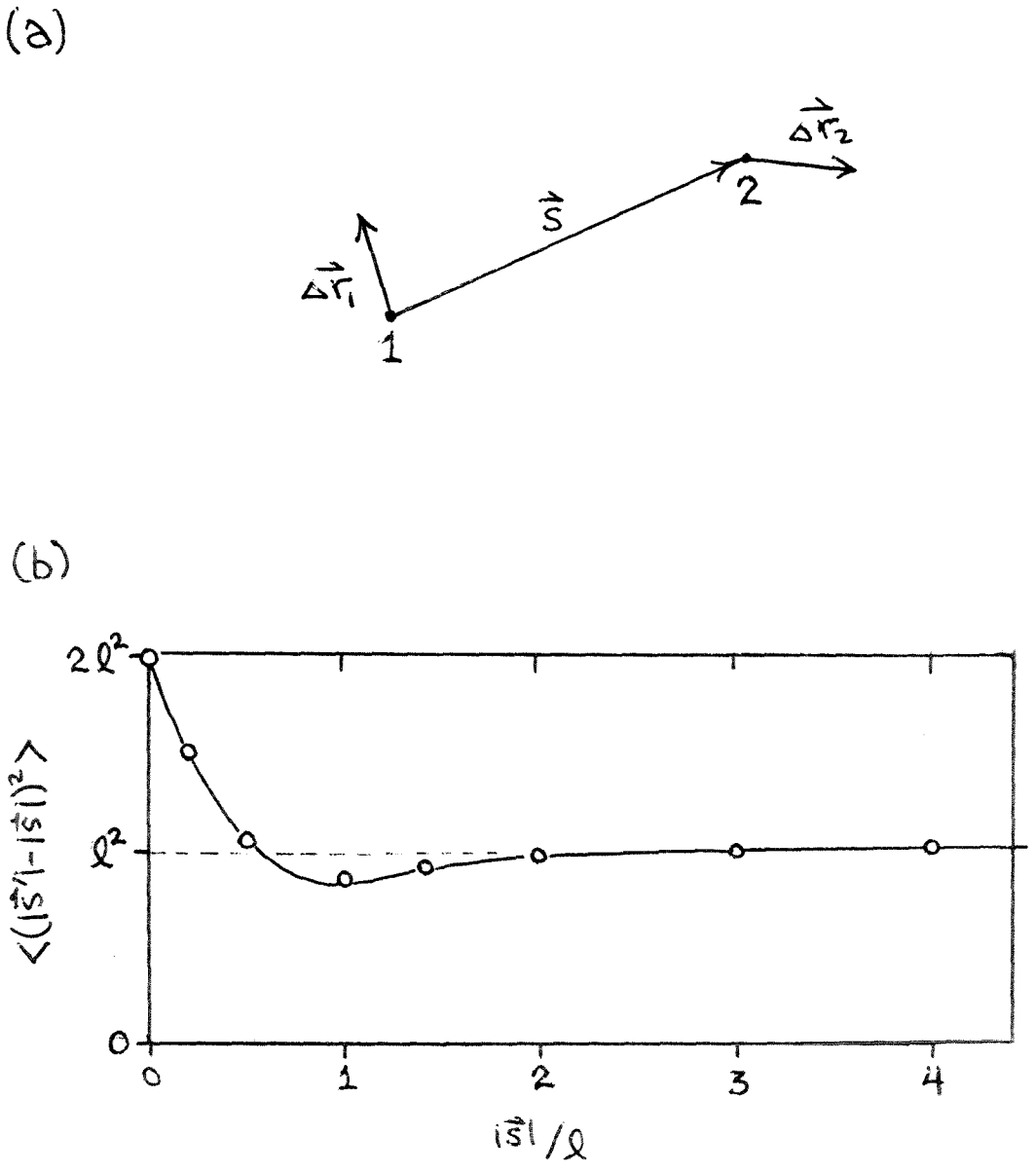


Figure A4-1: The mean-squared change in separation for randomly moving particles

component changes the vector separation, but it does not add significantly to the length.

Appendix V The Distribution of Nearest Neighbors in a Random Array

To determine the distribution of nearest neighbors in a random array, consider N points placed at random in an area A, of large dimensions. We select one of these points. Call it P. The probability that P will have a nearest neighbor within a radius $r \to r + dr$ can be found by imagining that all the other points are suddenly removed, and then replaced, one by one. Each falls with an equal probability, $1/A$, of landing in any given unit area. In order for P to have its nearest neighbor in the range $r \to r + dr$, these other points must come down in such a way that there is <u>none</u> closer than r, but <u>one or more</u> in the annulus $r \to r + dr$. Since the probability that all N points will miss the circle of radius r around P is $(1 - \frac{\pi r^2}{A})^N$; and since the probability of something falling in the annulus of radius $r \to r + dr$ is the sum of the probabilities for each of N identical tries; the probability of P's nearest neighbor being at $r \to r + dr$ is:

$$P(r)\,dr = \left(1 - \frac{\pi r^2}{A}\right)^N \cdot \left(\frac{N 2\pi r\,dr}{A}\right) \tag{A5.1}$$

This can be simplified by taking the limit where $A \to \infty$, but in such a way that the density, $n_o \equiv N/A$, is held constant, i.e.:

$$P(r)\,dr = \lim_{A\to\infty} \left(1 - \frac{\pi r^2}{A}\right)^N \left(\frac{N 2\pi r\,dr}{A}\right) = \lim_{N\to\infty} \left(1 - \frac{n_o \pi r^2}{N}\right)^{N} (2\pi n_o r\,dr) \tag{A5.2}$$

Or:

$$n(r)\,dr = 2\pi n_0 r\, e^{-n_0 \pi r^2}\, dr \tag{A5.3}$$

This is the expression used in the text. It is normalized to unity:

$$\int_{r=0}^{\infty} n(r)\,dr = 1 \tag{A5.4}$$

and has the standard modified-gaussian form which we have come to recognize as being characteristic of the distribution of random displacements in two-dimensions.

The most frequent nearest-neighbor distance (the peak of the distribution) is:

$$r_{max} = \frac{1}{\sqrt{2\pi n_0}} \tag{A5.5}$$

The mean-squared nearest-neighbor distance is:

$$\langle r^2 \rangle = \int_0^{\infty} r^2 n(r)\,dr = \frac{1}{n_0 \pi} \tag{A5.6}$$

References

Abramowitz, M. and Stegun, I. (1965) Handbook of Mathematical Functions (New York: Dover Publications).

Allen, C.W. (1973) Astrophysical Quantities (University of London: Athlone Press).

Babcock, H.W. (1953) "The Solar Magnetograph" Astrophys. J. 118, 387.

Babcock, H.W. and Babcock, H.D. (1955) "The Sun's Magnetic Field, 1952-1954" Astrophys. J. 121, 349.

Babcock, H.W. (1963) "The Sun's Magnetic Field" Annual Reviews of Astron. and Astrophys. 1, 41.

Beckers, J.M. (1971) "The Measurement of Solar Magnetic Fields" in IAU Symposium 43 (Howard, ed.), 3.

Bocchino, G. (1933) "The Drift of the Prominences During the Eleven Year Cycle of Solar Activity" (transl.) Mem. Societá Astronomica Italiana (N.S.) 6, 479.

Brunner, W. (1930) "Trends in the Arrangement of Sunspots in Groups" (transl.) Astron. Mitt. Zurich 124, 67.

Bruzek, A. (1967) "The Physics of Solar Flares: the Energy and Mass Problem" in Solar Physics (Xanthakis, ed.; New York: Interscience), 399.

Bumba, V. and Howard, R.F. (1965a) "A Study of the Development of Active Regions on the Sun" Astrophys. J. 141, 1492.

Bumba, V. and Howard, R.F. (1965b) "Large Scale Distribution of Solar Magnetic Fields" Astrophys. J. 141, 1502.

Butler, C.P. (1922) "Systematic Distribution of Solar Calcium Flocculi: I. Inclination of Elongated Groups" Monthly Notices Roy. Astron. Soc. 82, 334.

Butler, C.P. (1924) "Systematic Distribution of Solar Calcium Flocculi: II. Life History of Bright Solar Calcium Flocculi" Monthly Notices Roy. Astron. Soc. 84, 134.

Cowling, T.G. (1946) "The Growth and Decay of Sunspot Magnetic Fields" Monthly Notices Roy. Astron. Soc. 106, 218.

Cowling, T.G. (1953) "Solar Electrodynamics" in The Sun (Kuiper, ed.), 532.

Dravins, D. (1975) "Horizontal Velocities in the Solar Photosphere" Solar Phys. 40, 53.

Dunn, R.B. and Zirker, J. (1973) "The Solar Filigree" Solar Phys. 33, 281.

Frazier, E.N. (1972) "The Magnetic Structure of Arch-Filament Systems" Solar Phys. 26, 130.

Glackin, D.L. (1973) "Spatial Distribution of Emerging Flux Regions" Publ. Astron. Soc. Pacific 85, 241.

Glackin, D.L. (1974) "Differential Rotation of Solar Filaments" Solar Phys. 36, 51.

Glackin, D.L. (1975) "Emerging Flux Regions" Solar Phys. 43, 317.

Gnevishev, M.N. (1938) "On the Life-Length of Sun-Spots" (transl.) Pulkovo Obs. Circ. 24, 37.

Greenwich Observatory (1925) "Characteristic Movements of Sunspots" Monthly Notices Roy. Astron. Soc. 85, 553.

Hale, G.E., Ellerman, F., Nicholson, S.B., and Joy, A.H. (1919) "The Magnetic Polarity of Sunspots" Astrophys. J. 49, 167.

Hale, G.E. and Nicholson, S.B. (1938) Magnetic Observations of Sunspots, 1917-1924 (Washington: Carnegie Institution), 2 vol.

Harvey, K.L. and Martin, S.F. (1973) "Ephemeral Active Regions" Solar Phys. 32, 389.

Harvey, K.L., Harvey, J.W. and Martin, S.F. (1975) "Ephemeral Active Regions in 1970 and 1973" Solar Phys. 40, 87.

Howard, R. and Harvey, J. (1970) "Spectroscopic Determinations of Solar Rotation" Solar Phys. 12, 23.

Howard, R. and Stenflo, J. (1972) "On the Filamentary Nature of Solar Magnetic Fields" Solar Phys. 22, 402.

Howard, R. and Edberg, S.J. (1973) "On the Random Nature of the Eruption of Magnetic Flux on the Solar Surface" Solar Phys. 28, 73.

Howard, R. (1974a) "Studies of Solar Magnetic Fields: I. The Average Field Strengths" Solar Phys. 38, 283.

Howard, R. (1974b) "Studies of Solar Magnetic Fields: II. The Magnetic Fluxes" Solar Phys. 38, 59.

Janssens, T.J. (1970) "Long Term Observations of the Hα Chromospheric Network" Solar Phys. 11, 222

Kiepenheuer, K.O. (1953) "Solar Activity" in The Sun (Kuiper, ed.), 322.

Kraichnan, R.H. (1974) "Convection of a Passive Scalar by a Quasi-uniform Random Straining Field" J. Fluid Mech. 64, 737.

Leighton, R.B. (1957) "Some Observations of Solar Granulation" Publ. Astron. Soc. Pacific 69, 497.

Leighton, R.B. (1959) "Observations of Solar Magnetic Fields in Plage Regions" Astrophys. J. 130, 366.

Leighton, R.B., Noyes, R.W., and Simon, G.W. (1962) "Velocity Fields in the Solar Atmosphere: I. Preliminary Report" Astrophys. J. 135, 474.

Leighton, R.B. (1964) "Transport of Magnetic Fields on the Sun" Astrophys. J. 140, 1547.

Leighton, R.B. (1969) "A Magneto-Kinematic Model of the Solar Cycle" Astrophys. J. 156, 1.

Macris, C.J. (1962) "Studies on the Flocculi of the Solar Chromosphere" Mem. Societá Astronomica Italiana 33, 85.

Newton, H.W. and Nunn, M.L. (1951) "The Sun's Rotation Derived from Sunspots 1934-1944 and Additional Results" Monthly Notices Roy. Astron. Soc. 111, 413.

Noyes, R.W. (1963) Observations of Oscillatory Motions in the Solar Atmosphere (Doctoral Thesis: Calif. Inst. Tech.)

Parker, E.N. (1955) "The Formation of Sunspots from the Solar Toroidal Field" Astrophys. J. 121, 491.

Parker, E.N. (1963) "Kinematical Hydromagnetic Theory and its Application to the Low Solar Photosphere" Astrophys. J. 138, 552.

Roberts, P.H. (1970) Velocity Fields in Magnetically Disturbed Regions of the H_α Chromosphere (Doctoral Thesis: Calif. Inst. Tech.).

Rogers, E.H. (1970) "Lifetime of the H_α Chromospheric Network" Solar Phys. 13, 57.

Schröter, E.H. and Wöhl, H. (1975) "Differential Rotation, Meridional and Random Motions of the Solar Ca^+ Network" Solar Phys. 42, 3.

Sheeley, N. R. (1964) "Polar Faculae during the Sunspot Cycle" Astrophys. J. 140, 731.

Sheeley, N.R. (1965) Measurements of Solar Magnetic Fields (Doctoral Thesis: Calif. Inst. Tech.).

Sheeley, N.R. (1966) "Measurements of Solar Magnetic Fields" Astrophys. J. 144, 723.

Simon, G.W. (1963) Correlations between Large-Scale Solar Photospheric and Chromospheric Motions, Ca II (K) Emission, and Magnetic Fields (Doctoral Thesis: Calif. Inst. Tech.).

Simon, G.W. and Leighton, R.B. (1964) "Velocity Fields in the Solar Atmosphere: III. Large-Scale Motions, the Chromospheric Network and Magnetic Fields" Astrophys. J. 140, 1120.

Smithson, R.C. (1972) A Videomagnetograph Study of Diffusion of Solar Magnetic Fields in Weak Plage Regions (Doctoral Thesis: Calif. Inst. Tech.).

Smithson, R.C. (1973) "Videomagnetograph Studies of Solar Magnetic Fields: I. Magnetic Field Diffusion in Weak Plage Regions" Solar Phys. 29, 365.

Stenflo, J.O. (1972) "Evolution of Solar Magnetic Fields Over an 11-year Period" Solar Phys. 23, 307.

Stenflo, J.O. (1973) "Magnetic-Field Structure of the Photospheric Network" Solar Phys. 32, 41.

Tandberg-Hansen, E. (1967) Solar Activity (Waltham, Mass.: Blaisdell), 266.

Vorpahl, J. and Pope, T. (1972) "Solar Bright Points in 3840 Å and H_α" Solar Phys. 25, 347.

Waldmeier, M. (1941, 1955) Ergebnisse und Probleme der Sonnenforschung (Leipzig: Akademische Verlagsgesellschaft).

Waldmeier, M. (1946-1975) Astron. Mitt. Zurich (Yearly Reports of Solar Activity).

Waldmeier, M. (1947) Publ. Zurich Observ. 9, 1.

Waldmeier, M. (1960) "Circulation and Magnetic Field of the Sun's Polar Zone" (transl.) Zeitschrift für Astrophys. 49, 176.

Weart, S. (1972) "What Makes Active Regions Grow?" Astrophys. J. 177, 271.

Wilcox, J.M. and Howard, R. (1970) "Differential Rotation of the Photospheric Magnetic Field" Solar Phys. 13, 251.

Zirin, H. and Tanaka, K. (1973) "The Flares of August 1972" Solar Phys. 32, 173.