INTRODUCTION

TO

THE STUDY OF SUNSPOTS

BY

SALEM HANNA KHAMIS

Presented to the Department of Physics
of the American University of Beirut
in Partial Fulfilment of the Requirements
for the Degree of
Master of Arts

1941-1942

"Repeated observations have finally convinced me that these spots are substances on the surface of the solar body where they are continuously produced and where they are also dissolved, some in shorter and others in longer periods. And by the rotation of the sun, which completes its period in about a lunar month, they are carried round the sun; an occurrence important in itself and still more so for its significance."(1)

Galileo Galilei

---

(1) From an address by Galileo, to the Grand Duke Cosimo II, announcing his discovery of sunspots. (Handbuch der Astrophysik Band IV p. 57)

PREFACE

The following pages are simply a collection and a systematization of the most important developments in the field of solar physics which are thought to be of great importance to the study of the phenomenon of sunspots. Almost in all cases the original works of the investigators on the subjects discussed were consulted when they could be found in the books and magazines available in the library of the American University of Beirut. Slight alterations were made especially in the case of the hydrodynamical theory of Bjerknes, where a correction is made for a slight mistake.

A simple hypothesis for the explanation of the phenomenon of equatorial acceleration was put forward in a very elementary form; and necessary introductions were introduced to complete the discussion. However, other topics which were intended to be put in this report, were cancelled--such as the effect of solar phenomena on the terrestial atmospheric disturbances, and the researches of Störmer on solar electric vortices.

Effort was made to give the derivation of most of the important equations, yet the important gas equation remains without a proof for lack of references.

Figures are introduced when necessary, and an appendix on zonal harmonics is included.

In addition to his indebtedness to the various authors of the books and articles consulted, the writer of this report is greatly indebted to Dean Brown, Prof. Rubinsky and Dr. Chévrier for their various helps, suggestions, and comments.

Salem Hanna Khamis

American University of Beirut

May 4, 1942

TABLE OF CONTENTS

GENERAL INTRODUCTION

1 A Historical Background:

Two epoch making events are noted in the study of Solar physics. The first was the invention of the telescope by Galileo in 1609. This invention marked the beginning of the investigations in this field of study. Sunspots, though recorded before this invention, were discovered by Galileo in 1611, and independently in the same year, by T. Fabricius. Also at the same time Father Scheiner published his observations in his 'Rosa Ursina'.

But real progress did not begin until the second event, the application of spectrum analysis to astronomy. The solar light was first analysed by Newton in 1666, and in 1815 Fraunhofer discovered the dark lines in the solar spectrum. This, followed by Kirchhoff's researches and his well known law discovered in 1859, led to a new outlook in the methods of investigations in astrophysical studies. From this part of the 19th century up to our own times the progress was very rapid. The most prominent pioneers in this work since then may be only outlined here.

SCHWABE of DESSAU was the first to announce a period of 10 years to the frequency of sunspots, this was in 1843. In 1852, Sir Edward Sabin announced his discovery of the relation between the cycle of solar activity and the periodic variations in the terrestial magnetism.

R. WOLF was the first to make a wide statistical study of spots recorded since Galileo, and in 1850 he deduced a period of 11.11 years for the solar cycle.

CARINGTON and SPÖRER'S investigations were on the equatorial acceleration of the sun based on the motion of sunspots in different latitudes. They also studied the displacement in latitude of the spots in the eleven years cycle, and they worked on the determination of the position of the solar axis of rotation relative to the ecliptic.

In 1868 Janseen and Lockyer each independently discovered the possibility of studying the bright lines in the prominences' spectra during daylight by using a spectroscope of high dispersion. With this discovery we can say that a new period began in solar research.

The sun's chromosphere became of great interest to workers on the sun such as Huggins, Zöllner, Young, Respighi, Secchi and Tacchini. The last two founded the first astrophysical journal in 1871.

Angstrom then followed by Rowland, worked on the solar spectrum. Vogel worked on determining the solar rotation by means of the Doppler effect produced in the Fraunhofer lines.

With Hale's discovery of the spectroheliograph in 1889 began the new era in the field of solar physics. In 1891 he obtained his first photographs of prominences in the light of the H and K calcium lines. About the same time Evershed and Deslanders used the spectrohelioscope in their researches. Shortly after the

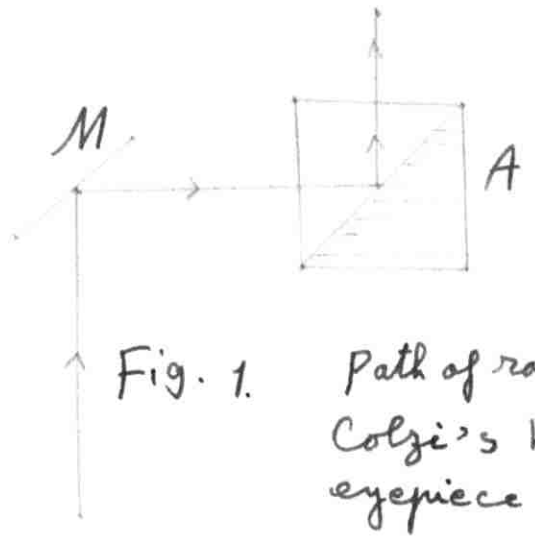

Fig. 1. Path of rays in the Colzi's helioscope eyepiece.

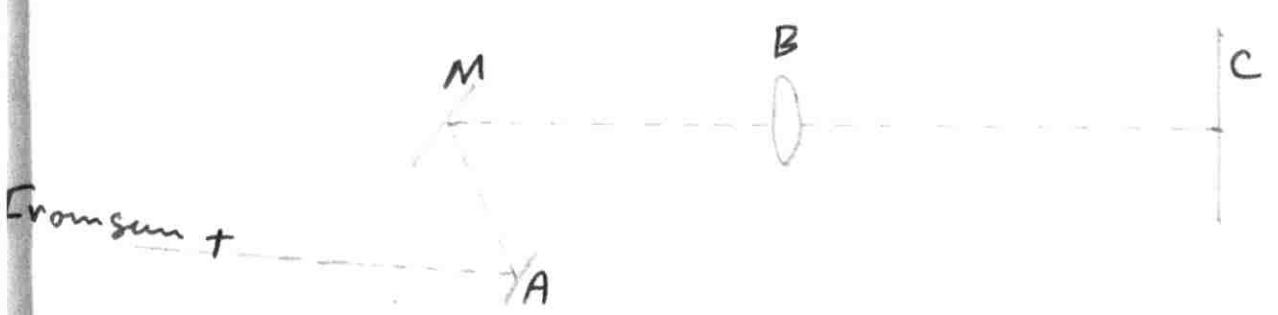

Fig. 2. System of mirrors in a horizontal Telescope.

invention of the spectroheliograph and spectrohelioscope, telescopes and spectrographs of great focal length and of high dispersion were manufactured and arranged horizontally or vertically.

These great works and discoveries and inventions led to a wide solar campaign in the first half of this 20th century. Great progress both in the experimental and in the theoretical fields is noticed. We will try now to give a brief summary of the apparatus of observation and of our present knowledge of the sun so that we may be able to go farther to our real problem the problem of sunspots.

2 The Apparatus for Observing the Sun:--

Had it not been for the nearness and thus for the high intensity of the sun's light, it could have been observed, like other stars, directly by the use of a reflector or a refractor. To escape the harmful bright image of the sun, the telescopes should be accompanied by some means for reducing the intensity of the light. For this end helioscopes are necessary.

The helioscope in its simplest form consists of a delicately shaded glass slip with its two surfaces parallel. In most cases dark glass is used so that it gives a slightly smoked image of the sun. Reflected or polarised light may also be used as in the case of Herschel and P. Cavalleri's helioscope. Colzi designed a helioscope in which the light is softened by causing it to pass through a double prism composed of a right angled crown prism and another rectangular liquid prism. The rays of light, being reflected by a mirror M, enter the crown prism A and some of it is absorbed by the liquid prism (usually contains vaseline oil) whose index of refraction differs very slightly from that of the glass. The other part, which is not absorbed, is reflected and gives an image of the sun which can be observed by the eye by means of an eye-piece (see figure 1).

The image can also be projected on a screen, as was done by Galileo and Father Scheiner. This is suitable especially when drawings of the disc are required. A photographic plate may replace the screen so that it is possible to photograph the phenomena observed; then the instrument is called a heliograph.

Sometimes, for spectrum analysis, it is necessary to attach to the telescopes spectroscopes of high dispersive power and of great focal length. It is rather difficult to attach such huge spectroscopes to the eyepiece of a telescope. For this reason fixed telescopes of vertical or horizontal axes are in general use. The image of the sun is reflected to the telescope by means of a system of mirrors and a coelostat. Then the instrument is called a horizontal or vertical helioscope. Such an instrument consists of a coelostat (A) (fig. 2), a reflecting mirror (M), and an objective lense (B) which gives an image of the sun at a screen (C). The coelostat and the second mirror reflect the solar rays in a constant direction regardless of the diurnal motion of the sun, this is done by driving the coelostat by a clockwork mechanism mounted on an axis parallel to the earths axis in such a manner as to follow the sun in its apparent motion keeping the reflected rays parallel to one suitable direction always. This is of the fixed horizontal type. The first telescope of the vertical type was first designed by Hale and was

called the "Solar Tower." At the top of the tower rests the coelostat and another mirror, while the reflector or refractor is placed vertically in a fixed position.

When the spectrum is required for analysis, the instrument is supplied with a spectroscope or a diffraction grating instead of the screen C. The diffraction grating is now in common use. The image of the sun can then be seen in the light of one spectral line. The instrument is then called a spectroheliosoope. If photographs are taken then the instrument is called a spectroheliograph. Thus spectroheliographs and spectroheliosoopes are used for photographing or observing the solar spectrum through a certain monochromatic radiation e.g. H_α hydrogen line or the K calcium line. This is done by isolating one line from the remaining spectrum by means of a slit. The image of the sun can be moved infront of the slit or the slit can be moved while the image is kept fixed and thus the whole solar disc can be observed or photographed in the light of a single spectral line.

For increasing the dispersion of such instruments and thus to obtain a more extensive spectrum, the number of prisms could be increased. By a combination of prisms of different indices of refraction, the spectrum could be observed in the direction of the incident rays without deviation just as in the case of Amici's "direct-vision prism." Instead of this combination of prisms, the diffraction grating stands as the best instrument for such a purpose.

The spectroheliograph may be fixed either to a reflector or a refractor, and when it is of large dimensions it is fixed into a horizontal or vertical telescope. The various types of spectroheliographs may be classified into two kinds, one in which the solar image and the photographic plate are kept fixed while the whole spectroheliograph is made to move, or the spectroheliograph is fixed while the image of the sun is made to move infront of the first slit, and the photographic plate is made to move infront of the second slit.

The spectroheliosoope is similar to the spectroheliograph with the difference that instead of having a photographic plate to get a photograph for the solar disc in one wave length, you look at the image of the sun directly through a small telescope.

Photographic observations have advanced a lot and are until now the most available means of studying monochromatic images. But visual observation also has its merits in being very rapid, and because details are observed better than photographed. Now we will have a bird's eye view of the general results of such observations and photographs that were done with such instruments.

3 General Description of the Sun:

(a) Distance, Mass, Dimensions, and Figure of the Sun:--

The measurement of the earth-sun distance stands as the fundamental problem from which we should start. This is seen from the fact that it is impossible to know any thing about the dimensions, mass, density, and other phenomena of the sun without knowing its distance accurately. The importance of this distance is also due to the fact that all celestial distances are measured

by means of it, and it is known by the name "astronomical unit" of distance. There are different methods for the measurement of this distance; these methods can be classified into three classes:

(1) Geometrical methods, which depend on the direct measurement of the parallax of some planet or asteroid (the asteroid Eros proved to be the most suitable for this purpose) whose distance in terms of astronomical units is known.

(2) Gravitational methods, which depend on the determination of the ratio of the mass of the earth to that of the sun from the effect of perturbations. Thus from the perturbations caused by the moon we can claculate the sum of the masses of the moon and the earth. When this mass of the earth-moon system has been found, then the solar parallax may be determined by use of gravitational laws, and from the parallax it is easy to find the distance.

(3) Methods depending upon the velocity of light which give directly the earth's orbital velocity and the radius of the earth's orbit. From this by the aberrational method the distance of the sun can be computed.

Thus we measure indirectly the parallax of the sun, and this will yield the distance of the sun. We cannot measure the solar parallax directly because this measurement is liable to great error due to the brightness of the image and its size.

The mean of the determinations of the solar parallax is $8''.803 \pm 0''.001$ which gives a mean distance of 149,450,000 ± 17,000 kilometers or 92,870,000 miles.Ψ

When we are given the distance of the sun and its apparent diameter (31' 59".3 ± 0".1), its dimensions can readily be computed. Taking the value of the solar parallax $P_\odot$ to be 8".80 and the earth's equatorial radius R to be 6377 kilometers, the distance D of the sun from the earth is given by the relation

$$D = 206265\, r_\oplus / P_\odot = 149,450,000 \text{ Km.} \quad \text{-----} \quad (1)$$

a distance travelled by light in 498 seconds. From the above mentioned value of the apparent diameter of the sun at mean distance, the radius ρ at mean distance will be 959".6. The radius $r_\odot$ of the sun expressed in kilometers is given by

$$D = \frac{206265}{\rho} r_\odot$$

so that from (1) by substitution for D we find

$$r_\odot = \frac{\rho r_\oplus}{P_\odot} = 695450 \text{ Km.}$$

a value which is 109 times the earth's radius. From this ratio the volumes is found to be 1,300,000 times the earth's volume.

From the ratio of the masses of the earth and the sun, the mean density of the sun can be compared with that of the earth and that of water. The ratio of the masses can be found as follows:

Let A be the acceleration of the central force which keeps the earth to its orbit, g the acceleration of gravity on the earth and $M_\odot$ and $M_\oplus$ the masses of the sun and the earth respectively, then

$$\frac{A}{g} = \frac{M_\odot}{D^2} : \frac{M_\oplus}{r_\oplus^2} \qquad (2)$$

Ψ A more recent determination gives 9300, 300 miles, a larger distance, Science News Letter Nov. 15, 1941.

Neglecting the eccentricity of the earth's orbit and taking the orbital velocity $V_\oplus$ of the earth to be

$$V_\oplus = \frac{2\pi D}{T}$$

where T is the number of seconds in a year, and since $A = V_\oplus^2/D$ we have, by substituting in (2) and taking g to be 981 cm/sec^2

$$\frac{M_\odot}{M_\oplus} = 332000$$

Thus the ratio of the masses is nearly four times less than the ratio of the volumes, and therefore the mean density of the sun is one fourth of that of the earth or 1.4 times that of water.

The force of gravity on the sun's surface is obtained by dividing the ratio of the two masses by the square of the ratio of the two radii, that is $\frac{332000}{(109)^2} = 27.9$ times the force of gravity at the earth's equator.

The figure of the sun is found above quantitatively may suggest that it is unchangeable. But Father Secchi and Father Rosa of the Roman College observatory discovered that the sun's diameter changes. Father Rosa then demonstrated that the diameter of the sun increases when the number of sunspots and prominecences is minimum. This Secchi-Rosa law was then confirmed by R. Wolf.

The slight difference between the equatorial and the polar diameters (the polar being the greater) makes the study of the variations of the solar diameter more complicated. This difference varies with the solar activity cycle. Lane Poor discovered that the ratio of the polar and equatorial diameters varies periodically; the period is not certain but may be equal to that of the spots. This variation may be also related to the varying height of the chromosphere and the corona. But the relation between the solar cycle and the variation of the diameter is inconclusive; yet in general we may say that the sun is a slightly pulsating star with a period of pulsation equal approximately to that of the sunspot cycle. L. Sussman (1) gave an explanation for the periodicity of sunspots which depends on this solar phenomenon. To this we will return later when we have acquired a basic knowledge of the solar phenomena and the solar composition and structure.

(b) Solar Structure, Rotation and Composition:

We have seen that there are two different ways for observing the sun: the photographic or visual and the spectroscopic observations.

The luminous image of the sun as observed by the naked eye, or as photographed, is the projection of the photosphere. The brightness of the image decreases from centre to limb because of the existence of an atmosphere which absorbs more light when the rays are inclined than when they are radial. The photosphere forms the interior of the sun. The photosphere is not white and perfectly plane, but it shows a well defined structure called granulations. These are numerous nuclei seen projected on a

---

(1) Popular Astronomy August 1940

dark background, lighting the whole photosphere. The background is not really dark, but it appears so in comparison with these nuclei. The most favourable view of granulation can be obtained around the center of the solar disc. These nuclei are usually round except near the sunspot region where they become somewhat oblong with a diameter not more than a second of arc. They are disturbed in such a way that they are separated from each other by a distance equal to that of their size; but near susnpots they are so crowded that they hide the darker background below. The photosphere is not even so uniform in its appearances but under favourable conditions for photographing the sun, the granulation appears in a number of detached groups of nuclei. The space between such groups is also filled by nuclei of various sizes. In the groups themselves the nuclei are either disturbed or very difficult to be seen, and in many cases they disappear giving place to streaks. However, there are spaces between these groups where no nuclei are visible, and instead, small dark specks of the same luminosity as that of the background are seen. Some of these specks increase in size, getting darker and darker until they become entirely black, and then they are called pores. These pores are nothing else than small sunspots. Several small pores of this kind join and form two large spots. To these spots we will return later, now paying attention to other phenomena which has some relation to these spots. Of the important features of the solar surface are the bright streaks.

These bright streaks are ramified in form and appear above the photosphere extending for thousands of kilometers over the sun's surface. These are also termed faculae; they often appear in the neighbourhood of sunspots. Though they may retain their position for several weeks, yet their shape changes in a few hours. The spectra of faculae show enhanced lines indicating higher temperatures or lower pressure (or both) than that of the surrounding photosphere.

When a spectroheliograph is used for photographing the sun, the granules are note seen, but the whole surface of the sun is shown covered with numerous dark and bright markings called by Hale, flocculi. Many of these flocculi overlie faculae directly and are similar to them in form, yet they are different from them. They are compressed or highly heated gas which absorb light of a certain ~~ion~~ wavelength only and are transparent to a large part of the solar light, thus rendering themselves invisible when the sun is directly observed by the visual or photographic methods. The calcium flocculi (those seen through the ionized H or K calcium lines) are bright over an extensive area, especially in the neighbourhood of sunspots. The hydrogen flocculi (those observed in the light of H_α or H_γ hydrogen lines) are more clearly defined than those of calcium; and the largest hydrogen flocculi are dark. These great dark hydrogen flocculi are in many cases prominences projected upon the photosphere, and may be seen as prominences when they are near the limb of the sun.

Prominences are vast eruptions of gas which rise to heights sometimes as great as 600,000 miles--two thirds of the sun's diameter--and extend over similar lengths along the solar disk with

the appearance similar to real sheets of flame. Small prominences appear to be directly on the solar atmosphere--the chromosphere--while large prominences are above the chromosphere but they are connected to it by columns like the trunks of trees. All prominences are in motion, a fact which could be detected in a few minutes; the gases in some prominences move with velocities as high as 200 miles per second. The bright line spectra of prominences include the Balmer hydrogen series, the D_3 of helium, and the H and K of ionized calcium; often in the case of bright active prominences metallic lines are found.

The spots, faculae, and flocculi supply us with a rather good method for measuring the solar rotation. Early observations show that these phenomena move from east to west in parallels of latitude on the solar disk. Thus, when a spot is formed on the side of the sun which is invisible from the earth, it appears afterwards on the eastern limb of the sun and disappears at the western limb. Certain spots of considerable dimensions may last for several rotations of the sun and therefore they can be used for measuring the period of this rotation. But it was also noticed that different spots gave different results for the period of rotation according to the latitude at which the spot is located. It was Carrington who first reached the conclusion that this difference in the results is a systematic one, and that the time of rotation of the solar equator is shorter than that for other parallels of latitude. Also gases of the lower portions of the solar atmosphere (whose motion towards or away from us may be measured spectroscopically) afford a different method for measuring the velocity of rotation. All these methods are seen to agree to a certain extent, and according to the investigations of Mr. and Mrs. Maunder the mean sidereal rotation period for spots on the equator is 24.65 days; in latitude 20°, 25.19 days, in latitude 30°, 25.95 days; and for the few spots in latitude 35° (because beyond latitude 40° N. or S. it is very rare and even impossible to see spots), 26.63 days.

This usual phenomena of equatorial acceleration was given many explanations. The latest was that it is required by the way in which heat is outpoured from the sun's interior, while another explanation considers it as a result of the dynamic encounter of two stars which gave rise to the solar system. To this equatorial acceleration we will pay more attention when we will discuss the radiation equilibrium of the sun and the outpour of energy from the interior of the sun, that is, the first explanation given here. Now we will return to the composition and structure of the sun.

We said that the visible disc of the sun is the projection of the luminous surface of the sun and is known as the photosphere. The photosphere forms the interior of the sun. The atmosphere of the sun lies above the photosphere and is composed of luminous but very transparent gases. This atmosphere can be studied telescopically only during a total eclipse, but most of its phenomena can nowadays be studied spectroscopically nearly at every time of the day. The solar atmosphere may be considered to be of two layers: the reversing layer (responsible for the Frannhofer lines), extending to a height of about 600 miles above

the photosphere and is composed of vapours of many identified terrestial elements, and then the Chromosphere which forms the higher part of the atmosphere and in which the prominences of various kinds have their origin. The Chromosphere extends to a height of several thousand miles, and is composed of light gases such as hydrogen and helium, yet calcium is also found there. By means of the flash spectrum during a total eclipse of the sun, the different layers at which the spectral lines are emitted can be determined. Mitchell of Virginia found that ionized calcium which produces the H and K lines extends to a height of 14,000 kilometers, hydrogen produces the H at 10,000 and other Balmer lines at 8,000 kilometers, helium extends about 7500 kilometers while neutral calcium which produces the spectral line λ 4227 Angstroms extends about 5000 kilometers. Other lines, chiefly of metallic origin are at lower heights, Fraunhofer lines are originated at a height less than 500 kilometers. that is in the reversing layer, yet there is no sharp edge between the two layers.

The outer envelope of the sun above the chromosphere is termed the corona. Its height is tremendous, but its density is very small, and it is very difficult to be observed except during an eclipse of the sun. It extends at least 500,000 miles, and some of its streamers have been observed to a height of 5,000,000 miles. The spectrum of the corona consists of: a continuous one due probably to incandescent fine liquid or solid corpuscles, a dark line spectrum of the sun but very faint, and thirdly the bright line spectrum due to luminous gases within the corona.

Now having acquired a general and brief description of the sun let us go a little deeper and discuss some of the most important theoretical and practical developments in the field of solar physics.

PART I

INTRODUCTORY THEORETICAL PROBLEMS

IN

THE PHYSICS OF THE

SOLAR ATMOSPHERE

CHAPTER I

Ionization in the Solar Atmosphere:

Since Kirchhoff's interpretation of the Fraunhofer lines in the spectrum of the sun, astrophysicists began to identify the elements found in the solar atmosphere. Of the 20,000 lines catalogued by Rowland only 6000 lines were identified with known lines of terrestial sources. By this method thirty six elements were proved to exist in the Sun, while many other terrestial elements (such as: Rubidium, Nitrogen, Phosphorous, Boron, Antimony, Bismuth, Sulpher, Thallium, and Praseodymium) were not identified. Other elements gave weak evidences for their existence in the sun, while on the other hand, other elements(such as calcium, iron and others) are very prominent.

To assume that these phenomena are due to the chemical composition of the sun will yeild to the unsatisfactory conclusion that the elements of which no lines are found in the flash or Fraunhofer spectrum are absent from the sun, and that the other lines which were unidentified are helio-elements which do not exist on the earth. There is no a priori reason why the sun should have some elements to the exclusion of the other elements, or why the sun should have elements different from those known on our planet. Another explanation was given according to which certain elements are confined to the photosphere because of their high atomic weight; but this view is unsatisfactory because it solves only one part of the question and also certain light elements do not show themselves.

Moreover another difficulty was at hand. It was observed by Lockeyer that certain linws are relatively strengthened in the chromospheric spectrum, and which were called by the same person "enhanced lines". This phenomenon is similar to what was observed in the spark spectra as compared with those of the arc. Lockeyer's explanation was, therefore, that the reason for this in the chromospheric spectrum is also due to a localized increase in temperature. But this explanation should necessarily lead to the untenable conclusion that the temperature of the sun increases radially outward.

It was Saha(1) who made it clear that "the varying records of different elements in the Fraunhofer spectrum may be regarded as arising from the varying response of these elements with regard to the stimulus existing in the sun. The stimulus existing in the sun is the same for all elements,-----, but owing to different internal structure, elements will respond in a varying degree to this stimulus". (2) Saha also claimed that the strengthened lines and most of the unidentified lines are not due to normal but to

---

(1) Philosophical Magazine volume 40 pages 472 and 809 et seq., 1920 and " " " 41 " 267, 1921

(2) ibid volume 40 page 811

ionized atoms. This leads to the conclusion that the chromosphere is a center of intense ionization. This attitude is also confirmed by the theoretical development of the theory of spectroscopy. For example the spectroscopic constant K in the equation

$$\nu = K\left(\frac{1}{m^2} - \frac{1}{n^2}\right) \quad \text{where } n = m+1, m+2, \cdots$$

for an ionized atom is four times its value for the normal atom, and the wavelength found by this way agrees with the experimental value. This was developed by Saha into his famous theory of ionization. We will now try to derive his famous equation.

Consider a reaction according to the formula

$$\upsilon_a A + \upsilon_b B + \cdots = \upsilon_m M + \upsilon_n N + \cdots$$

where A,B,--,M N,--,are the elements or compounds or ions and the υ's are their relative abundance, then K, the reaction-isober or equilibrium constant (at constant pressure) is given by

$$K = \frac{P_M^{\upsilon_m} P_N^{\upsilon_n} \cdots}{P_A^{\upsilon_a} P_B^{\upsilon_b} \cdots} \qquad (1)$$

where P_M, P_N, --- are the partial pressures of the reacting substances M, N, ---- then

$$\log K = \log \frac{P_M^{\upsilon_m} P_N^{\upsilon_n} \cdots}{P_A^{\upsilon_a} P_B^{\upsilon_b} \cdots} \qquad (1)a$$

or (1) $$\log K = -\frac{U}{4.571T} + \frac{\Sigma \upsilon C_p}{R} \log T + \Sigma \upsilon C \qquad (1)b$$

where the logarithms are to the base 10 and

T = absolute temperature
U = heat of dissociation or ionization
C_p = specific heat at constant pressure
C = Nernst chemical constant

For a reaction of the form

$$Ca \rightleftharpoons Ca^+ + e - U \qquad (2)$$

we have

$$\Sigma \upsilon C_p = (C_p)_{Ca^+} + (C_p)_e - (C_p)_{Ca}$$

but $(C_p)_{Ca^+}$ is approximately equal to $(C_p)_{Ca}$ and $(C_p)_e$ is equal to $\frac{5}{2}R$ (where R is the gas constant) considering the electron as a monatomic gas, therefore we have

$$\Sigma \upsilon C_p = 2.5\,R$$

The Nernst constant according to Eggert's calculation is given by the formula

$$C = -1.6 + \frac{3}{2}\log M \qquad (3)$$

---

(1) Nernst (the new Heat Theorem) pp. 125 and 133

where M is the molecular Wt. of the element , and the pressure is given in atmospheres. For calcium C is the same for Ca^+ and Ca and therefore

$$\Sigma \upsilon C = C_e$$

for an electron the molecular weight M can be taken as 5.5×10^{-5} and therefore C = -6.5 and

$$\Sigma \upsilon C = -6.5 \qquad (4)$$

substituting for $\Sigma \upsilon c$ and $\Sigma \upsilon c_p$ in (1)b we get

$$\log K = \frac{-U}{4.571T} + \frac{5}{2}\log T - 6.5$$

Now if x is the fraction of the calcium atoms that is ionized and if P is the total pressure, then the partial pressure of the free electrons is $\frac{x}{1+x}P$, and this multiplied by the ratio of the ionized to the neutral atoms, i.e., $\frac{x}{1-x}$ gives the right hand side of equation (1)a and therefore we get

$$\frac{P_e \times P_{Ca^+}}{P_{Ca}} = \frac{x}{1+x}P \cdot \frac{x}{1-x} = \frac{x^2}{1-x^2}P = K$$

or on substitution in the last equation we get

$$\log K = \log \frac{x^2}{1-x^2}P = \frac{-U}{4.571T} + 2.5\log T - 6.5 \qquad (5)$$

which is the famous equation derived by Saha for the "reaction-isobar". U, the heat of dissociation can be calculated easily from the i nization potential V. For an ionization potential of one volt, U is equal to 8.502×10^4 calories. Putting equation (5) in terms of the ionization potential V we get

$$\log K = \log \frac{Px^2}{1-x^2} = -\frac{5036}{T}V + 2.5\log T - 6.5 \qquad (6)$$

since the ionization potential for a given gas is a constant, then the "reaction-isobar," as in formula (6), is a function of the absolute temperature only. Hence, a gas with a smaller ionization potential is more ionized than that of an element of a higher ionization potential, on condition that the pressure is constant.

Saha calculated the degree of ionization of some elements whose ionization potential is known, by using equation (6). For calcium he got the following results taken from his original work (table IV). Here he expresses percentage ionization, temperature in the absolute scale, and pressure in atmospheres.

Table I. Percentage Ionization of Calcium V = 6-12 volts

Pressure / Temperature	10	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-6}
4000°	—	—	—	2.8	9	26	93
5000°	—	2	6	20	55	90	100
6000°	2	8	26	64	93	99	
7000°	7	23	68	91	99	100	
8000°	16	46	84	98.5	100		
9000°	29	70	95	100			
10000°	46	85	98.5				
11000°	63	95	100				
12000°	76	96.5					
13000°	84	98.5					
14000°	90	100					

Region of Complete Ionization

Calcium has a relatively low ionization potential, and it seems interesting to compare the results given in table I for calcium with those in table II [1] for atomic hydrogen (V = 13.5 volts).

Table II

Pressure / Temperature	1	10	$10^{}$	$10^{}$	$10^{}$	$10^{}$	$10^{}$
7000	—	—	—	1	4	12	
8000	—	—	2	5	18	50	
9000	—	2	6	20	63	90	
10000	2	6	17	49	87	99	
12000	9	28	68	94	100	100	
14000	27	65	93	100			
16000	55	90	100				

Complete Ionization

(1) Hand Buch Der Astrophysik p. 205. Band IV

Thus higher temperatures are required for producing the same ionization in hydrogen as in calcium. Caesium, which has the lowest ionization potential is completely ionized at about 4000° with a pressure of 10^{-4} atmospheres ; helium, having the highest ionization potential known, cannot be ionized completely at a temperature less than 20,000° for a pressure of 10^{-4} atmospheres.

Taking the temperature of the chromosphere and the reversing layer to be between 6000° and 7000° and since the pressure is relatively low, then we can apply Saha's equation to know the degree of ionization of different elements. The theory shows that molecular hydrogen in the sunshould be dissociated into atoms, but since atomic hydrogen has a high ionization potential, its ionization cannot be complete. Also helium leaves traces of enhanced lines because it cannot be appreciably ionized in the solar atmosphere. A summary of Saha's results may be inserted here.

(1) The higher part of the chromosphere is the seat of ionized atoms chiefly of calcium, Barium, Stronium, Scandium, Titanium, and iron. In the lower part both neutral and ionized atoms exist.
(2) the pressure has a great influence on the degree of ionization. The almost complete ionization of Ca, Sr, and Ba in the upper chromospheric layers is due to the low pressure there.
(3) Hydrogen should be completely dissociated into atomic hydrogen in the solar atmosphere.
(4) The greater the ionization potential, the less is the degree of ionization.

But Saha's equation is a special case, and it is true for the ionization of a gas consisting of atoms of the same element. For mixtures of gaseous elements, Russell(1) obtained by extending Saha's equation a more general case. Let us now proceed with Russell's extension of the theory.

If P' is the partial pressure of the ionized atoms of any sort, P that of the corresponding nonionized atoms, and P" that of the free electrons we have as before

$$\frac{P'P''}{P} = K \qquad (1)_C$$

Now, if the relative numbers of the different atoms be $a_1, a_2, a_3, \cdots$, and if $x_1, x_2, x_3, \cdots$ the fractions ionized, then the sum of the number of atoms present which is ionized and that which is unionized is

$$a_1(1+x_1) + a_2(1+x_2) + \cdots$$

and the partial pressures will be

$$P_1' = \frac{a_1 x_1}{\Sigma a + \Sigma a x} P, \quad P_1 = \frac{a_1(1-x_1)}{\Sigma a + \Sigma a x} P, \quad P'' = \frac{\Sigma a x}{\Sigma a + \Sigma a x} P = \frac{\bar{x}}{1+\bar{x}} P \qquad (7)$$

where $\bar{x} = \frac{\Sigma a x}{\Sigma a}$ = the fraction of all the atoms present which is ionized. Equation (a) then becomes

$$\frac{x_1}{1-x_1}\frac{\bar{x}}{1+\bar{x}} = \frac{K_1}{P}, \quad \frac{x_2}{1-x_2}\frac{\bar{x}}{1+\bar{x}} P = K_2 \qquad (8)$$

(1) Astrophysical Journal vol. 55 p. 119 1922

where K_1, K_2 are given by (6) . We, therefore, obtain from (8)

$$\frac{x_1}{1-x_1} = \frac{K_1}{K_2}\,\frac{x_2}{1-x_2} \qquad (9)$$

but from (5) we have

$$\log K_1 = \frac{-U_1}{4.571T} + 2.5 \log T - 6.5 \qquad (5')$$

and

$$\log K_2 = \frac{-U_2}{4.571T} + 2.5 \log T - 6.5 \quad \text{(If T is the same)} \qquad (5'')$$

subtracting (5'') from (5') we get

$$\log \frac{K_1}{K_2} = \frac{U_2 - U_1}{4.571T} = 5036\,\frac{V_2 - V_1}{T} \qquad (10)$$

From (9) and (10) we find that "the ratio of the number of ionized atoms to that of nonionized atoms for any two elements in a gaseous mixture bear a fixed proportion to one another, which depends only on the temperature, and is independent of the pressure, the relative abundance of the two elements, or the presence of the other elements." (1)

The latter conditions affect the amount of ionization of each element in such a way that the above mentioned ratio is not affected by them.

Equation (8) may be written in the form

$$\left.\begin{aligned} \frac{x_1^2}{1-x_1^2} &= \frac{K_1}{P_1} \\ \text{where } P_1 &= P\,\frac{\bar{x}}{x_1}\,\frac{1+x_1}{1+\bar{x}} \end{aligned}\right\} \qquad (11)$$

Now from (11) when $x_1 > \bar{x}$, (i.e. when the ionized fraction of the element is more than the average), then P_1 will be less than P; and if this element were present alone then the percentage of ionization will be greater for the same total pressure. But for elements which are ionized with greater difficulty than the average, that is when $x_1 < \bar{x}$, then for that element P_1 will be more than P, and the percentage ionization will be less than it would be if that element were present alone with the same total pressure. But, in all cases, the ionization will be less than if this element were present alone with its actual partial pressure, because P" will always be greater than P_1 in equation (9)c

It might be interesting to see the results of this theory in some special cases :

(1) At the beginning of ionization all the x's are small, and the previous equations may be approximated to give:

$$x_2 = \frac{K_2}{K_1}\,x_1 \text{ , etc , } \bar{x} = \frac{x_1 \Sigma a K}{K_1 \Sigma a} .$$

$$P_1 = P\,\frac{\bar{x}}{x_1} = P\,\frac{a_1}{\Sigma a}\left(1 + \frac{K_2 a_2}{K_1 a_1} + \frac{K_3 a_3}{K_1 a_1} + \cdots\right) \quad -?$$

(1) ibid page 122

The quantity outside the parenthesis is the partial pressure of the element 1. From these equations it appears that at the beginning of ionisation the element of lowest ionization potential behaves (approximately) as if it were alone present at its actual partial pressure. For elements of higher ionization potentials the ionization at the beginning is less and in most cases is much smaller.

(2) When the ionization is nearly complete, then $x_1^{(1)} \sim 1$

and we have $$P_1 = P \frac{2\bar{x}}{1+\bar{x}}$$

Thus, when the proportion of atoms of elements of higher ionization potentials is small, then the effective pressure for this element is nearly equal to the total pressure. Therefore, for an element of low ionization potential the range of temperature or pressure within which both ionized and uncharged atoms will be present in sensible proportions is extended by the presence of other constituents in the gaseous mixture.

(3) For an element of difficult ionization x_1 is much less than $\bar{x}$ and P_1 is greater than P at the beginning of ionization. The pressure or temperature range within which both ionized and nonionized atoms are present is reduced in this case. If we put

$$\frac{\bar{K}}{P} = \frac{\bar{x}^2}{1-\bar{x}^2}$$

we find that

$$\bar{K}\frac{1-\bar{x}}{\bar{x}} = K_1\frac{1-x_1}{x_1} = K_2\frac{1-x_2}{x_2} = \cdots = \frac{\Sigma aK(1-x)}{\Sigma ax}$$

and

$$\bar{K} = \frac{\Sigma aK(1-x)}{\Sigma a(1-x)}$$

So $\bar{K}$ is the mean of the individual K's with a weight equal to the ratio of the nonionized atoms remaining. At low temperatures atoms which have the largest K's and of easy ionization contribute most to the mean, $\bar{K}$. At high temperatures, atoms of high ionization potential will contribute to this mean most. Thus the effective mean ionization potential, corresponding to $\bar{K}$, will increase with the temperature.

Second Stage Ionization: Saha dealt with this subject satisfactorily, but Russel extended Saha's work to the case of more than one element and added a correction to Saha's results. Here we will continue Russels work. Consider a gas of the same atoms which are liable to loosing two electrons successively. Let V be the ionization potential for first stage ionization, and V' for the second stage. Let also x be the ratio of the atoms which are singly ionized, and y of those that are doubly ionized; then the

---

(1)($\sim$) means "is not very different from"

ratio of the free electrons will be $x+2y$ and we have in this case for the two stages:

$$\left.\begin{aligned} \frac{x(x+2y)}{(1-x-y)(1+x+2y)} &= \frac{K}{P} \\ \text{and}\quad \frac{y(x+2y)}{x(1+x+2y)} &= \frac{K'}{P} \end{aligned}\right\} \qquad (12)$$

The first of these equations may be written

$$\frac{(x+y)^2}{1-(x+y)^2} = \frac{K}{P}\,\frac{(x+y)^2(1+x+2y)}{(x^2+2xy)(1+x+y)} > \frac{K}{P}$$

or

$$\frac{x^2}{1-x^2} = \frac{K}{P}\left(1-\frac{y(2-x+x^2+2xy)}{(1-x^2)(x+2y)}\right) < \frac{K}{P}$$

and the second equation of (12) can be written in the form

$$\frac{y^2}{1-y^2} = \frac{K'}{P}\,\frac{x}{1-y}\,\frac{2y^2+xy+y}{2y^2+xy+2y+x} < \frac{K'}{P}$$

From these equations it is easy to see that the numbers of both neutral and singly ionized atoms are less than they would be if the second ionization did not take place, but the number of doubly ionized atoms is less than it would be if the second ionization was the only one that occurred.

Dividing the second equation by the first one of equation (12) we get

$$\frac{y(1-x-y)}{x^2} = \frac{K'}{K} \qquad (13)$$

and therefore

$$\log\frac{K'}{K} = -5036\,\frac{V'-V}{T} = \log\frac{y(1-x-y)}{x^2} \qquad (14)$$

Since V'-V in all cases, so far known, (1) is positive and is equal to at least five volts, therefore K'/K is a small quantity, and it is impossible for both y and (1-x-y) to be considerable simultaneously. When the temperature rises K'/K increases, but then x becomes small and the number of neutral atoms very small.

The above equations are true in the case of a gas composed of the same atoms. In a mixed gas we will have

$$\frac{x\bar{x}}{(1-x-y)(1+\bar{x})} = \frac{K}{P}$$

$$\frac{y\bar{x}}{x(1+\bar{x})} = \frac{K'}{P}$$

and $$\bar{x} = \frac{\Sigma a(x+2y)}{\Sigma a}$$

(1) Astrophysical Journal 55 124 1922

This presence of different elements will decrease the amount of second stage ionization because the ionization potential V' is always higher than V, but equation (13) will hold in all cases.

Equation (10) which was derived by Russell for single ionized atoms will become in this case

$$\frac{x}{1-x-y} = \frac{K}{K_2}\frac{x_2^2}{1-x_2}, \quad \frac{y}{x} = \frac{K'}{K_2}\frac{x_2}{1-x_2} \qquad (15)$$

where x_2 is the percentage ionization for any singly ionized element.

This is a summary of H. N. Russell's theory of ionization which is an extension of Saha's theory. E. A. Milne (1) published before Russell (but the latter developed his theory independently) a beautiful summary of this ionization theory; in it he points out the influence of the free electrons resulting from the ionization of atoms of low ionization potential in diminishing the ionization of other atoms.

Evidences in favour of the theory were given by Russell when he compared the sun-spot and the solar spectrum with reference to the relative intensity in the hotter and the cooler spectrum of lines associated with ionized and nonionized atoms. The results, in general, were found to be in agreement with observed data, yet certain discrepancies (such as the presence of the large fraction of ionized Barium atoms) suggest that the theory is not complete and that some modification is necessary because the theory neglects the effects of radiation on ionization.

The lines of the alkali metals (all of which are due to the neutral atoms) are greatly strengthened in the spot spectrum. The same is true in the case of the lines of alkaline earths which are due to neutral atoms. Calcium lines (for example) are strengthened in the sunspot spectrum; but Barium "neutral" lines are absent from both spot and solar spectra. But lines of ionized atoms of calcium, Barium, and Stronium are very strong in both spectra. These results are all in agreement with Saha's theory because sunspots are at a lower temperature than their surroundings. The only two elements which are much stronger ionized that what the theory indicates are Barium and Lithium. Russell (2) extended his theoretical work on the theory of ionization by taking into consideration the effect of radiation. Russell says that the absorption of photospheric radiation by "the atoms of the solar atmosphere tends to increase the degree of ionization, both directly, by shifting an electron into a position from which its removal is easier, and indirectly, when enhanced lines are absorbed, by getting the ionized atoms into states in which they are probably less likely to combine with electrons. Both these influences operate strongly in the case of Barium and weakly, if at all, for sodium." (3) This might explain the fact

---

(1) "Observatory" September 1921 volume 44 page 261
(2) Astrophysical Journal 55 354 1922
(3) A. P. J 55 354 1922

that Barium is almost completely ionized in the solar atmosphere while sodium is much less ionized than Barium, although both elements have the same ionization potentials. But this does not explain the behaviour of lithium.

The behaviour of the following elements in the spectrum of the photosphere and of sunspots, compared with the furnace, arc, and the spark spectra, makes it possible to arrange them in the order of their increasing atomic numbers as follows:

Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, and Zn.

which is the same as ~~the same as~~ their arrangement according to increasing ionization potential from 6 volts for Ca, to 9.4 volts for Zn[1] This suggests that the ionization potential is a periodic function of the atomic number.

Saha's equation also suggests that on account of higher temperature in the faculae, the percentage ionization increases and therefore enhanced lines are strengthened in the faculae spectra. This prediction of the theory was confirmed by the work of St. John.[2]

Now we may reflect on what was mentioned in the introduction about the structure of the sun in the light of this ionization theory. We see now that the chromosphere consists of gases supported by radiation pressure acting on the individual atoms. The pressure and density increase with a slow rate inwards where the effect of gravitation becomes important. Pressure at the bottom of the chromosphere is about 10^{-7} atmospheres; below this level gravity begins to overcome radiation pressure, and pressure increases rapidly with increasing depth while temperature remains approximately constant because the gases are still transparent-- This lower region is known as the reversing layer. There is no sharp line dividing the two layers, but we can say in general that the chromosphere is the layer where first stage ionization is complete, or at least where metallic vapours are completely ionized.

When the pressure reaches 10^{-2} atmospheres the transparency of the gases becomes less and the opacity increases, and the increase in pressure is so rapid that the reversing layer merges very rapidly into the photosphere. The photosphere is now seen to be that opaque mass which constitutes the observed solar disc. With the increase of opacity the temperature rises in accordance with the theory of radiative equilibrium. The theory of radiative equilibrium will be discussed briefly in the following pages. But one thing remains to be said about the consequences of the ionization theory. It is an immediate and necessary result of the previous discussion that the sun should be surrounded by a large layer of free electrons. The existence of the corona surrounding the chromosphere with the streams of electrons ejected from it is a proof of the existence of this atmosphere of free electrons. The variation of this layer or the direct emission of electrons from it may influence terrestial phenomena, a fact with which we will deal later.[3] Before leaving this topic let us summarize the results in a general statement.

***

(1) A. P. J 61 225 1925
(2) Pop. Astr. vol. 30 p. 228 1922
(3) See preface.

The fundamental principle of this theory is that since arc spectral lines can be produced only by the neutral atom, and enhanced lines only by ionized atoms, then the relative intensities of the two kinds of lines must give some indication of the relative abundance of neutral and ionized atoms which emit the analyzed light. Also since the relative numbers of these atoms can be calculated from the theory for given conditions of pressure and temperature then some evidence may be obtained as to the actual temperatures and pressures in the reversing layers of the stars. The application of the ionization theory to the sun was done until now by comparison between the spectra of the solar chromosphere and the reversing layer, and between the spectra of the reversing layer and of sunspots. It was shown that the differences in intensity of lines arise either from differences of temperature (sunspots and their surroundings) or differences of pressure (between chromosphere and reversing layer). But this is not the only method of applying the theory. Another more general application was developed by Milne and Fowler (1) In this method attention is devoted to the change of intensity of a given line or group of lines through the sequence of stellar spectral types, a method into which we will not enter in this small paper. Further applications and modifications of the theory were made by Milne and others, but they are beyond the scope of this paper and it is sufficient here to give the following table of references for such works:

I Monthly Notices of the Royal Astronomical Society volumes 84, 85, 86, 86, 88, pp. 354, 111, 8, 578, 188 years 1924, 1929, 1925, 1926, 1928, tespectively. (by Milne)

II Publications of the Lick observatory vol. 17, 1931 "A Study of the Solar Chromosphere" by Menzel

III Monthly Notices of R. A. S. 89, 483, 1929 by McCrea
" " " " " " 94, pp. 14 and 726 (1933--34) by Chandrasekhar

IV Astrophysical Journal 69, p. 209, 1929, Unsold

But the Subject is still under discussion and researches are still going on in this field.

---

(1) Monthly Notices of the Royal Astronomical Society volume 83 p. 403, 1923

CHAPTER II

RADIATIVE EQUILIBRIUM

(1) **Introduction:**

The problem of radiation equilibrium will be considered first in the case of a stationary star, and then it will be extended to the case of a rotating star. But before we begin, some of the properties of radiation have to be considered because of their importance in the development of the subject in question.

According to the modern theories on radiation, it is composed of photons each of which is associated with an amount of energy proportional to the frequency. This energy, E, is equivalent to a mass of E/C^2, where C is the velocity of light, as given by the famous equation

$$\Delta E = \Delta m \cdot c^2 \quad \text{or} \quad E = mc^2$$

Since a photon moves with a velocity C, then it is associated with a momentum mc or, according to the above mentioned formula, E/C. Whenever this photon is absorbed its momentum is also transferred to the absorbing medium; this gives rise to what is known as radiation pressure. Supposing that photons of energy E ergs per cubic centimeter impinge normally on a perfectly absorbing surface, the energy absorbed by the surface per second will be EC and the momentum will be E dynes per square cm. If the absorbing surface is an imperfect absorber, the momentum of the photons which are scattered, or reflected, or transmitted should be accounted for.

For a perfect reflector the pressure is 2E because the direction of motion of the photons is reversed. If the surface is semi-transparent and transmits photons of energy E' per cubic centimeter, then the force exerted on the surface is E - E'. Thus the pressure of normally incident rays of radiation is equal to the energy-density. But this is not true in the case of inclined rays. Consider again a column of radiation travelling in a certain direction, and let it impinge on a screen of area A with an angle of incidence equal to θ. The screen is obstructing a cross-section of the beam equal to $A\cos\theta$, and the force on the screen will be now $EA\cos\theta$ instead of EA for normal incidence. Resolving this force into its two components normal and tangent to the absorbing surface we get

$$F_{\perp} = EA\cos^2\theta$$

and

$$F_{\parallel} = EA\cos\theta\sin\theta$$

Supposing that the radiation is isotropic, then the resulting force would be

$$F_{\perp} = EA \times (\text{average of } \cos^2\theta)$$

and $$F_{\parallel} = E A \times (\text{average of } \cos\theta \sin\theta)$$

Where the averaging is taken over a sphere. But since the average of $\cos^2\theta$ is 1/3 and of $\sin\theta \cos\theta$ is zero, over a sphere, it follows that

$$F_{\parallel} = 0$$

and $$F_{\perp} = \frac{1}{3} E A$$

or $$P = \frac{1}{3} E \quad \text{where } P = \text{pressure} = \frac{F_{\perp}}{A}.$$

Or, in other words, the pressure of radiation is one third of its energy-density. This pressure is exerted normally to any absorbing surface and is completely analogous to the hydrostatic pressure of a fluid. If the radiation is not isotropic, then instead of the average, we have to consider the weighted mean of $\cos^2\theta$

In terms of the temperature using Stefan's law

$$E = aT^4 \qquad (16)$$

where a is the Stefan-Boltzmann constant and T is the absolute temperature, we can put our previous result in the form

$$P = \frac{1}{3} E = \frac{1}{3} aT^4 \qquad (17)$$

Now, since a star is radiating energy into space continually, its surface cannot provide this energy for a long time unless it receives energy from inside. Therefore, we have to study the way in which heat is supplied to the radiating surface. There are three ways for the transfer of heat through a certain medium. The first two which are found in a medium in static equilibrium are radiation and conduction. In both ways the net flow is in the direction of increasing temperature. The conduction of heat in a ster is so small that it can be neglected in comparison with radiation transfer.

The third mode of transfere of heat is convection. This is possible in a star which is not in static equilibrium, and we will study it in more detail afterwards. Now we will consider only radiation transfer and we will neglect all other modes.

(2) <u>Radiative Equilibrium in a Stationary Star</u>:

As mentioned before, for simplifying the theoretical work it will be assumed that the only mode of transfer of heat is by radiation, and that radiation is isotropic and therefore has an energy density given by (16) and a hydrostatic pressure given by (17) The radiation will possess a momentum of E/C units in the direction of propagation.

Take the x-axis in the direction of the temperature gradient and consider a slab of stellar material perpendicular to the ox-axis having one square cm area and thickness dx. Let the temperature of the two faces be T and T+dT on the absolute scale. The radiation pressure acting normally on the two faces will give forces $+P_R$ and $-(P_R + dP_R)$, so that the resultant force in the direction of ox is $-dP_R$. This resultant gives the amount of momentum which is being acquired per second by the region occupied by the slab owing to the flow of radiation. Since there is equi-

librium, this momentum passes into the matter by the process of absorption.

Let K be the mass absorption coefficient, that is, a thin screen of material of mass W gms per sq. cm. absorbs the fraction KW of the radiation passing through it normally. If ρ is the density of matter in that region then $W = \rho dx$. If H ergs per second pass through a square centimeter of the slab at an angle of incidence θ then the distance travelled through is dx sec θ and the absorbed energy is $KH\rho \sec\theta\, dx$. Therefore the momentum gained in the direction of the x-axis is $[KH\rho \sec\theta\, dx \cdot \cos\theta]/C$ or $\frac{KH\rho\, dx}{C}$ This is true for the angle of incidence in the interval 0° to 90°, but for an angle greater than 90° the sign has to be changed because the radiation there is coming from the other side. This momentum should be equal to $-d P_R$ and therefore

$$-dP_R = \frac{HK\rho\, dx}{C}$$

or

$$H = -\frac{C}{K\rho}\frac{dP_R}{dx} \qquad (18)$$

But since $P_R = \frac{1}{3}aT^4 = \frac{1}{3}E$, then

$$\frac{dP_R}{dx} = -\frac{1}{3}a\frac{dT^4}{dx}$$

and therefore

$$H = -\frac{ac}{3K\rho}\frac{dT^4}{dx} \qquad (19)$$

Equation 18 shows that the net flow of radiation is directly proportional to the internal pressure gradient and inversely proportional to a factor $K\rho$ measuring the obstructive power of the material screen through which it is being forced to travel. This equation is analogous to that governing the flow of material fluid through a channel and it breaks down under the same circumstances as the corresponding hydrodynamical equation, namely, when the flow is so rapid that the pressure gradient can no longer be calculated hydrostatically. This is the case near the surface of the star. Therefore this equation should not be applied for the transparent part of the solar globe, but at a little depth below the surface of the photosphere the approximation given by this equation is very good. Yet in the case of even the outer layers of the sun's chromosphere this equation unexpectedly is found to give a rather good approximation. (1)

Let us now seek another proof of the equation, which enables us to discuss certain points of detail. (2)

---

(1) Eddington "The internal constitution of the Stars" page 101-footnote.
(2) ibid p. 107

Figure 3.

Consider an elementary solid angle, $d\omega$ through which passes isotropic radiation of energy-density E. The density of that part of the radiation which passes through this solid angle is

$$E \frac{d\omega}{4\pi}$$

But since the flow is not perfectly isotropic, and depends on the angle θ between the direction of the radius and the angle $d\omega$ then the energy-density E is a function of θ ; let us denote it by $E(\theta)$. The energy density of radiation whithin this solid angle is therefore:

$$E(\theta) \frac{d\omega}{4\pi}$$

Now, consider a small cylinder of length ds and cross-section $d\S$ with its length in the direction θ ,that is along $d\omega$. The amount entering the cylinder per second through the base is

$$E(\theta) \frac{d\omega}{4\pi} \cdot c\, d\S \qquad (20)$$

where c is the velocity of light.
The amount leaving at the top is

$$\left[E(\theta) + \frac{d}{ds} E(\theta) \cdot ds\right] \frac{d\omega}{4\pi} \cdot c\, d\S \qquad (21)$$

disconsidering infinitesimals of the second or higher order.
The amount absorbed in the cylinder is

$$E(\theta) \frac{d\omega}{4\pi} \cdot c\, d\S \cdot K\rho\, ds \qquad (22)$$

Also a certain amount of radiation is emitted by the material in the cylinder. This amount will be emitted isotropically and therefore the amount within $d\omega$ is the fraction $\frac{d\omega}{4\pi}$ of the whole. If j is the radiation emitted per gram per second, then the amount emitted from the mass $\rho ds\, d\S$ of the cylinder within the solid angle $d\omega$ is

$$j \frac{d\omega}{4\pi} \cdot \rho\, ds\, d\S \qquad (23)$$

Balancing the gains (20) and (22) and the losses (21) and (23), supposing that the cylinder is in a steady state, we get

$$\frac{d}{ds} E(\theta) = \frac{j\rho}{c} - E(\theta) \cdot K\rho \qquad (24)$$

When $E(\theta)$ is a function of r and θ only, as in the case of a star we have then, as seen from fig. (3)

$$\frac{dr}{ds} = \cos\theta$$

and

$$-r \frac{d\theta}{ds} = \sin\theta$$

or

$$\frac{d\theta}{ds} = \frac{-\sin\theta}{r}$$

but since

$$\frac{d(\,)}{ds} = \frac{d(\,)}{dr}\frac{dr}{ds} + \frac{d(\,)}{d\theta}\frac{d\theta}{ds}$$

therefore on substitution from above we get

$$\frac{d(\,)}{ds} = \frac{d(\,)}{dr}\cdot\cos\theta - \frac{d(\,)}{d\theta}\cdot\frac{\sin\theta}{r}$$

or

$$\frac{d}{ds} = \cos\theta\,\frac{d}{dr} - \frac{\sin\theta}{r}\,\frac{d}{d\theta} \qquad (25)$$

applying the right hand side of 25 to the left hand side of (24) instead of $\frac{d}{ds}$, we get

$$\cos\theta\,\frac{dE(\theta)}{dr} - \frac{\sin\theta}{r}\,\frac{dE(\theta)}{d\theta} = j\rho/c - E(\theta)\cdot K\rho \qquad (26)$$

Equation (26) is the general one existing between $E(\theta)$ and r. Special cases of this equation are important. For example, let us consider first the case near the surface of the star. Here r is very large and the curvature can be neglected. Equation (26) reduces therefore into

$$\cos\theta\,\frac{d}{dr}\,E(\theta) = j\rho/c - K\rho\,E(\theta) \qquad (27)$$

Now, if E is the total energy-density of radiation then it is given by

$$E = \frac{1}{4\pi}\int E(\theta)\,d\omega \qquad (28)$$

because $\frac{d\omega}{4\pi}$ is that fraction of the sphere which we were just considering. If also H stands for the net outward flow per second across unit surface perpendicular to r, then (according to the introduction of this chapter)

$$\frac{H}{c} = \frac{1}{4\pi}\int E(\theta)\cos\theta\,d\omega \qquad (29)$$

Also, if P_R' is the actual pressure of radiation in the radial direction then (also according to the introductory discussion of this chapter)

$$P_R' = \frac{1}{4\pi}\int E(\theta)\cos^2\theta\,d\omega \qquad (30)$$

Now, multiplying (27) by $\frac{d\omega}{4\pi}$ we get

$$\cos\theta\,\frac{d}{dr}\,E(\theta)\frac{d\omega}{4\pi} = \left[\frac{j\rho}{c} - K\rho E(\theta)\right]\frac{d\omega}{4\pi}$$

or

$$\cos\,\frac{d}{dr}\,E(\theta)\,d\omega = \left[\frac{j\rho}{c} - K\rho\,E(\theta)\right]d\omega$$

integrating over the sphere we get (by using also (29) and (28)

$$\frac{1}{c}\,\frac{dH}{dr} = \frac{j\rho}{c} - E\rho K \qquad (31)$$

multiplying again (27) by $\frac{\cos\theta}{4\pi} d\omega$ and integrating we get by using (29) and (30) $\frac{dP_R'}{dr} = -\frac{K\rho H}{C}$ (32)

Equation (32) is in agreement with equation (18) except that the actual operative stress-component P_R appears instead of the hydrostatic approximation p_R.

Equation (31) may be written in the form

$$cE = \frac{j}{K} - \frac{1}{K\rho}\frac{dH}{dr} \qquad (33)$$

In strict thermodynamical equilibrium with no outward flow (33) becomes on equating $\frac{dH}{dr}$ to zero

or $$cE = j/K \qquad (34)$$
$$j = cEK$$

but since E is equal to aT^4 according to Stefan's law, then

$$j = cKaT^4 \qquad (35)$$

equation (35) gives the well-known law that the emission coefficient j, is proportional to the absorption coefficient, K, for different kinds of matter at the same temperature.

Let us now expand $E(\theta)$ in terms of legendre's polynomials (1) we have $E(\theta) = A + BP_1(\cos\theta) + CP_2(\cos\theta) + DP_3(\cos\theta) + \cdots$ (36)

multiplying (36) by $d\omega/4\pi$ and integrating over a sphere we get

$$\frac{1}{4\pi}\int E(\theta)\, d\omega = A \qquad (37)$$

according to the properties of this expansion. But from equation (28) we get by contrast with (37)

$$E = \frac{1}{4\pi}\int E(\theta)\, d\omega = A \qquad (38)$$

Before going on a property of the zonal harmonics is necessary to be put here leaving the proof to the appendix. This is a recursion formula which enables us to compute the different functions $P_1(\cos\theta)$, $P_2(\cos\theta)$, $P_3(\cos\theta)$ and so on. We have putting (x) for $(\cos\theta)$

$$P_n(x) = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}\left[(2x)^n - \frac{2n}{2n-1}\cdot\frac{n-1}{1}(2x)^{n-2} + \frac{(2n)(2n-2)}{(2n-1)(2n-3)}\cdot\frac{(n-2)(n-3)}{1\cdot 2}(2x)^{n-4} + \cdots\right] \qquad (39)$$

From this recursion formula we can calculate $P_1(\cos\theta)$ $P_2(\cos\theta)$, etc., These are found to be

$$P_1(\cos\theta) = \cos\theta \qquad (40)$$
$$P_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1) \qquad (41)$$
$$P_3(\cos\theta) = \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta) \text{ and so on.} \qquad (42)$$

For a proof of this reference may be made to the appendix. Now multiplying (38) by $\cos\theta\frac{d\omega}{4\pi}$ we get

---

(1) Known in a special case by the name zonal Harmonics: see appendix

$$E(\theta)\cos\theta\frac{dw}{4\pi} = A\cos\theta\frac{dw}{4\pi} + B\cos^2\theta\frac{dw}{4\pi} + \frac{C}{2}(3\cos^2\theta - 1)\cos\frac{dw}{4\pi} + \ldots \quad (43)$$

integrating (43) over a sphere and using (29) we get

$$\frac{H}{C} = \frac{B}{4\pi}\int E(\theta)\cos^2\theta\, dw = \frac{1}{3}B \quad (44)$$

multiplying again (36) by $\frac{dw}{4\pi}P_2(\cos\theta)$ or $\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\frac{dw}{4\pi}$ we get

$$\frac{1}{4\pi}E(\theta)\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)dw = \left[A P_2(\cos\theta) + B P_1(\cos\theta) P_2(\cos\theta) + \frac{C'}{4}(3\cos^2\theta - 1) + \ldots\right]$$

integrating over a sphere and using (28) and (30) we get

$$\frac{3}{2}\left(P_R' - \frac{E}{3}\right) = \frac{3}{2}\left(P_R' - P_R\right)$$

$$= \frac{1}{4\pi}\int\left[\frac{3}{2}E(\theta)\cos^2\theta - \frac{1}{2}E(\theta)\right]dw$$

$$= \frac{C'}{4\pi}\int\{P_2\cos(\theta)\}^2 dw$$

$$= \frac{C'}{5}$$

$$\therefore \frac{3}{2}\left(P_R' - P_R\right) = \frac{C'}{5} \quad (45)$$

Therefore the coefficient A, B, C, of the expansion are found to be

$$A = E,\quad B = \frac{3H}{C},\quad \text{and } C' = \frac{15}{2}\left(P_R' - P_R\right) \quad (46)$$

Going back to the general formula (26) and substituting for the above mentioned expansion we get

$$\left(\cos\theta\frac{d}{dr} - \frac{\sin\theta}{r}\frac{d}{d\theta}\right)\left[A + B P_1(\cos\theta) + C P_2(\cos\theta) + \ldots\right]$$

$$= \frac{jp}{c} - K p\left[A + B P_1(\cos\theta) + C P_2(\cos\theta) + \ldots\right] \quad (47)$$

Using another property of zonal harmonics expressed by

$$\cos\theta\cdot P_m(\cos\theta) = \left\{m P_{m-1}(\cos\theta) + (m+1)P_{m+1}(\cos\theta)\right\}\frac{1}{2m+1} \quad (48)$$

and

$$-\sin\theta\frac{dP_m(\cos\theta)}{d\theta} = \frac{m(m+1)}{2m+1}\left[P_{m-1}(\cos\theta) - P_{m+1}(\cos\theta)\right] \quad (49)$$

and substituting in 47, then by equating the coefficients on both sides thus obtained, we get the series of equations

$$\frac{1}{3}\frac{dB}{dr} + \frac{2B}{3r} = -KpA + \frac{jp}{C} \quad (50)$$

$$\frac{dA}{dr} + \left(\frac{2}{5}\frac{dG}{dr} + \frac{6}{5}\frac{G}{r}\right) = -K\rho B \qquad (51)$$

$$\left(\frac{2}{3}\frac{dB}{dr} - \frac{2}{3}\frac{B}{r}\right) + \left(\frac{3}{7}\frac{dD}{dr} + \frac{12}{7}\frac{D}{r}\right) = -K\rho G \qquad (52)$$

$$\left(\frac{3}{5}\frac{dG}{dr} - \frac{6}{5}\frac{G}{r}\right) + \text{----} \qquad = -K\rho D \qquad (53)$$

and so on to any number of required terms in the expansion, Eddington shows (1) that G can be neglected in comparison with A and that B can also be neglected in comparison with KPA The same is for $\frac{dB}{dr}$ and $\frac{dG}{dr}$ Doing this in (50) and (51) we get

$$K\rho A = \frac{j\rho}{c} \qquad (54)$$

$$\frac{dA}{dr} = -K\rho B \qquad (55)$$

and by putting $\dot{G} = 0$ in (45) we get

$$P_R' = P_R \qquad (56)$$

These results are accurate to within 18 decimal places(2) Since

$$A = E, \text{ and } B = \frac{3H}{c} \text{ we get}$$

$$\left.\begin{aligned} cE &= j/K \\ \text{and} \quad H &= \frac{-c}{3K\rho}\frac{dE}{dr} \end{aligned}\right\} \qquad (57)$$

These equations agree with the ~~previous~~ previous results obtained in a different way.

Absorption and Opacity: It was assumed in the previous derivations of the equations of radiative equilibrium that the coefficient of absorption K is independent of the direction θ of the absorbed radiation. This would have been true, had the radiation been of the same composition all through. But actually the radiation emitted by the higher levels is of a slightly lower frequency than that emitted by the lower levels. This is caused by the difference in temperature between higher and lower levels, a difference

---

(1) See Eddington "Internal constitution of Stars" p. 106
(2) ibid p. 107

which affects the frequency of the emitted light. Since K depends upon the kind of radiation that is absorbed, then it should be different for the different frequencies. The importance of this variation of K in the final results of applying (57) was pointed out by Rosseland (1)

To see this effect let us consider a radiation of frequency between υ and $\upsilon + d\upsilon$. In this case K will be constant and we will denote it by K_υ . Equations (57) will be true now, in the differential form with H, j, and E peculiar to this kind of radiation of frequency between υ and $\upsilon + d\upsilon$. The equations will take the form

$$\left.\begin{aligned} dj &= c\,K_\upsilon\, I(\upsilon)\,d\upsilon \\ dH &= \frac{-c}{3K_\upsilon \rho}\,\frac{dI(\upsilon)}{dr}\,d\upsilon \end{aligned}\right\} \qquad (58)$$

where $I(\upsilon)\,d\upsilon$ is put instead of $dE \cdot I(\upsilon)d\upsilon$ stands therefore for the energy-density for radiation of the range υ and $\upsilon + d\upsilon$. Integrating equations (58) we get

$$\left.\begin{aligned} j &= c\int_0^\infty K_\upsilon\, I(\upsilon)\,d\upsilon \\ H &= \frac{-c}{3\rho}\int_0^\infty \frac{1}{K_\upsilon}\,\frac{dI(\upsilon)}{dr}\,d\upsilon \end{aligned}\right\} \qquad (59)$$

Putting equations (59) in the form of equations 57 we get

$$\left.\begin{aligned} j &= c\,K_1 \int_0^\infty I(\upsilon)\,d\upsilon \\ H &= \frac{-c}{3K_2\rho}\int_0^\infty \frac{dI(\upsilon)}{dr}\,d\upsilon \end{aligned}\right\} \qquad (60)$$

where $$E = \int_0^\infty I(\upsilon)\,d\upsilon$$

and

$$\frac{dE}{dr} = \int_0^\infty \frac{dI(\upsilon)}{dr}\,d\upsilon$$

Since $$\frac{dI(\upsilon)}{dr} = \frac{\partial I(\upsilon)}{\partial T}\,\frac{dT}{dr}$$

---

(1) Mibt Monthly Notices vol. 84 (525) 1924

we can put equations 59 and 60 into the forms

$$\left.\begin{aligned} j &= C\int_0^\infty \kappa_\upsilon I(\upsilon)\, d\upsilon \\ H &= \frac{-C}{3\rho}\int_0^\infty \frac{1}{\kappa_\upsilon}\frac{\partial I(\upsilon)}{\partial T}\frac{dT}{dr}\, d\upsilon \end{aligned}\right\} \qquad (61)$$

and

$$\left.\begin{aligned} j &= C K_1 \int_0^\infty I(\upsilon)\, d\upsilon \\ H &= \frac{-C}{3K_2\rho}\int_0^\infty \frac{\partial I(\upsilon)}{\partial T}\frac{dT}{dr}\, d\upsilon \end{aligned}\right\} \qquad (62)$$

comparing (61) and (62) we conclude that

$$K_1 = \frac{\int_0^\infty \kappa_\upsilon I(\upsilon)\, d\upsilon}{\int_0^\infty I(\upsilon)\, d\upsilon} \qquad (63)$$

and

$$\frac{1}{K_2} = \frac{\int_0^\infty \frac{1}{\kappa_\upsilon}\frac{\partial I(\upsilon)}{\partial T}\, d\upsilon}{\int_0^\infty \frac{\partial I(\upsilon)}{\partial T}\, d\upsilon} \qquad (64)$$

(64) follows because $\frac{dT}{dr}$, the temperature gradient, is independent of the frequency and can be considered as a constant and can, therefore, be cancelled.

From (63) and (64) we see what Rosselland (1) pointed out. Rosselland pointed that the equations of radiative equilibrium derived by Eddington require (according to our last two equations) K to be averaged in two different ways. Thus when the two averaging methods are used simultaneously, the results will be erroneous to the extent of an averaging factor. To distinguish the two methods of averaging, Eddington calls K_1 as the arithmetic mean coefficient of obsroption because it is averaged in the same way as finding the arithmetic mean. "K_2" is given the name of the coefficient of opacity and is averaged in accordance with equation 64, thus it is given by

$$K_2 = \frac{\int_0^\infty \frac{\partial I(\upsilon)}{\partial T}\, d\upsilon}{\int_0^\infty \frac{1}{\kappa_\upsilon}\frac{\partial I(\upsilon)}{\partial T}\, d\upsilon} \qquad (65)$$

---

(1) M. N. 84 525 et seg. 1924

Suppose, for example, that we have studied a range of frequency υ_1 to υ_2 which contains 2/3 of the whole weight of K_2, that i s

$$\int_{\upsilon_1}^{\upsilon_2} \frac{\partial I(\upsilon)}{\partial T} d\upsilon = \frac{2}{3} \int_0^{\infty} \frac{\partial I(\upsilon)}{\partial T} d\upsilon$$

Let the weighted mean value of $1/K_\upsilon$ for this range of frequency be $1/K'$. We can now set an upper limit to K_2, because in the worst possible case even when K_υ is infinite outside this range equation 64 gives

$$\frac{1}{K_2} = \frac{2}{3} \cdot \frac{1}{K'} + \frac{1}{3} \cdot \frac{1}{\infty}$$

or

$$K_2 = \frac{3}{2} K'$$

This is the upper limit of K_2 and whatever happens beyond the limits υ_1 and υ_2 the opacity cannot be increased more than 50 per cent.

The upper limit to K_2 is especially valuable because the danger is that we may be unaware of some important mechanism of absorption and emission. Thus the result of the previous example shows that we can narrow down our research to processes capable of absorbing and emitting frequencies between υ_1 and υ_2 and it relieves us from an exhaustive discussion of very low and very high frequencies, which might be difficult and uncertain.

From Planck's Law

$$I(\upsilon) = \frac{C\upsilon^3}{e^{h\upsilon/RT} - 1} \qquad (66)$$

we get

$$\frac{\partial I(\upsilon)}{\partial T} = \frac{Ch}{RT^2} \frac{\upsilon^4 e^{h\upsilon/RT}}{(e^{h\upsilon/RT} - 1)^2} \qquad (67)$$

putting x for $h\upsilon/RT$ we get

$$\frac{\partial I(\upsilon)}{\partial T} = \frac{C R^3 T^2}{h^3} \frac{x^4 e^x}{(e^x - 1)^2} = K \frac{x^4 e^x}{(e^x - 1)^2} \qquad (68)$$

At temperature T the weight of any range dx to be used in forming the mean value K_2 is proportional to

$$\frac{x^4 e^x}{(e^x - 1)^2} dx \qquad (69)$$

From (69) Table No III has been calculated giving in the second column the relative weight for each value of x and in the third column the weight of the range from 0 to x (1)

---

(1) Eddington "The internal constitution of the stars" p. 112

TABLE NO. III

$x = h\nu/RT$	Weight at x	Weight o to x
o	0.000	0.0000
1/2	0.244	0.0016
1	0.921	0.0121
1 1/2	1.872	0.038
2	2.897	0.084
2 1/2	3.806	0.150
3	4.467	0.230
4	4.864	0.413
5	4.270	0.591
6	3.229	0.736
7	2.194	0.840
8	1.375	0.908
9	0.810	0.949
10	0.454	0.973
∞	0.000	1.000

From this table it is noticed that the weight reaches its maximum at about $x = 4+$ and decreases again. From the third column we see that 69 per cent of the weight is contributed by frequencies between

$2.5\ RT/h$ and $7\ RT/h$

At 10 million degrees the frequency corresponds to a wavelength 14.3 Å so that 69 per cent of the weight is between 6 and 2 Angstroms (A). For elements of moderate atomic weight this corresponds to the L radiation. Therefore in the experimental work on opacity we have to pay attention to this region of the spectrum in particular.

(3) The Radiative Equilibrium of a Rotating Star

Introductory:-- We can tell something about the general effect of rotation on the star before entering into the analytic treatment of the subject. It is clear that when a star starts spinning, the equatorial regions will lift slightly under the influence of centrifugal force. Thus the centrifugal force helps the radiation pressure in supporting the equatorial regions against gravity, and therefore if the rate of liberation of heat is not changed a greater mass can be supported, and the star becomes spheroidal. Since centrifugal force is helping the radiation pressure near the equator, and not near the poles it follows that the force due to radiation pressure should be more at the poles than at the equator; or in other words the net outflow of energy must be greater at the pole than at the equator. This means that the effective temperature at the pole is larger than that at the equator. This result will be discussed in the following pages.

In our previous discussion, the equation of radiative equilibrium was devoted for a non-rotating star at rest under the action of no external forces. It was then found that the relation of the pressure to temperature was of the form

$$P = \frac{1}{3} a T^4 \qquad (70)$$

It was assumed that the production of energy and the absorption coefficient are constant throughout the star. E. A. Milne (1) generalized Eddingtons proof to the case of a rotating star stating from the assumption that the liberation of energy is constant throughout the star. This assumption which in Eddington's theory led to the simple adiabatic law (70) , in Milne's theory no more this pressure-temperature relation has such --simple consequences.

A rotating system of gaseous masses will be in mechanical equilibrium when the gravitational and centrifugal forces are balanced completely by the gas pressure and the radiation pressure. But a system of given masses, mean densities, and angular velocity may be in mechanical equilibrium in an infinite number of different ways, each way being characterized by a functional relation, between the pressure and the temperature, of the form

$$P = f(T) \qquad (71)$$

where f(T) may be an arbitrary function.

H. Von Zeipel (2) dealt with this problem of a rotating star in a more general way than that of Milne. He began with the assumption that the nature of the gas is constant over every level surface, but it may change from one level surface to another.

---

(1) M. N. 83 118 1923

(2) M. N. 84 665 1924

H. Von Zeipel proves then that the state of the gas is constant along a level surface, and that the production of energy, the coefficient of absorption, and the mean molecular weight are also constant on a level surface. The most important consequence of his fundamental hypothesis was that the production of energy per second and gram within the a gaseous mass that rotates like a rigid body is restricted by the following formula

$$\epsilon = \text{constant} \times \left(1 - \frac{\omega^2}{2\pi G\rho}\right) \qquad (72)$$

where ω is the angular velocity, ρ the density, and G the constant of gravitation.
For the importance of this theorem in our coming problems, a simple proof will be given for it here (1)

Von Zeiple's Theorem:

Since the star is in mechanical equilibrium, then the equation of equilibrium will take the form

$$-\frac{\partial P}{\partial x} + \rho\frac{\partial \varphi_0}{\partial x} + \rho\omega^2 x = 0 \qquad (73)$$

$$-\frac{\partial P}{\partial y} + \rho\frac{\partial \varphi_0}{\partial y} + \rho\omega^2 y = 0 \qquad (74)$$

$$-\frac{\partial P}{\partial z} + \rho\frac{\partial \varphi_0}{\partial z} = 0 \qquad (75)$$

where φ_0 is the gravitational potential, and the z-axis is taken along the axis of rotation. These equations of mechanical equilibrium are the same if we consider the mass to be at rest under the influence of the combined potential φ of both the centrifugal and the gravitational forces; then we have the relation

$$\varphi = \varphi_0 + \frac{1}{2}\omega^2(x^2+y^2) \qquad (76)$$

And now the above equations will take the simpler but equivalent forms

$$\frac{\partial P}{\partial x} = \rho\frac{\partial \varphi}{\partial x} \qquad (77)$$

$$\frac{\partial P}{\partial x} = \rho\frac{\partial \varphi}{\partial y} \qquad (78)$$

$$\frac{\partial P}{\partial z} = \rho\frac{\partial \varphi}{\partial z} \qquad (79)$$

---

(1) Eddington "The Internal Constitution of the Stars" p. 282 and Milne--"HandBuch der Astrophysik" Band III Teil 3 page 236

multiplying (77) by dx, (78) by dy, (79) by dz and adding the results we get

$$\frac{\partial P}{\partial x}dx+\frac{\partial P}{\partial y}dy+\frac{\partial P}{\partial z}dz=\rho\left(\frac{\partial \varphi}{\partial x}dx+\frac{\partial \varphi}{\partial y}dy+\frac{\partial \varphi}{\partial z}dz\right)$$

or more simply, since both sides are complete differentials

$$dP=\rho\, d\varphi \tag{80}$$

so that where $d\varphi$ is zero, that is on a level surface, dP also vanishes. Therefore on a level surface P is constant or in other words P is a functi n of φ alone, so that

$$P=f(\varphi) \tag{81}$$

From (80) we have also

$$\rho=\frac{dP}{d\varphi}$$

but from 81 $\frac{dP}{d\varphi}=f'(\varphi)=F(\varphi)$ and therefore

$$\rho=F(\varphi) \tag{82}$$

ρ is also found to be a function of φ alone and therefore it will be constant over an equipotential surface.

Since the temperature is a function of both pressure and density, and since these in turn are also functions of φ ~~alone as in~~ (81) alone as in (81) and (82), it follows that T is also a function of the potential φ only.

All other physical characteristics defining the statistical state of the material depend only on the two variables T and ρ and therefore they are functions of φ alone. From this it follows that all these characteristics will be constant over a level surface. Also it follows that the gradients of any of these quantities will be normal to the level surface.

Applying Poisson's equation

~~$\nabla^2\varphi=\nabla^2$~~ $\nabla^2 \text{Pot } V=-4\pi V$

to (76) where pot V is equal to φ_0 we get by applying the operator ∇^2

$$\nabla^2\varphi=\nabla^2\varphi_0+\nabla^2\left[\tfrac{1}{2}\omega^2(x^2+y^2)\right]$$

or by Poisson's equation

$$\nabla^2\varphi=-4\pi G\rho+2\omega^2 \tag{83}$$

we found before that for a star at rest

$$H=\frac{-C}{K\rho}\frac{dP_R}{dx}$$

Since the flow H is along the normal to the level surface and since P is a function of φ we may write

$$H=\frac{-C}{K\rho}\frac{dP_R}{dn}=\frac{-C}{K\rho}\frac{dP_R}{d\varphi}\frac{d\varphi}{dn} \tag{84}$$

where dn is along the outward normal to the level surface Equation (84) may be put in the form

$$H=-f(\varphi)\frac{d\varphi}{dn} \tag{85}$$

where $$\frac{c}{K\rho}\frac{dP_R}{d\varphi} = f(\varphi) \tag{86}$$

resolving (86) into its rectangular components we get

$$\left.\begin{aligned} Hx &= -f(\varphi)\frac{\partial \varphi}{\partial x} \\ Hy &= -f(\varphi)\frac{\partial \varphi}{\partial y} \end{aligned}\right\} \tag{87}$$

and

$$Hz = -f(\varphi)\frac{\partial \varphi}{\partial z}$$

where Hx, etc., stand for the net flow of radiation across a unit area normal to the x-axis, etc; this is because the lines of flow cross such an area obliquely at an angle whose cosine is $\frac{\partial \varphi}{\partial x} / \frac{\partial \varphi}{\partial n}$

If no additional radiation were being generated the equation of continuity would be

$$\frac{\partial Hx}{\partial x} + \frac{\partial Hy}{\partial y} + \frac{\partial Hz}{\partial z} = 0 \tag{88}$$

But since the rate of generation of energy is $\rho\epsilon$ per unit volume then the conditition of continuity becomes

$$\frac{\partial Hx}{\partial x} + \frac{\partial Hy}{\partial y} + \frac{\partial Hz}{\partial z} = \rho\epsilon \tag{89}$$

From (87) we have by differention

$$-\frac{\partial Hx}{\partial x} = f(\varphi)\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial}{\partial x} f(\varphi)\frac{\partial \varphi}{\partial x}$$

$$= f(\varphi)\frac{\partial^2 \varphi}{\partial x^2} + f'(\varphi)\left(\frac{\partial \varphi}{\partial x}\right)^2$$

Doing the same for Hy and Hz and subsaitituting in (89) we get

$$-f(\varphi)\left[\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2}\right] - f'(\varphi)\left[\left(\frac{\partial \varphi}{\partial x}\right)^2 + \left(\frac{\partial \varphi}{\partial y}\right)^2 + \left(\frac{\partial \varphi}{\partial z}\right)^2\right] = \rho\epsilon \tag{90}$$

And on using Vector notation we may put (90) in the form

$$-f(\varphi)\nabla^2\varphi - f'(\varphi)\left(\frac{d\varphi}{dn}\right)^2 = \rho\epsilon \tag{91}$$

(This is because $(\frac{d\varphi}{dm})^2$ is the square of the resultant force and therefore equal to the sum of the squares of its components $\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}, \frac{\partial\varphi}{\partial z}$, and because $\nabla^2 = \nabla\cdot\nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$)

Substituting in 91 for $\nabla^2\varphi$ as given by equation 83 we get

$$-f(\varphi)\left[-4\pi G\rho + 2\omega^2\right] - f'(\varphi)\left(\frac{d\varphi}{dm}\right)^2 = \rho\epsilon \qquad (92)$$

But $\frac{d\varphi}{dm}$ in a rotating star cannot be constant over a level surface, and also in (92) all the other terms are constants on a level surface and this is possible only if the coefficient of $(\frac{d\varphi}{dm})^2$ is zero, that is if

$$f'(\varphi) = 0$$

or

$$f(\varphi) = \text{Constant} \qquad (93)$$

Equation (92) now may take the form

$$\rho\epsilon = \text{constant} \times (4\pi G\rho - 2\omega^2)$$

or

$$\epsilon = K\left(1 - \frac{\omega^2}{2\pi G\rho}\right) \qquad (94)$$

Which is the famous H. Von Zeipel theorem.

A famous controversy took place on the derivation of this theory between J. Jeans and A. Eddington and Von Zeipel. (1) Sir James Jeans criticized the equation because it is built on Eddington's equation for radiative equilibrium which are only approximate. But generally Von Zeipels theorem was accepted and we will now see some of its consequences.

For slow rotation equation (94) becomes approximately

$$\epsilon = \text{Constant}$$

except in a very small region near the surface of the star. In this region the density is very low. Thus in the case of the sun with a period of rotation equal to 25.5 days we have from 94

$$\epsilon \propto \left(1 - \frac{0.0000195}{\rho}\right) \qquad (95)$$

So that ϵ is constant to within 10 per cent, in all parts where the density is more than 0.0002. Thus as ω becomes smaller and the rate of liberation of energy tends to be constant; but where ρ becomes equal to 0.0000195 in the case of the sun then ϵ becomes zero, that is there will be no liberation of energy on the level surface where ρ is constant and equal to 0.0000195 approximately. Moreover, when ρ becomes less than this value, that is very near the surface of the sun or of a slow rotating star, then the rate of liberation of energy becomes negative. In other words, at this region energy is absorbed instead of being liberated; there is a sink instead of a source.

(1) M. N. 85 ;;. 526, 933, and 678 1925

This conclusion which was inferred by Von Zeipel is physically untenable. To solve this difficulty Eddington infers that the outer parts of the star or sun cannot be rotating as a rigid body in statical equilibrium, thus escaping Von Zeipel's theorem.

It is now wellestablished empirically that the angular velocity of the soalr surface varies with the latitude and there is also no doubt that this variation extends into a great depth in the interior of the sun. Therefore the angular velocity of the sun cannot be a constant, ω. The sun, therefore, might be supposed to have gone on altering its distribution of ω until it reached a state satisfying Von Zeipel's condition. Here we arrive at what was suggested by Milne,(1) that is, the cause of the variation of the solar rotation with latitude is that process which the sun undergoes so that it satisfies von Zeipel's condition and at the same time escapes the unexpected result of a negative liberation of energy.

Yet this is not the only way to look at the problem. The irregular rotation of the sun is equivalent to a superposition of the regular rotation of currents circulatng about the axis of rotation. Circulating currents in other planes also will solve the difficulty. And even it seems more likely that the primary circulating currents are in planes along the meridians. This may be seen from the following consideration(2)

The star can adjust itself in accordance with the ordinary conditions of radiative equilibrium, so that the average temperature over a level surface is maintaned constant. In a non-rotating star not only the average but also the local temperature is constant on a level surface, owing to symmetry conditions; but in a rotating star the condition that the local temperature is also constant on a level surface leads as we have seen to von Zeipel's condition. If Von Zeipels condition is violated then the temperature on a level surface begins to rise at the equator and fall at the poles or vice versa. This will upset the constancy of the pressure over a level surface; and there will be created a pressure-gradient from the equator to the poles (or vice versa) tending to cause a flow of matter. This flow must continue, and take the form of a permanent circulating current. Though by this means the distribution of matter is readjusted, yet it would not bring about equilibrium because no static equilibrium is possible with Von Zeipel's condition unsatisfied. But when the currents become of considerable speed a steady state will be reached because the viscosity of the stellar material is considerable and the fundamental equations of equilibrium will be modified by the addition of viscous stresses. The star then acquires a steady state of circulation. Although the primary currents are set up in planes through the meridians the currents will be deflected east and west by the

---

(1) Observatory 48 p. 73 1925
(2) Observatory 48 73 1925
"Internal constitution of the stars" p. 285 ---Eddington

star's rotation, and thus a secondary phenomenon of east-west or ~~west~~ east currents will be produced. This secondary phenomenon will cause the different periods of rotation of different parts of the sun. Thus the fact that heat and radiation are transferred through the interior of a rotating star leads to unequal heating along the polar and along the equatorial radius, so that a small permanent circulation is maintained. This alteration of temperature between the pole and equator was also inferred from the fact that the sun rotates before discussing von Zeipel's condition, and more-over we found that the temperature at the poles will be more than that at the equator.

Though the east and west currents are of immediate observationed importance, yet the primary currents are also of great interest to us. This kind of curculation, according to Bjerknes will tend to become stratified, this will be treated later when we come to the assumption of V. Bjerknes of the existence of such currents in the sun, ~~sun~~ assumption which enables him to solve most of the solar problems especially sunspots.

(4) A Remark:

A small remark may not be out of place here. If we consider a sphere rotating with a uniform angular velocity ω, then the velocity at the equator will be $V_E = 2\pi R/P$ where P= period $= \frac{2\pi}{\omega}$ and R is the equatiorial radius. The velocity at a point of latitude φ will be $V_\varphi = \frac{2\pi R\cos\varphi}{P}$ Suppose now that the velocity is suddenly decreased all over the sphere for a certain reason. Let the decrease be denoted by v then we have

$$V_E - v = \frac{2\pi R}{P} - v = \frac{2\pi R}{P_1}$$ where P_1 is the new period at equator . also $V_\varphi - v = \frac{2\pi R\cos\varphi}{P} - v = \frac{2\pi R\cos\varphi}{P_2}$

where P_2 is the new period at latitude φ. Therefore we have

$$v = 2\pi R\left(\frac{1}{P} - \frac{1}{P_1}\right) = 2\pi R\cos\varphi\left(\frac{1}{P} - \frac{1}{P_2}\right)$$

or

$$\cos\varphi\left(\frac{1}{P} - \frac{1}{P_2}\right) = \frac{1}{P} - \frac{1}{P_1}$$

From this result we notice that P_2 is more then P_1 and therefore ω_1 is more than ω_2. The dynamic encounter theory gives an explanation to the change of the angular velocity, and the velocity is gained by the re-attracted masses which are more concentrated at the equator giving it more angular velocity. Apart from this explanation, there might be a connection between the previous result (when taken in a more general form by considering infinitesimal changes) and the von Zeipel theorem or in a more general way between it and the liberation of energy in the star. The connection as it appears to me now is that according to our previous result the temperature near the pole is more than near the equator. From Equation 94 we have

$$\omega^2 = \frac{K-\epsilon}{K} 2\pi G\rho \qquad (96)$$

If we suppose that the liberation of energy is caused to be more near the poles so that the radiaion pressure will be more than near the equator, then from (96) it follows that ω should be less to the north of the equator. How this change in ω is brought out may be according to Eddington's and Milne's suggestions. But to suppose that the circulation brings out the fact that ω should vary seems to be an indirect or forced explanation. But the reverse that ω should vary and this variation causes the circulation seems to be more legitimate. However this needs further development.

The general problem of radiative equilibrium of a rotating star has been treated by E. A. Milne (1) and H. von Zeipel (2) Milne adopted the approximation ϵ = constant and von Zeipel adopted condition (94). This will not be treated here.

---

(1) M. N. 83 p. 118 1923
(2) M. N. 84 p. 665 1924

CHAPTER III

The General Magnetic Field of the Sun (1)

Introduction:

Rowland's success in 1876 in producing a magnetic field, by whirling an electrically charged disc at a high velocity(2) was the first experimental proof for Maxwell's hypothesis that an electrified body in motion is equivalent to an electric current. This success marks the beginning of the investigations on the magnetic fields of rotating bodies. Schuster suggested in 1891 that a rotating body may behave like a magnet(3), a suggestion that was then expressed by Lord Kelvin in his opinion that the earth's magnetic field must be due to its rotation.

In 1894, J. J. Thomson remarked that if the atoms exert different attractions on positive and negative electric charges, then a large rotating body ought to produce a magnetic field. The maximum magnetic force at the surface of a rotating sphere would be proportional to ωr^2 where ω is the angular velocity of rotation and r the radius. Assuming the earth's magnetic force to be the maximum attainable by the rotation of a sphere the size of which is the same as that of the earth, he calculated, that the magnetic force of a sphere one foot in radius rotating with an angular velocity of one hundred rotations per second would be about one hundred-millionth part of that of the earth, a quantity which cannot be detected in the labratory (4) This idea was then applied by Sutherland in his hypothesis of terrestial magnetism, and he pointed out that the external electric effect would be overcome by the presence within the earth of equal charges of positive and negative electricity. These he supposed to be spread over concentric spheres of radii that differ only by a very small quantity.

Baner adopted a similar view in the paper he read before the American Association on December 31st 1912. The symmetrical part of the earth's field can then be accounted for by supposing the radius of the sphere containing the positive charge to be only 0.44×10^{-6} cm. smaller than that of the sphere conteining the negative charge. This difference is about four tenth's of the radius of a molecule.(5) A similar result is obtained for the portion

(1) Astrophysical Journal 38 pp. 27-125 1913 and HandBuck der Astrophysik Band IV pp. 200-204
(2) Rowland "On the magnetic effect of electric connection" American Journal of Science (3) 15, 30-38, 1878
(3) Proce physical society of London 3, 57, 1879
(4) "On the Electricity of Drops" Phil. Mag. (5) 37, 358, 1894
(5) Wilson "Structure of Atoms" science, N. S. 35, 511, 1912

of the earth's field due to the effect of the atmosphere. If the positive and negative electrons differ in mass and possess inertia, the earth's centrifugal force may produce a spheriodal rather than a spherical distribution of the charges, thus accounting for the observed increase in the equivalent intensity of magnetization towards the equator. Assuming the sun's field to be due to the same cause as that of the earth; Baner computed that the vertical magnetic intensity at the sun's poles is about 300 gausses. This value is in close agreement with that of Schuster who had previously shown that the magnetic intensity of the sun should be about 440 times greater than that of the earth(1).

Schuster (though rejects the hypothesis that a neutral molecule in its motion behaves as if it carried a charge) preferred a very different theory which assumed that every molecule is a magnet. If this magnetism is accounted for as the effect of the rapid revolution of electrons within the molecule, a gyrostatic action may be anticipated. That is, each molecule would tend to set itself with its axis parallel to the axis of the earth, and the earth's magnetic field would result from the combined effect of all the molecules. This theory, like the preceding one, has its weak points. But its chief advantage lies in the possibility that it may explain the secular variation of the earth's magnetism by a precessional motion of the magnetic molecules.

The suggestion of Schuster, that every rapid rotating body may produce a magnetic field is a very important physical problem. The existence of the earth's magnetic field is a favourable phenomenon to this hypothesis, but its experimental proof appears to be beyond the reach of our present scientific possibilities due to the limitations of size and rotational velocity imposed by labratory conditions. Therefore only an inductive method may help us, and that is to investigate whether the rotating heavenly bodies exhibit a similar magnetic phenomena as that of the earth.

The best heavenly body suited for the purpose is, evidently, the sun. Its great radius and its angular velocity of rotation sho ld give rise to a magnetic field which is more than four hundred times as intense as that of our planet. Also the solar atmosphere contains self-luminous gases giving line-spectra capable of revealing the magnetic field by observing the Zeeman effect. The brightness of the sun is sufficient to permit the use of the very high dispersion required to detect a field so much weaker than the fields usually employed in laboratory studies of radiation. Finally its axial rotation and large anguler diameter facilitate observations at a great number of points on its surface, while the position of its axis, allowing displacements to be measured near both poles, enables the observer to apply the most perfect test to the Zeeman effect--the reversal of the sign of the displacement with the polarity. Yet a very important limitation exists and that is the existence of free electrons and ions in the solar atmosphere and near sunspots; these electrons and ionsmay produce by their motion local and general magnetic fields. Nevertheless, it is possible to determine accurately the part played by free electrons in spots, and their effect on the general

---

(1)Schuster "A critical examination of the possible causes of terrestrial magnetism" Proc. Physical Society of London 24, 127, 1912

magnetic field may be determined.

The possibility of a general observational work on solar magnetic phenomena was strengthened by the discovery of the intense magnetic fields in sunnapots by Hale(1). The Zeeman effect was the fundamental principle of much investigations, and it was found to extend well beyond the limits of the penumbra of spots; and the configuration of the hydrogen flocculi suggested that with suitable polarizing apparatus local fields might be dtected in regions where no spots were visible. The next step would therefore be the investigation of the solar general magnetic field.
Bigelow applied an indirect method in 1889 in which he used the coronal streamers to indicate the existence of the magnetic field. He inferred that the sunsmust be a magnet because the coronal streamers agree well in form with the lines of force of a magnetized sphere (2). Störmar claculated the trajectories of electric corpuscles moving out from the sun under the influence of an assumed magnetic field, and the resulting curves closely resemble the structure of the corona.(3) Also Delanders applied the same idea in the case of prominences, and he concluded from their forms and their radial velocities that the ions which compose them are moving under the influence of a magnetic field. (4)

Hale pointed out that the obove mentioned indirect methods are not decisive, because the phenomena used are related to the magnetic field at very high levels of the solar atmosphere, and this field may differ in intensity and may even be opposite in polarity to the field of the sun lying withing the photosphere; The only direct method which is so far known is therefore the Zeeman effect, and its results only can be taken as reliable. Before entering into the results of Hales works, a brief discussion of the Zeeman effect would not be out of place here.

The Zeeman Effect:

Faraday's last experimental work in 1862 was on the relation between light and magnetism. Having previously discovered the rotation of the plane of polarization of light travelling along the lines of magnetic force through certain substances placed in a magnetic field, he sought to detect any change in the lines of the spectrum of a flame when the flame was put between two strong magnetic poles. He examined the spectrum of the "D" lines given out by a sodium flame, but he failed to notice any effect on the lines when a current was passed in the coils of the electromagnet.

However, in 1892, Zeeman tried with success the same effect on which Faraday worked. The improved apparatus at the time of Zeeman made him able to repeat FaradaY's experiments and he noticed

---

(1) Astrophysical Journal 28, 315 - 348, 1908
(2) Bigelow "The Solar(Grerna) corona" Smithonian Institution 1889
(3) Stormer Comptes Rendus Feb., 20, 1911
(4) Comptes Rendus December 30, 1912

the change in the "D" lines of sodium. The lines were widened when the flame was put between two powerful magnetic poles. The results of Zeeman's experiment were communicated to Lorentz, who pointed out the explanation according to his electronic theory, and in addition he also predicted that the edges of the modified lines ought to be circularly polarized, when the rays of light are travelling along the lines of magnetic force . This prediction was then varified experimentally by Zeeman himself.

The effect observed depends on the direction of vision with respect to the magnetic field, which gives two distinct cases:

(1) When the rays of light whose spectrum is formed are perpendicular to the lines of force of the magnetic field, the effect is then called the "transverse" Zeeman effect.

and (2) When the rays of light are parallel to the direction of the magnetic field the effect is called the "longitudinal" Zeeman effect.

Both the transverse and the longitudenal effects are shown in figure (4) where "a" represents a line before the magnetic field is excited. In the transverse effect the original line is split up into a triplet b_1, a_1, b_2 as shown in the figure; this is called the Zeeman normal triplet. In the longitudinal effect the original line is split up into two components c_1 and c_2 called the Zeeman doublet. In the transverse effect the line a_1 has the same frequency as the original line "a", In addition to this splitting the components are polarized even though the original line is not polarized. The central component in the transverse effect is plane polarized with the electric vector parallel to the field; while the other two components are displaced by an amount of $\pm \Delta\upsilon$ occupying the same positions as the Zeeman doublet and at the same time they are also plane polarized but with the electric vector perpendicular to the field. On the other hand, the Zeeman doublet, caused by the longitudinal effect, are circularly polarized but each in opposite sense, and they are displaced by an amount $\pm \Delta\upsilon$, where "υ" is the frequency, equal to the displacement of the transverse effect.

Though the Zeeman effect was explained according to the Lorentz theory and the classical theory of light, yet the application of the quantum theory to this problem was more sucessful. The first application of the quantum theory was done by Hertzfeld and Bohr, and lately by Debye and Sommerfeld. However, we shall consider here the simple deduction of the frequency change given by Sommerfeld.(1)

According to Bohr's law of radiation, radiated energy is emitted only in the transitions of electrons between two stationary states; the frequency υ of the emitted spectral line being determined by the relation $h\upsilon = E_1 - E_2$ where E_1 and E_2 are the initial and final energies of the electron causing the emission of that line. The effect of the magnetic field is to produce a change in the energy in both initial and final states of an amount ΔE_1 and ΔE_2. Therefore the change of the frequency of the emitted radiation is determined by

$$h\Delta\upsilon = \Delta E_1 - \Delta E_2 \qquad (97)$$

---

(1) "The Quantum" by Stanley Allen page 209

The problem to be solved is now to determine the change in the energy of the stationary state due to an applied magnetic field of strength H. Sommerfled assumed that the magnetic field produced orientation in space in the form of the orbit, while the radius of the orbit might not be changed. The change in energy is equal simply to the change ΔE in the kinetic energy and the amount of this change is given by

$$\Delta E = \frac{m h}{2\pi} \omega \qquad (98)$$

where ω is the angular velocity corresponding to the Larmor precession, m is the equitorial or orbital magnetic quantum number.
But also ω is given by

$$\omega = \frac{1}{2} \frac{e}{m_0} \frac{H}{c} \qquad (99)$$

where e is the electronic charge in e.s.u., H in e.m.u. and m_0 is the mass of the electron and c is the velocity of light . Substituting the value of ω from (99) into 98 we get the value of ΔE given by

$$\Delta E = \frac{mh}{2\pi} \cdot \frac{1}{2} \frac{e}{m_0} \frac{H}{c} = mh \frac{e}{m_0} \frac{H}{4\pi c} \qquad (100)$$

substituting the value of ΔE from (100) into (97) we get

$$h \Delta \nu = h (m_1 - m_2) \frac{e}{m_0} \frac{H}{4\pi c} \qquad (101)$$

or

$$\delta \nu = (m_1 - m_2) \frac{e}{m_0} \cdot \frac{H}{4\pi c} \qquad (102)$$

We notice from this expression that "h", the planck's constant, disappears, which makes Sommerfeld to remark, "In our final formula the quantum theory has, in a certain sense, became latent, in that its characteristic feature, the quantity h, has disappeared."
The change in frequency, therefore depends on the magnetic quantum numbers. According to the selection principle, in ordinary cases the change in the orbital magnetic quantum number is limited to the values ± 1 or zero.

In the first case when the difference between these numbers is ± 1, the change in frequency of the resulting components of the spectral lines is therefore

$$\Delta \nu = \pm \frac{e}{m_0} \frac{H}{4\pi c} \qquad (103)$$

and the change in the frequency makes the new frequency to be

$$\nu = \nu_0 \pm \frac{e}{m_0} \frac{H}{4\pi c} \qquad (104)$$

Each of the two components, corresponding to the increased and decreased frequency, is plane polarized with the plane of polarization perpendicular to the field.

In the second case when the change is zero, we have $\Delta\nu = 0$, (105) and therefore the linee has another component of the same frequency and thus no displacement is observed. This undisplaced component is plane polarized with its electric vector parallel to the magnetic field and is seen only when looked at transversely.

The above treatment though elementary and incomplete yet may serve as an explanation to what is happening in such cases. Thus if the sun has a magnetic field it could be detected by means of this Zeeman effect. This caused Hale in 1908 to investigate the magnetic field of suspots, and again in our case, to investigate the general magnetic field of the sun.

Hale's Work:

G. Hale selected for his investigations the region λ 5800 to λ 6000 Angstroms in which the solar lines λ 5812.139, λ5828.097 and λ 5851.821 are found. This selection was determined by three principal considerations:

(1) The less refrangible region of the spectrum is advantageous because, on the average, the separation of the components of lines by a magnetic field varies directly as the square of the wavelength.

(2) However, too great a wavelength is undesirable since the average sharpness of the solar lines decreases as the wavelength increases.

(3) The larger separation $\frac{\Delta\lambda}{\lambda_2}$ observed in the labroratory for certain spark lines indicate that these should be tried if sharp enough in the sun.

A detailed description of apparatus and method of observation and of data obtained therefrom may be found as given by Hale himself in the Astrophysical Journal 38 27 1913. But we will mention here Hale's results only.

The evidence presented by Hale's work seems sufficient to prove that the observed displacements are caused by magnetic fields in the sun. The next problem is to decide whether these fields are due to local phenomena or represent the magnetic effect of a rotating sphere; It is known that sunspots show the Zeeman effect and that the widening of the lines frequently extends beyond the boundary of the penumbrae. Magnetic fields may also be caused by invisible spots or by whirls in which no umbrae or penumbrae have appeared. Also there is some evidence to support the view that the pores are small vortices, which develop into spots under favourable conditions.

It is therefore necessary to determine the effect of such causes on the displacement of the spectral lines.

Hale solves this difficulty by negating any considerable effect of such causes on the general magnatic field. This he claims for the following reasons:

(1) The observations of sunspots made by Hale indicate that right-handed and left handed whirls are about equally common in the northern and southern hemisphere. The great majority of spots consist of two principal members, frequently attended by satellites, the line joining the chief spots usually making a small angle with the equator. In general, Hale found

that these groups are of the bipolar type. Hence there is no reason to suppose that the influence of spots visible or invisible could be of such a character as to produce Zeeman displacements which on the average are of opposite sign in the northern and southern hemispheres.

(2) The observations were made during a low minimum of solar activity, and in the great majority of cases no spots whatever and few K_2 flocculi, were visible on the sun.

(3) If the pores are electric vortices, like the spots, there is no reason to suppose that the pores of one polarity preponderate in the northern hemishphere and those of opposite polarity in the southern hemisphere.

(4) Even if there were a clear preponderance of pores of opposite sign north or south of the equator, it would be difficult to account for such a curve of displacements as the plotted observations represent.

(5) Assuming however, that such a curve could be plausibly explained as originating in the pores, it is evident from the character of the curve that we should be dealing with a general magnetic field of the sun, though not one caused by the solar rotation.

These and other reasons make Hale to conclude the existence of the gneral magnetic field of the sun; and he is more inclined that it is caused by the rotation of the sun. The results of Hale's work indicate that the field strength of the sun's magnetic field at the poles is of the order of 50 gausses.

The three lines on which attention has been concentrated in Hale's work are probably produced at a comperatively low level. These clearly show the effect of the suns field, while various higher level lines, promising from the laboratory stand point, failed to do so. Thus Hale concludes that it is probable that the intensity of the general magnetic field falls off rapidly in passing upward through the reversing layer.

A serious objection to the application of all theories of terrestial magnetism is that they cannot be applied without modification to the sun because of the soler high temperature, low density, and gaseous condition. In the case of sunspots, neutral molecules cannot produce the observed fields unless an improbable degree of centrifugal separation of the positive and negative electrons is assumed. However Hale suggests that Harker's investigation, on the two carbon electrodes, may be of direct application to this problem.(1)

This investigation was done onttwo carbon electrodes which were mounted within a carbon resistance furnace 5 mm apart and connected with a galvanometer. At atmospheric pressure and temperature of 2500° c a current of nearly 2 amperes was observed when one electrode was cooled by temporarily removing it from the hot part of the furnace. The cooler electrode, on which much carbon was deposited, was the positive one.

The same may be said in the case of sunspots. Since they are cooler than their sunroundings, a flow of negative electrons should, therefore, take place on all sides towards the umbra.

---

(1) Harker "Very High Temperatures" Nature July 18, 517 1912

These electrons whirled in the vortex may account for the strong magnetic fields observed in the regions of spots. This we shall treat later in our discussion on sunspots.

Hale Summarizes his results in the following:

(1) The lines λ 5812.139, λ5828.097, λ5851.821, and λ5929.898 show distinct displacements not shared by atmospheric lines nor by certain other solar lines.

(2) The sign of the displacements is opposite in the northern and southern hemispheres of the sun.

(3) (3) The maximum displacements are observed about 45° north and south of the solar equator. From this point they decrease to zero at the equator and near the poles of rotation.

(4) A curve representing the displacements as a function of the latitude corresponds closely with a theoretical curve, showing the displacements of a normal Zeeman triplet observed at various latitudes in the field of a magnetized sphere.

(5) In view of this agreement, and the apparent impossibility of accounting for the observed displacements on other grounds, it is probable that they represent the Zeeman effect due to the sun's general magnetic field.

(6) Assuming this to be true, we find that the magnetic poles of the sun lie at or near the poles of rotation.

(7) The polarity of the sun corresponds with that of the earth.

(8) On the hypothesis that the magnetism of the sun is due to the axial rotation of a body acting as though it carried a residual volume charge, the sign of the charge comes out to be negative.

(9) The preliminary results indicate that the general magnetic field decreases rapidly in intensity at levels in the solar atmosphere higher than those represented by the lines in part (1) of this summary.

(10) A first approximation value for the vertical intensity of the sun's general field at the poles is 50 gausses.

These are the preliminary results of Hale and now we shall discuss the work of Seares on this subject by giving a summary of his theoretical discussion.

Seares Investigations (1)

On the supposition that the distribution of magnetic force on the solar surface is similar to that on the earth its intensity at magnetic latitude φ' is given by

$$H = H_e \sqrt{1 + 3\sin^2\varphi'} \qquad (106)$$

where H_e is the equatorial strength of the field. Taking δ as the separation of the outer components of a normal Zeeman triplet from the central one, and γ as the angle between the line of sight and the lines of force, Seares deduced the relative shift Δ in adjacent strips from the formulae which give the distribution of intensity of triplets

$$\Delta = 2\delta \cos\gamma \qquad (107)$$

---

(1) Seares "The Displacement Curve of the Sun's General Magnetic Field" Astrophysical Journal 38 page 99 et. seq. 1913

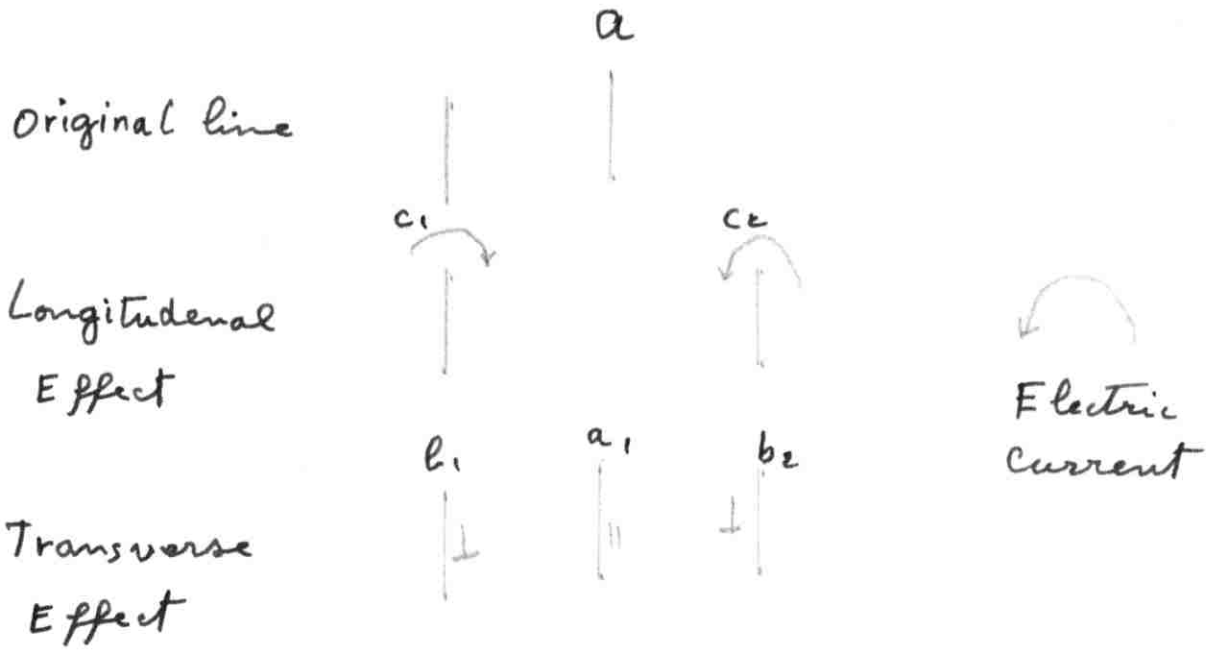

Figure 4. Zeeman Normal Separation

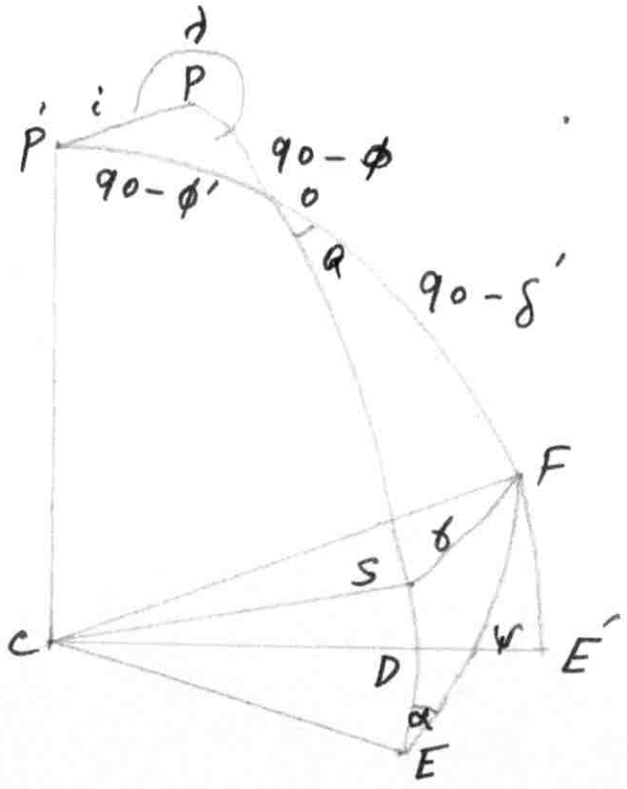

Fig. 5. Determination of the sun's General magnetic field.

and he gets

$$\Delta = 2\,c\,He\cos\gamma\sqrt{1+3\sin^2\varphi'} \qquad (108)$$

where c is a constant depending upon the units adopted.
To determine the angle γ reference is made to figure (5) (1) and having in mind that the observation are always made with the slit in coincidence with the central meridian of the solar image. We have in the figure

P, = sun's north pole,
P' = adjacent magnetic pole
C = observed point, always on the central meridian
E, = intersection of central meridian with solar equator.
E' = intersection of magnetic meridian through C with magnetic equator.
S = intersection of line of sight with sphere.
F = intersection with sphere of tangent to line of force at observed point.
φ = EC = heliographic latitude of C
φ' = E'C = magnetic latitude of C
D = ES = heliographic latitude of sun's centre
Ω = angle at C between central meridian and magnetic meridian
S' = angle between sun's surface and line of force at observed point.
i = pp' = inclination of magnetic axis to solar axis of rotation.
γ = SF = angle between line of sight and line of force at observed point.
$\lambda = \angle OPP' =$ Longitude of magnetic pole referred to central meridian and measured in the direction of sun's rotation.
$\psi = \angle ECF = \text{arc } EF$
$\alpha = \angle SEF$

From the spherical triangle SEF we have :

$$\cos\gamma = \cos D\cos\psi + \sin D\sin\psi\cos\alpha \qquad (109)$$

From triangle EOF

$$\cos\psi = \cos\varphi\sin\delta' + \sin\varphi\cos\delta'\cos\Omega$$
$$\sin\psi\cos\alpha = \sin\varphi\sin\delta' - \cos\varphi\cos\delta'\cos\Omega \qquad (110)$$

The inclination of the lines of force to the sun's surface at the observed point is connected with the magnetic latitude by the relation

$$\tan\delta' = 2\tan\varphi' \qquad (111)$$

Whence we find

$$H'\sin\delta' = 2\sin\varphi', \quad H'\cos\delta' = \cos\varphi' \qquad (113)$$

Where

$$H' = \sqrt{1+3\sin^2\varphi'} \qquad (113)$$

Substituting (113) into (110) we get

$$H'\cos\psi = 2\cos\varphi\sin\varphi' + \sin\varphi\cos\varphi'\cos\Omega$$
$$H'\sin\psi\cos\alpha = 2\sin\varphi\sin\varphi' - \cos\varphi\cos\varphi'\cos\Omega \qquad (114)$$

From triangle OPP we have

---

(1) ibid page 106.

$$\left.\begin{aligned} \sin\varphi' &= L \sin\varphi + M\cos\varphi \\ \cos\varphi'\cos\varphi &= L\cos\varphi - M\sin\varphi \end{aligned}\right\} \quad (115)$$

Where

$$L = \cos i \qquad M = \sin i\cos\lambda \quad (116)$$

Substituting (115) into (114) we get

$$\left.\begin{aligned} 2H'\cos\psi &= 3L\sin 2\varphi + 3M\cos 2\varphi + M \\ 2H'\sin\psi\cos\alpha &= 3M\sin 2\varphi - 3L\cos 2\varphi + L \end{aligned}\right\} \quad (117)$$

Then by substituting (117) into (110) we have

$$2H'\cos\gamma = 3L\sin(2\varphi - D) + 3M\cos(2\varphi - D) + L\sin D + M\cos D \quad (118)$$

combining (108) and (118) and resubstituting the values of L and M we

$$K\Delta = \left[\begin{aligned} &\{3\sin(2\varphi - D) + \sin D\}\cos i \\ &+\{3\cos(2\varphi - D) + \cos D\}\sin i\cos\lambda \end{aligned}\right] \quad (119)$$

Where the factor K is a new constant depending upon the units and the field-strength at the sun's magnetic equator. This equation is the required one.

Since the coefficients of $\cos i$ and $\sin i\cos\lambda$ in (119) may be replaced by expressions of the form $(n\cos N)$ and $(n\sin N)$ it appears that, whatever are the values of D, i, and λ, the displacements always define a sine curve. But the latitude of the sun's centre, D, is small, never exceeding 7°, and the observations show that i is also small, if not actually zero. The displacement curve is therefore approximately $3\sin 2\varphi$. Its ordinates are zero near the equator and the poles; they have opposite signs in the two hemispheres, and maximum absolute values near $\varphi = 45^\circ$ north and south of the equator.

Equation (119) contains three unknowns K, i, and λ. To determine them, let us consider first values of the displacements for numerically equal northern and southern latitudes, observed on or near the same date.

The angle D will then be sensibly unchanged. Denoting the displacements by Δ_m and Δ_s we find

$$K(\Delta_m - \Delta_s) = 6 \sin 2\varphi (\cos D \cos i + \sin D \sin i \cos \lambda) \quad (120)$$

The quantity in parenthesis in the right hand side of (120) differs from unity by a quantity of the second order in D and i, and it is quite sufficient to write

$$K = \frac{6 \sin 2\varphi}{\Delta_m - \Delta_s} \quad (121)$$

The denominator has its maximum near $\varphi = 45^\circ$, and the equation is most advantageously used for points near this latitude.

Writing now (119) in the form

$$K\Delta = A \cos i + B \sin i \cos \lambda \quad (122)$$

we have, with sufficient approximation,

$$\sin i \cos \lambda = \frac{K\Delta - A}{B} \quad (123)$$

Equations (121) and (123) serve for the determination of K, i, and λ. Although the introduction of K affords a convenient arrangements of the formulae for the calculation of i and λ, we are really interested in the value of the solar field-strength H_e upon which K depends.

By comparing (107) and (108), it appears that when Δ is expressed in Angstrom units, c represents the separation in Angstroms of a line whose plane of polarization is normal to the slit and a line whose plane of polarization is parallel to the slit (1) split from the line observed, and this separation is produced by a field of one gauss. When this quantity has been determined by appropriate laboratory investigations, the solar field-strength can be determined. Because (108) may be written in the form

$$\Delta = c\, H_e\, F \quad (124)$$

where F is the right hand side of (119) and Δ is supposed to be expressed by Angstrom units. When i and λ have been found by the method outlined, H_e may be calculated by (124) or by an equation analogous to (120), namely

$$\Delta_m - \Delta_s = 3c H_p \sin^2 \varphi (\cos D \cos i + \sin D \sin i \cos \lambda), \quad (125)$$

into which the polar field-strength $H_p = 2H_e$ has been introduced. For first approximation we may write $i = 0$, $\cos D = 1$, and applying (125) to $\varphi = 45^\circ$ we get for H_p

$$H_p = \frac{2\Delta_{45}}{3c} \quad (126)$$

in which Δ_{45} is the mean displacement at 45°.

This derivation was given by Seares on the assumption that the line observed is a normal Zeeman triplet. But Seares took

---

(1) In the apparatus used by Hale, if a beam of circularly polarized light passes through the slit, it will be transformed into a series of plane polarized beams whose vibration planes are alternately parallel to and perpendicular to the slit.

then into consideration the general case of the anamolous Zeeman effect where greater complexity than triplets is found. Also he continues his modification and takes into consideration the effect on the displacement curve of elliptical polarization due to reflection on the silvered mirrors of solar towers. We will not enter here into Seares further deductions, and it is sufficient to summarize his theoretical results and their relation to Hale's results.

For a normal Zeeman triplet the theoretical displacement curve, equation (119) is a function of the heliographic latitude, the position of the observer, and the solar magnetic elements. It is a sine curve differing but little from $K\Delta = 3\sin 2\varphi$ and has therefore, zero values near the equator and the poles, and absolute maxima near 45^0 north and south. It can be adapted to the calculations of the solar magnetic elements.

More extensive observations have been made by Hale, Seares, Van Maanan, and Ellerman to determine the elements of the suns general magnetic field and the probable variation of its intensity at different levels of the solar atmosphere (1). A summary of the results will be given here.

It was mentioned before that only four lines were shown to have displacements that could be attributed to the general magnetic field of the sun. Also it was mentioned that a number of other lines, mainly stronger lines known from laboratory investigation to have large Zeeman separations, showed no corresponding solar displacements. The explanation given was that the displaced lines probably originate at a lower level in the solar atmosphere, while the others correspond to higher levels where the field is too weak to be detected. Now since the intensity of a line is probably a function of its level in the solar atmosphere, then it should be possible to determine the decrease in intensity of the field with increasing elevation. Therefore measurement of displacements and intensities of other spectral lines in the solar spectrum will make it possible to determine the elements of the field at different levels, by using the previous equations.

The results of Hale and collaborators indicate the presence of a magnetic field, confirm the results in Hale's previous work, and place beyond doubt the conclusion that the sun bhhaves approximately as a uniformly magnetized sphere with the magnetic axis only slightly inclined to the solar axis of rotation and a polarity corresponding to that of the earth. A comparison of the results given by various lines shows that the field decreases with increasing values of the lines intensities. With the exception of one of the titanium lines and one of vanadium, all the results give approximately a value of 55 gausses for $H\rho$ at a level of 250 km, and 10 gausses at a level of 420 km. It would appear from this that only the lines in a layer of small depth are sensibly affected by the magnetic field and are therefore measurable by this method.

The coordinates of the magnetic poles were determined by the dissymmetry of the displacement curve drawn for different epochs; and since the dissymetry is very small, it can be inferred that the magnetic poles cannot be far from the poles of rotation.

---

(1) A. P. J. 47 206--254 1918

Yet the causes of the existence of a magnetic field in the sun remain unknown, the relation of a rotating body to its magnetic field is still an interesting hypothesis which needs investigation.

PART II

THE SUNSPOT PHENOMENON

AND OTHER RELATED STUDIES

CHAPTER IV

Introductory Studies

It seems indispensable for the study of sunspot to give a somewhat satisfactory description of other phenomena connected with the problem of sunspots. The brief and elementary discussion of the Saha ionization theory, the radiative equilibrium, and the general magnetic field of the sun is sufficient for our purpose; but the extremely brief discussion of prominences, flocculae, and the solar spectrum is very unsatisfactory in comparison to their immediate relation to sunspots. For this reason these phenomena will be discussed in this introductory chapter in more detail.

Some Spectroscopic Results: The Fraunhofer Spectrum:

According to our previous discussion of the solar spectrum and of the ionization in the solar atmosphere it is apparent that most of the Fraunhofer lines are due to neutral atoms of those elements which have very high ionization potential and to ionized atoms of those elements which have low ionization potentials. From our present knowledge of the temperature and pressure of the solar atmosphere and of the spectrum of the sun and of specral analysis, the following elements are certainly present in the sun (1)

Hydrogen	Calcium	Ythrium	Neodymium
Helium	Scandium	Zirconium	Samarium
Lithium	Titanium	Columbium	Europium
Beryllium	Vanadium	Molybdenum	Gadolinium
Barium	Chromium	Ruthenium	Dysprosium
Carbon	Manganese	Rhodium	Erbium
Nitrogen	Iron	Palladium	Ytterbium
Oxygen	Cobalt	Silver	Lutetium
Sodium	Nickel	Cadmium	Tungsten
Magnesium	Copper	Indium	Osmium
Aluminium	Zinc	Antimony	Iridium
Silicon	Gallium	Barium	Platinum
Phosphorus	Germanium	Lanthanum	lead
Sulphur	Rubidium	Cerium	
Potassium	Stronium	Praseadymium	

The presence of the following elements in the sun is probable: Tin, Terbium, and Thalium (2)

---

(1) Abetti " The Sun " Page 93

(2) The element Thalium is finally detected in the sun Science News Letter. October 18, 1941

The following elements are not yet detected in the sun:

Flourine	Krypton	Tantalum	Radon
Neon	Masurium	Rhenium	Radium
Chlorine	Tellurium	Gold	Actinium
Argon	Iodine	Mercury	Thorium
Arsenic	Xenon	Bismuth	Proteactinium
Bromine	Holmium	Selenium	Caesium
Polonium	Uranium		

Yet there is recently a theory which leads to the fact that the surface temperature of the sun should be millions of degrees centigrade (1); this temperature will cause multiple stage ionization of many elements and renders their detection in the sun very difficult, and therefore, if this temperature is near the truth then the above unidentified element may be all existing in the sun, especially there is no a priori reason why certain elements to the exclusion of others should exist in the sun.

In addition to the separate lines of the Fraunhofer spectrum, molecular bands are observed. The presence of such bands tells us that the temperature of that part of the photosphere where the elements giving rise to them are found should be, together with the pressure, of such a value as to allow these elements to exist in their combined state without decomposition into their constituent atoms. According to the hitherto accepted identifications, the principal bands are due to cyanogen (C_2N_2), carbon molecules (C_2, or Swan bands), the hydrides (CH, NH, OH), the hidrides of aluminium, titanium, and Zirconium. Some of these bands, as we shall see, are more intense and only exist in the spectrum of sunspots.

Spectra of the Solar Limb, Spots, Reversing Layer, and the Chromosphere

If we pass gradually from the spectrum of the center of the solar disc to that of the limb, a gradual variation in the spectrum will be observed; when we arrive nearest to the limb, three characteristics of the spectrum will be apparent; these were first noticed by Hasting and studied later more carefully by Hale and Adams. (2) The characteristics are:

(1) Some of the most intense lines at the centre of the disc (as the H and K lines of calcium) do not show sharp edges in the continuous spectrum; but a gradual decrease in intensity, forming what is called the wings of the line, is observed. These wings become fainter and fainter as we approach the limb and even in some cases they completely disappear.

(2) Many of the lines widen at the limb, diminishing at the same time their contrast from the continuous spectrum.

(3) Some of the lines become more intense at the limb, others less so, according to the energy level at which the lines are ~~(become more intense at the)~~ produced.

The differences in the relative intensities of the lines between centre and the limb are more marked in the blue, violet, and ultraviolet regions of the spectrum, and are less

---

(1) Science News Letter June 14, 1941
(2) A. P. J. 25 300 1907

The most extensive work on the sunspot spectrum has been done at Mt. Wilson. The following table [1] of the lines affected by the various elements in the region between λ 5900 and λ 7000 summarizes these investigations.

TABLE IV

Element	Total No. of lines	N lines strengthened: one Element	N lines strengthened: Compound lines & Blends	N lines weakened: One Element	N lines weakened: Compound lines & Blends	Percentage of T.No.: strengthened	Percentage of T.No.: weakened	Percentage of T.No.: Affected
Ca	60	48	16	--	--	98	--	98
Cr	386	200	75	36	31	71	17	88
Co	118	26	25	17	14	43	26	69
H	4	---	--	4	--	--	100	100
Fe	1108	300	127	258	98	39	32	71
Mg	8	3	--	1	--	38	12	50
Mn	167	68	31	15	9	59	14	73
Ni	251	48	24	106	26	29	53	82
Sc	45	30	--	3	--	67	7	74
Si	9	--	--	8	1	--	100	100
Na	8	8	--	--	--	100	---	100
Ti	432	247	73	46	28	74	17	91
V	176	114	37	9	5	86	8	94

The behaviour of the enhanced lines is a very interesting characteristic of the sunspot spectrum. All of the enhanced lines appear weakened. Out of a total of 144 contained in the above mentioned region, 130 are weakened, 14 are not affected and none are strengthened.

II FLOCCULI:

If we obtain a monochromatic photograph of the sun in the light of the H calcium line by giving the spectroheliograph a continuous motion, the spectroheliograph will show the distribution of calcium vapour over the sun's disc. The whole of the sun's surface is found to be covered with small clouds of calcium vapour of about one second of arc in diameter with dark spaces in between, similar to the granulation of the atmosphere. Yet there is a great difference between these vapours and the granulation described in the introduction.

According to Langley's hypothesis the granulation of the solar surface are extremities of columns of vapour which rise from below.

(1) Hand Buch Der Astrophysik Band IV p. 133.

They characterize the regions where the convective currents raise the heated vapour up to a height where the temperature is so reduced as to give rise to condensation. Hale suggested that over-lying these columns are other vapours which are not so easily condensed and which therefore continue to rise, so that the granulation seen in the spectroheliogram might be columns of ascending calcium. Even we may go further and say that the higher and more extensive calcium clouds are just such columns of vapour rising to such a height above the chromosphere as to form the prominences which are actually composed almost of calcium and hydrogen vapours. The calcium vapours seen in the spectroheliogram are known by the name flocculi.

The structure of the flocculi is investigated by photographing them at different levels above the photosphere. The calcium vapours emitted by the sun are comparatively dense at the lowest levels; they expand at higher levels because of decreased pressure and therefore they become less dense. Laboratory results show that very dense calcium vapour produces wide spectral bands which decrease in width with decreasing density, until they become sharp and well defined lines. The H and K lines in the solar spectrum denote the presence of calcium vapour at varying densities. The wide dark bands, which according to Hale's and Deslandere's notation are known as H_1 and K_1, are due to dense calcium vapour at low levels. In between are the two narrower and better defined bright lines H_2 and K_2, and between these two again are dark and still finer lines H_3 and K_3, which are due to double reversal, that is to the absorption of the cooler calcium vapour at higher levels. With the spectroheliograph we are therefore able to photograph calcium vapour at different densities and hence at different levels above the photosphere.

When the flocculi are near regions of active eruptions they are very brilliant because of their high temperatures or of other causes. These very luminous vapours are erupted from the sun with such a great velocity that rapid changes in their shapes are observed. The ordinary flocculi undergo slow changes as a rule, indicating less disturbed conditions. The brilliant eruptive flocculi always appear in active regions of the solar surface especially near sunspots, and are nothing more than eruptive prominences when they are seen on the limb. But in the majority of cases the flocculi seen in the spectroheliograms represent vapours at a comparatively low level, while the prominences which extend above the level of the chromosphere do not show as bright objects projected on the disc.

These results are obtained by taking a monochromatic photograph of the sun through the light of the H or K calcium lines; but lines of other elements could also have been used. Hale and Ellerman first used hydrogen $H\beta$, $H\gamma$ and , $H\delta$ lines; these lines revealed dark hydrogen flocculi instead of bright flocculi. Bright hydrogen flocculi may be seen in disturbed regions and some of these are of eruptive type similar to calcium flocculi, and are seen usually near sunspots' regions.

Spectroheliograms taken with the $H\alpha$ line show the structure of the hydrogen flocculi more delicately and in more detail than that shown by other lines. This is because hydrogen attains its maximum height in the visible radiation. The obtained photographs present different appearances depending upon whether the

whole or only part of the $H\alpha$ line is used. (1) If the photograph is taken through the whole line (1.7 Å) then we get in it all the flocculi whatever their levels may be. But if a photograph is obtained through parts of the $H\alpha$ line, we get then (with the necessary dispersive power) the different levels and different structures, as in the case of the H_1 and H_2 lines for the calcium flocculi.

Hale was the first to not another important characteristic of spectroheliograms taken with the $H\alpha$ line. He noted the cyelonic or vortical appearance especially of the dark flocculi near sunspots. This lead him, as we shall see later, to the discovery of the magnetic field in sunspots. (2) The appearance of the simple vortex indicates a clockwise rotation in the southern hemisphere, counter-clockwise in the northern, of the vapours producing them, assuming that the motion is inward towards the spot.

In certain instances these vortices appear to exert a powerful attraction on the surrounding gases. Thus, for example, ST. John observed a long dark filament near a well defined vortex centered on a sunspot; the nucleus of the spot resolved itself into two, and in a few hours spectroheliograms taken with the $H\alpha$ line showed that the filament had not only extended towards the spot, but that on reaching it had divided into two branches, each of which came in contact with one of the nuclei as if these formed centers of attraction. The mean velocity of motion towards the spot was over 100 kilometers per second. Photographs taken on the following days showed a bright hydrogen ring round the spot; this was probably due to the cool hydrogen, which, after sinking into the spot where it became heated again, returned to the surface escaping from the lower portions of the vortex. Yet this is not a rule, and in many a case no motion towards spots is noticed. This elementary description of some of the characteristics of flocculi is enough for our purpose, and reference for more information may be made to

Hand Buck Der Astrophysik IV pp. 118-127
and Monthly Notices of the A. R. S. 82 334 1922

III <u>Prominences and Eruptions</u>:

When the slit of a spectroscope is directed to a point on the limb of the solar disc where there is a prominence, we observe its spectrum of reversed or emission lines. These lines are more or less intense according as the prominence is more or less bright, and their number depends upon the nature of the prominence. Usually only the more intense lines of the solar spectrum, hydrogen, calcium, and helium, are seen. $H\alpha$ is very conspicous visually, and H and K calcium lines photographically. Sometimes metallic lines appear, and these are very numerous and comprise practically all the lines visible in the chromosphere. This is possible when the observation is made during total eclipses because then with the absenceof light from the photosphere, the background of the sky is dark, and the contrast is enhanced.

---

(1) M.N. 85 464 1925
(2) A. P. J. 28 100 1908

The visual method of observing the prominences in full sunlight is to open the slit of the spectroscope a few seconds of arc so as to obtain, not the spectrum of the prominence, but its image in a given monochromatic light. If we set the slit tangentially to the disc, on the line Hα for example, and on a prominence, then if the slit is a narrow one, as is usual in the observation of the Fraunhofer lines, we see Hα reversed only in the region occupied by the prominence; beyond it the line is dark. The length and position of the bright line in the spectroscopic field corresponds to the size and position of the region of the prominence under investigation. If the prominence is made to take up successive positions we obtain a complete image of the prominence and we can estimate the intensity of the various parts. By opening the slit a few seconds of arc, Hα will appear confused and faint, but the true shape of the luminous prominence will stand out against the dark background of the sky: this is the monochromatic image of the prominence in the light of Hα

Prominences are usually classed as quiescent or eruptive. The quiescent prominences change their shape and position slowly so that they may be watched for several days on end; they consist essentially of hydrogen, calcium, and helium. Eruption prominences change their forms and positions rapidly, and their spectra show numerous metallic lines. These spectral lines also frequently show rapid changes of distortion and displacement due to the motion of vapours along the line of sight. Their form is of many variaties as their height which may rise from a minimum of 30" to 2' and even 3' and in exceptional cases may reach as high as 10', about 35 times the earth's diameter. Prominences may be very narrow at their base, a few tenths of a solar degree, or they may extend over several degrees and in exceptional cases to 20° or 30° , 20 or 30 times the diameter of the earth, or even more. The broad base is usually noted in quiescent prominences, while eruptive prominences are characterized by higher elevations.

A number of eruptive prominences and the rapid changes which they undergo have been photographed in recent years, thus giving us the opportunity of studying their movements and the forces which operate on them.

Bright spots in the flocculi near to sunspots are often shown by calcium spectroheliograms. These bright spots were called b y Hale and Ellerman, eruptions because of their brilliancy and variability, and also because they generally appear in the hydrogen spectroheliograms. Observations by Fox (1) seem to show that these eruptions are really the cause of the eruptive prominences which predominate, as Young observed, near the outer edge of the spots' penumbra. These gaseous eruptions are at times so brilliant and so intense as to stand out, when viewed with the spectroscope, against the background of the solar surface as prominences do on the edge of the disc.

These observations lead us to believe that the birth of a spot is always accompanied, and generally preceded, by one or more eruptions. In the first few hours of the life of a sunspot the eruption may partially or completely cover the spot, and often may precede it in the direction of the sun's rotation. An eruption rarely precedes a single and full grown spot, though it may follow

---

(1) A. P. J. p. 235 (1908)

it on the edge of the penumbra. If the spot is a very active one, eruptions are almost certain to be found on the following edge; these eruptions accompany spots which are in rapid decline, appearing often at the ends of the bridges. According to Fox, one can predict with certainty the birth of a spot, and a well developed spot itself gives rise to new eruptions.

Centres of attraction and repulsion act on the prominences; this is observed at times in the neighbourhood of spots (1), moreover the free ends frequently appear to be driven in a definite direction by horizontal currents, which leads to a belief in a real circulation of the solar atmosphere.

This circulation was hinted at and observed by Secchi, and recent researches by Slocum (2) seem to show that currents exist in the solar atmosphere at about 30,000 km. above the photosphere. These currents tend to draw the prominences towards the poles in mean latitudes between $\pm 20^\circ$ and $\pm 55^\circ$, and towards the equator in high latitudes, about $\pm 65^\circ$, while at the equator practically neutral conditions prevail. Yet Evershed notes that it is difficult to imagine that the solar atmosphere possesses sufficient density to produce such currents at the elevations attained by the prominences on the other hand, prominences are often seen open out like a tree at high elevations, or near to each other and inclined in opposite directions, so that it is difficult to attribute these shapes to atmospheric circulation.

It has been empirically found that the ascending velocity of the constituents of the eruptive prominences increases with their elevation. Pettit's (3) results of his researches on the forces operating on prominences found that, for example in the case of the prominence of 29th May and of 15th July 1919, the motion was uniform during a certain period, after which it increased abruptly as if it had received a sudden impulse, but the motion remains uniform. The mean velocity of the prominences investigated is 153 km/sec., with a maximum velocity of 400 km/sec. in three cases; this seems to be the highest observed photographically or visually. The motion of the prominences is by no means similar to a vertically projected body under gravity. For example in the case of the prominence of 29th May, the initial velocity was 5.5 km/sec. If subject to the force of gravity it would have risen to a hight of 83 km with decreasing velocity and would then have fallen back to the sun. Instead, it continued to rise with uniform velocity to a height of 50,000 km, there a second impulse increased the velocity to 9.2 km/sec. up to a height of 119,000 km, a third impulse raised the velocity to 13.2 km/sec. up to a height of 191,000 km and a final impulse gave it a velocity of 32.1 km/sec. and a height of 230,000 km; at the end of 8 hours the prominence reached its maximum height of about 600,000 km.

It is observed that knots and streamers in the same prominence frequently rush towards sunspots or towards certain centres of attraction with accelerated motion as if drawn by a local force,

---

(1) Slocum A. P. J. 36 265 1912
(2) A. P. J. 33 108 1911
(3) Pub. Yerkes Obs. P IV 1925

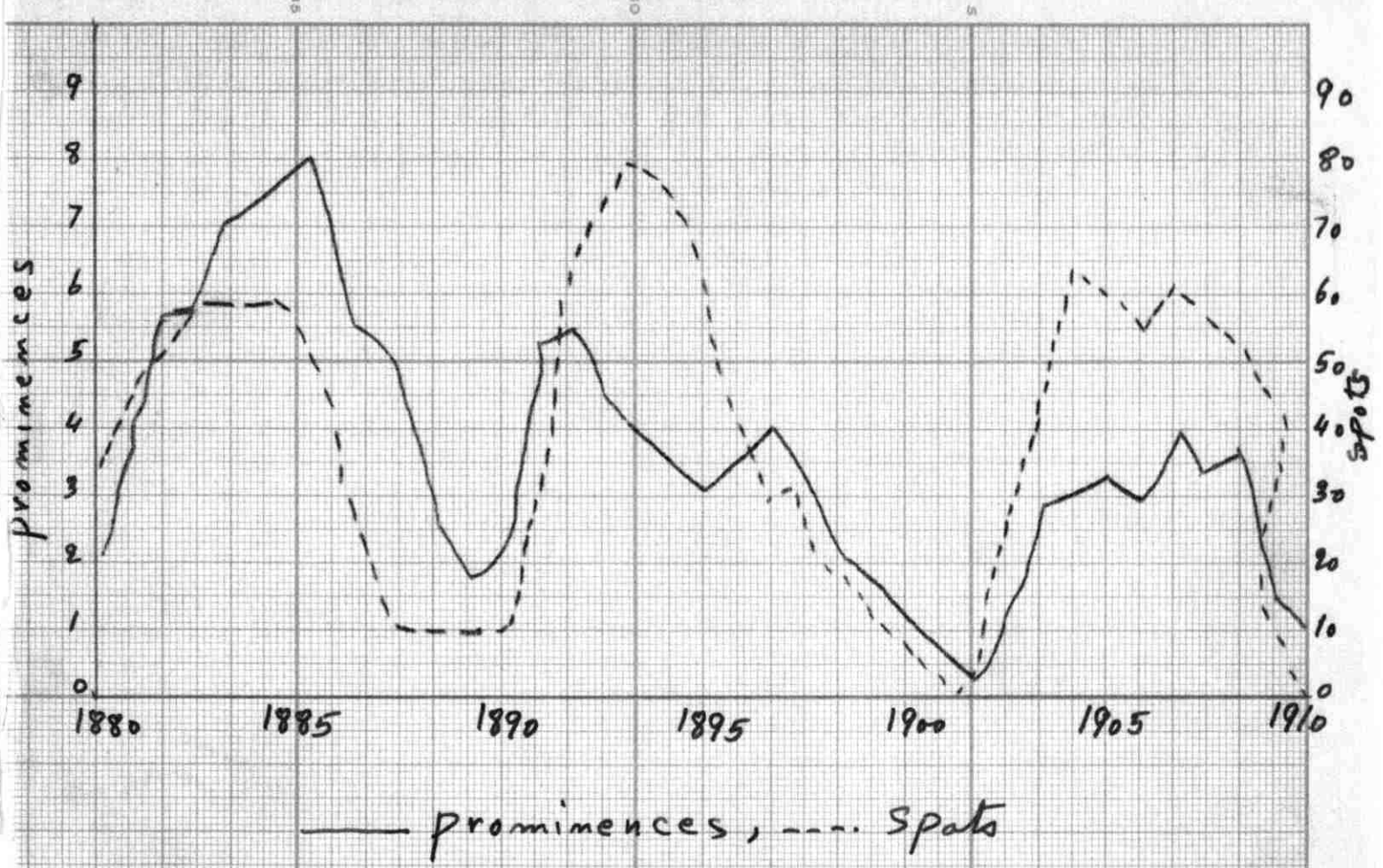

Fig. 6. Frequency of prominences (Ricco) and of spots (wolfer) from 1880 – 1912.

and with a velocity whose vertical conponent is about one third that of gravity. Hence gravity can have but little effect on these phenomena, and the prominences must be under the influence of magnetic and electric forces, as indicated by the down-pouring streamers which rush into the spots and centres of attraction inspite of the high velocity with which the prominence ascends. The matter ejected from the prominences and its changing velocity is probably due to a periodic force which acts for a brief period but with continually encreasing energy. Many explanations have been given to this mysterious phenomin[illegible] of the motion of the prominences and of the motion of the matter ejected from them. The best explanations are those given by Sur and E. A. Milne which depend on radiation pressure theory. These will not be discussed here , but reference may be made to the original works of these scientists.(1)

Statistical investigations into the area, form, and distribution of prominences during various cycles of solar activity has been and is being undertaken, as in the case of sunspots, faculae, and flocculi, with international cooperation. The statistics of the prominences throughout a period of three cycles, from 1880 to 1912 have been collected from the records of Ricco who compared the mean daily number of prominences observed in each year with Wolfer's relative numbers for the sunspots.(2) An examination of the graphs of the two phenomena (figure 6)(3) indicate the similarity of their general activities, which is also manifested in certain details. From the figure one notices that their maxima and minima occur nearly at the same time, and consequently the period of the prominence cycle is nearly the same as that of the sunspots; Also the increase in their activity, as with sunspots, is more rapid from a minimum to the following maximum than the decrease from a maximum to the next minimum. Yet one also notices a lag of the prominence cycle relative to the sunspot cycle, and other differences which indicate a certain amount of independence between the two phenomena.

The distribution in latitude of the prominences in the above mentioned three periods leads to the following conclusions:

1) There exists in both hemispheres a zone of maximum frequency which lies between 20° and 40° heliographic latitude, that is in the sunspot zone, and it continues throughout the 11 year cycle, except near the period of minimum activity.
2) Another zone of maximum frequency lies between latitude 40° and 60° which manifests itself after a maximum and nearly up to the next minimum, it then moves to higher latitudes and reaches the polar regions at the maximum period.
3) Metallic eruptive prominences are formed in the low latitude zones, in other regions the prominences are usually of the quiescent type.
4) The mean heliographic latitudes of the prominences are higher near a minimum period and lower after a maximum.

Also Lockeyer's researches on the relation between the prominences and the corona (4) lead to the following conclusions:

---

(1)a) Sur A. P. J. 63 111 1926
(b) Milne M. N. 86 591 1926
(c) Pike M. N. 88 3 1927
(2) These numbers will be explained in the next chapter.
(3) Hand Buch Der Astrophysik Band IV page 153
(4) M. N. 63 481 1903
M. N. 82 323 1922

(1) That the various coronal forms are clearly connected with those belts of latitude where the centres of action of the solar prominences are found.
(2) That there is no direct relation between sunspots and the production of coronal streamers.

The relation between prominences and sunspots and flocculi is important and so also is the relation between these phenomena and the atmospheric terrestial disturbances. To this we may refer later

CHAPTER V

Some Properties of Sunspots

I Duration, Shape, and Level of Sunspots:

The general appearance of the solar disc has been discussed in some of the previous chapters. It was mentioned then that small pores may join one another and form two large spots. The preceding spot is, as a rule, more compact of the two and has a greater motion in longitude. Then the two spots are connected by a bridge of smaller spots; the bridge then disappears together with the following spot. The leader spot then becomes circular in shape and begins to diminish gradually in size and finally split up into a number of pores from which new spots are frequently reproduced. The life of the spots is very variable, some spots last for a few hours only while others are visible for months. Like their life, the size of sunspots is also very variable. The largest spot recorded may probably be that seen in 1858 which had a breadth of 336,000 km, 18 times the diameter of the earth. During periods of solar activity spots of breadth of 2' has been measured; this is equivalent to a diameter of about 90,000 km. However, such dimensions are very rare, and a sunspot whose diameter, including the penumbra, is 50,000 km is not of common occurence. Spots of about 40,000 km and more are visible to the naked eye and many of them are seen during active periods. The shape of a spot is usually circular, but in larger spots complicated shapes appear and are often composed of smaller spots and pores.

A completely developed spot is composed of two parts: the umbra which is the inner or central dark portion, and the penumbra which surrounds it. The penumbra is usually the distinguishing factor between spots and pores. Granulation around sunspots differs from that away from them; near spots, granulation is very dense and the nuclei are larger and packed so closely near each other that the darker background is often barred and the granulation becomes hardly distinguishable. The comparison between the luminous photosphere and the spot's penumbra is very remarkable; the background of the penumbra is darker and the nuclei are elongated and grouped radially towards the centre of the spot with comparatively large intervals between, and they end as sharply as they begin. In the centre of the penumbra is the uniformly dark umbra, which appears dark only by contrast with other parts of the sun; in fact it is very brilliant and this can easily be confirmed by the transit of one of the inferior planets, Venus or Mercury, across the sun. There is no agreement between the different estimates of the brightness of the umbra as compared with the photosphere. The different estimates differ greatly and vary from one tenth to one fiftieth of the brightness of the photosphere. In large spots a bright bridge is often seen across the umbra.

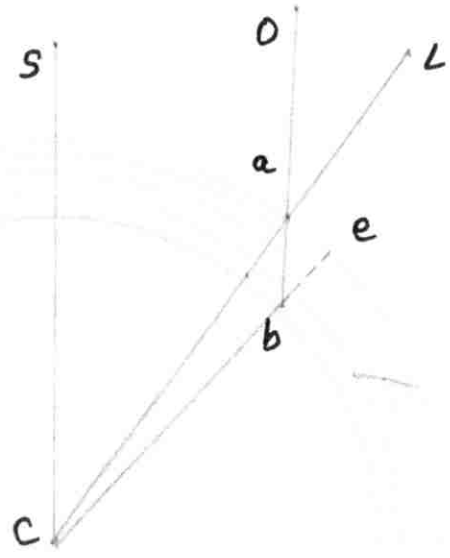

Fig. 7. Determination of the depth of sunspots.

The shape of sunspots may vary in two ways; a real variation due to an internal developement of the spot, and an apparent variation due to the position of the sunspot on the surface of the solar disc.

Sunspots are seen on the surface of a spherical body, and therefore their appearance varies with their position on that surface. It also should be remembered that the spots are not two but three dimensional objects; and therefore the problem is to determine their real form and level in the surface of the photosphere.

A. Wilson of Glasgow was the first to observe in 1774 that when a sunspot is in the centre of the sun's disc, the umbra and penumbra are circular and concentric; but as the spot approaches the limb, its shape changes. The umbra together with the eastern part of the penumbra become smaller, while the western part of the penumbra turned toward the limb also becomes smaller but decreases at a slower rate. When the spot is near the edge of the limb, the umbra disappears completely, and only a fine dark streak of the western part of the penumbra is seen. The whole spot is then reduced to a dark line. Thirteen days later, when the spot reappears on the eastern limb the whole phenomena is repeated but in an inverse order. This lead Wilson to the simple conclusion that sunspots are funnel shaped and are situated at a lower level than the photosphere, that their depth can be measured, and that the change in their shape is due to perspective effect. Figure 7 illustrates this. (1) CS in the figure is the line of sight directed towards the centre of the sun; Oab is the line from the edge of the spot when the eastern part of the penumbra has appeared owing to the motion of the spot towards the limb; CZ is a line joining (a) and the centre of the sun. The angle $S\hat{C}Z$ is known and hence its complementary $b\hat{a}e$. The width ae of the penumbra is first measured when the spot is at the centre of the sun's disc, and assuming that it does not alter its shape in moving towards the limb, the triangle (abe) gives the depth (be) of the spot. Wilson found that the depth was about one third of the earth's radius; this has since been confirmed by Herschel, Secchi, Tacchini, and Chevalier, although many spots were found not to conform to Wilson's theory.

Out of 69 normal spots, 72 were found by De La Rue (in his discussion of the observations made at Kew) to agree with Wilson's theory. After a long series of observations made at Zô-sè by Chevalier lead to the conclusion that generally spots are depressions at various depths of the photosphere, and on the average are at a less depth than that found by Wilson and this depth rarely exceeds 1" or 750 km. If, in the majority of cases, sunspots are real cavities, a depression more or less deep according to the size of the spot, should be observed when the spots are on the limb; this has been actually seen by Cassini in 1719 and by others. However, some spots present an inverse phenomenon and therefore the supposition that the umbra is at a lower level than the photosphere cannot be laid down as a general rule. A hypothesis has been put forward which assumes that both the umbra and the penumbra (though the

---

(1) Hand Buch Der Astrophysik IV 90

umbra is a depression in the penumbra) are at a higher level than the photosphere. To this we shall refer later.

II Distribution and Periodicity of Sunspots:

The distribution of sunspots is limited to two zones lying between 5° and 40° north and south of the equator. They rarely appear on the equator, and have never been seen beyond latitude 45 north or south of the equator. The maximum frequency of sunspots lies between 10° and 20° north and south latitudes, their number and size vary greatly from day to day and from month to month and even from year to year. The regular period of 11 years was first discovered by Schwabe. The number of spots taken over a long period will be fairly regular; the maximum number to be seen is from 25 to 50 per day, but days may pass when no spots are to be seen on the solar surface. A mean period of 11.1 years for the sunspot cycle was found by R, Wolf, when he examined all the sunspots observations since Galileo. He arrived at this period with his relative numbers which are of common use today. The Wolf relative number r is given by the formula

$$r = K(10g + f)$$

where g is the number of groups of spots visible daily
f is the number of spots counted
and K is a coefficient depending upon the instrument used and the observer.
K was taken as unity in the case of Wolf's researches. It is apparent that Wolf assumed a weight of 10 for the groups of spots and one for the number of spots. Wolf's observations and determination have been kept up by A. Wolfer and W. Brunner in collaboration with numerous observatories and under international organization which publishes the relative numbers annually. The relative numbers for every month have been recently published by R. Wolfer in the " Terrestial Magnetism and Atmospheric Electricity " June 1925. Because of their importance in researches on solar physics and on the relation between solar and terrestial phenomena the following table of the yearly relative numbers from 1741 to 1927 is given below. (1)

---

(1) Hand Buck Der Astrophysik IV 95

Table V observed Sunspot Relative Numbers, 1749--1927, by A. Wolfer.

Year	R. N.	Year	R. N.	Year	R. N.
1749	80.9	1799	6.8	1850	66.5
1750	83.4	1800	14.5	51	64.5
51	47.7	1	34.0	52	54.2
52	47.8	2	45.0	53	39.0
53	30.7	3	43.1	54	20.6
54	12.2	4	47.5	55	6.7
55	9.6	5	42.2	56	4.3
56	10.2	6	21.1	57	22.8
57	32.4	7	10.1	58	54.8
58	47.6	8	8.1	59	93.8
59	54.0	9	2.5	1860	95.7
1760	62.9	1810	0.0	61	77.2
61	85.9	11	1.4	62	59.1
62	61.2	12	5.0	63	44.0
63	45.1	13	12.2	64	47.0
64	36.4	14	13.9	65	30.5
65	20.9	15	35.4	66	16.3
66	11.4	16	45.3	67	7.3
67	37.8	17	41.1	68	37.3
68	69.8	18	30.4	69	73.9
69	106.1	19	23.9	1870	139.1
1770	100.8	1820	15.7	71	111.2
71	81.6	21	6.6	72	101.7
72	66.5	22	4.0	73	66.3
73	34.8	23	1.8	74	44.7
74	30.6	24	8.5	75	17.1
75	7.0	25	16.6	76	11.3
76	19.8	26	36.3	77	12.3
77	92.5	27	49.7	78	3.4
78	154.4	28	62.5	79	6.0
79	125.9	29	67.0	1880	32.3
1780	84.8	1830	71.0	81	54.3
81	68.1	31	47.8	82	59.7
82	38.5	32	27.5	83	63.7
83	22.8	33	8.5	84	63.5
84	10.2	34	13.2	85	52.2
85	24.1	35	56.9	86	25.4
86	82.9	36	121.5	87	13.1
87	132.0	37	138.3	88	6.8
88	130.9	38	103.2	89	6.3
89	118.1	39	85.8	1890	7.1
1790	89.9	1840	63.2	91	35.6
91	66.6	41	36.8	92	73.0
92.	60.0	42	24.2	93	84.9
93	46.9	43	10.7	94	78.0
94	41.0	44	15.0	95	611.0
95	21.3	45	40.1	96	41.8
96	16.0	46	61.5	97	26.2
97	6.4	47	98.5	98	26.7
98	4.1	48	124.3	99	12.1
		49	95.9		

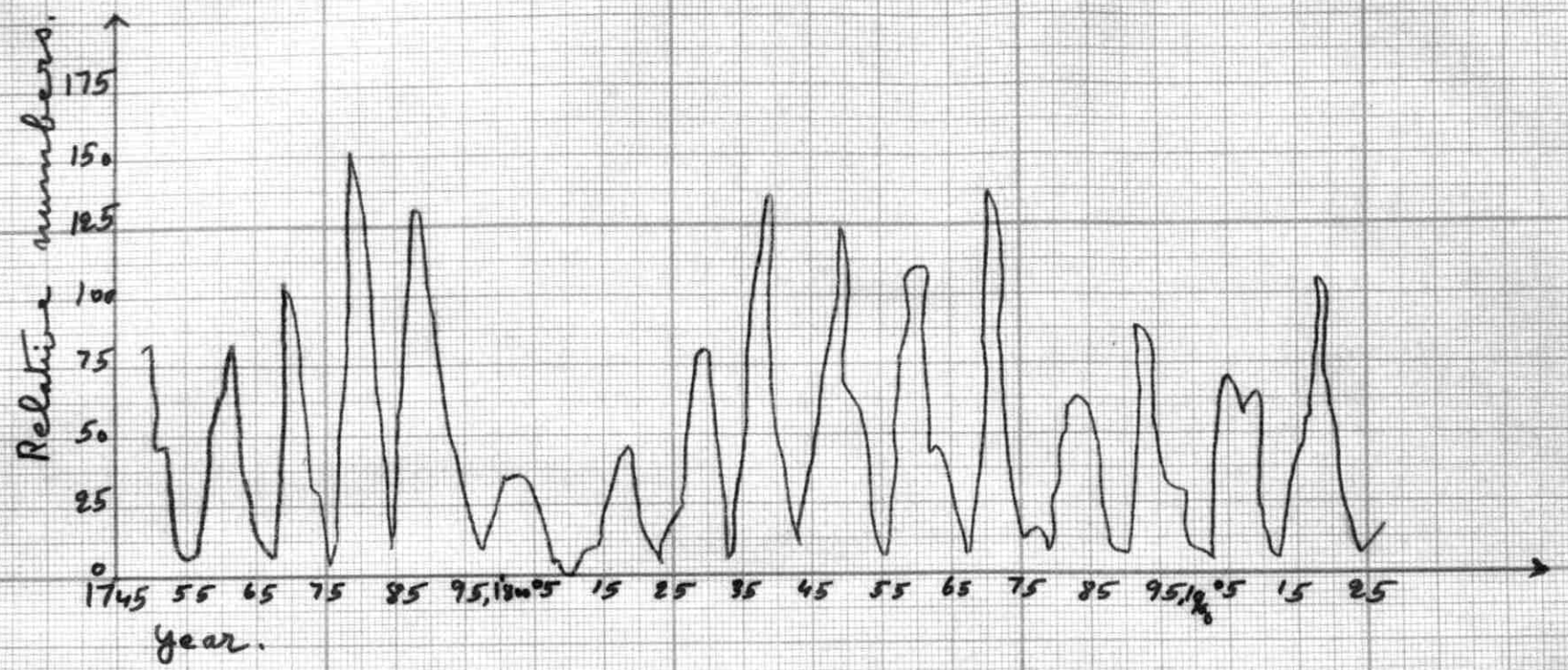

Fig. 8. Cycle of solar activity determined with Wolf's relative numbers.

Table V "Continued"

Year	R. N.	Year	R. N.	Year	R. N.
1900	9.5	1908	48.5	1918	80.6
1	2.7	9	43.9	19	63.6
2	5.0	1910	18.6	1920	37.6
3	24.4	11	5.7	21	26.1
4	42.0	12	3.6	22	14.2
5	63.5	13	1.4	23	5.8
6	53.8	1914	9.6	24	16.7
1907	62.0	15	47.4	25	44.3
		16	57.1	26	63.9
****		1917	103.9	1927	69.0

A more precise determination of the ~~period~~ Period of solar activity is by measuring the total areas covered by the spots (umbra and penumbra) and the faculae. This investigation which was undertaken by Carrington and De La Rue is still continued at Greenwich, where the areas are measured from the daily photographs of the sun submitted by various observatories. The areas are expressed in millionth parts of the visible solar surface. For example, at the maximum of 1859 the daily average was about 1400 millionths, while at the following minimum in 1867 an average of only 200 millionth's of the sun's surface was covered by spots.

It is clear from the curve of solar activity and the relative numbers for the periods investigated, or from the areas covered, that the ascent from a minimum to a maximum is steeper than the descent to minimum, and that the average interval between a maximum and a minimum is 6.6 years while from a minimum to the succeeding maximum the interval is 4.5 years. However, the intervals between two maxima or a minimum and a maximum, or viceversa, are irregular, yet an eleven year period is accepted. An inspection of the curve of figure (8) (1) we find that a period of 11.5 years is not quite satisfactory, while a double period of 23 years with two unequal halves appears to be more probable. This has been confirmed by Hale in his discovery of a period of 23 years for the magnetic period of sunspots.

In addition to the variation of the number and dimensions of the spots during the eleven year cycle, also the position and distribution of the spots on the solar surface vary in this period. Spörer investigated this phenomenon because of its great theoretical importance in connection with the physical constitution of the sun and its relation to the magnetic period of the spots. Two or three years before the disappearing of spots of a certain cycle, a new series of spots are seen to be found in northern and southern latitude of about 30 . Thus at the minimum of a cycle 4 belts of disturbances are observed: two near the equator 8° - 5° north and south where spots are dying out, and two belts in higher latitudes 30° - 40° north and south, where new spots are being formed. The two belts near the equator die out and the other belts near latitude 30° N. and S. drift towards the equator and when they are in latitude about ± 16° , the solar activity is at its maximum. As these two zones continue approaching the equator the number of spots becomes less and less until they almost die out near latitude ± 8° at the end

(1) Hand Buch Der Astrophysik Band IV p. 94

Fig. 9. Spörer's curves of sunspot latitude. (pringsheim, Physik der Sonne.)

of the cycle when a new cycle is observed to be beginning in latitude about $\pm 30^{\circ}$, and so on.

The results of Sporer which have been confirmed by other observations are illustrated in figure (9)[1]. The upper curve of the figure is like the curve of figure 8, with the ordinates corresponding to areas covered by spots instead of relative numbers; the two lower curves represent the heliographic latitudes of the spots counted in the years marked on the abscissa.

Maunder [2] undertook further investigations of Sporer's law and he plotted his results in a figure showing the distribution in latitude of the centres of the spots during those years. Maunder observes that a ripple disturbance on the solar surface, caused by the drifting of the sunspots may indicate a pulsating photosphere, as in the case of the cepheids. A similar pulsation in the external envelopes of the sun is probable.

Further statististical research was done by Turner.[3] Turner's harmonic analysis of the four sunspot cycles between 1876 and 1923 lead to the following differences between the "even" cycles (1888--1899 and 1912--1923) and the odd cycles (1876--1887 and 1900--1911):

(1) The spots, in the "even" cycle, are in the mean one degree further from the equator in both hemispheres, and the spotted area larger in the ratio of 143 to 126 than the "odd" cycle.

(2) The phases of variation in the "even" cycle are earlier by about 5.2 months than the "odd" one, both as regards latitude and area.

(3) And finally the latitude drift indicates a mean period of 11.35 years while the areas give a period of 11.65 years. Thus the mean for the 48 years investigated is 11.5 years, which is the period generally adopted for the frequency of sunspots.

We notice from above that considerable differences are found between the various periods of solar activity, deduced from the time when regular or sufficiently regular sunspots observations were begun. These differences are given in table VI [4] which gives the epochs of sunspots maxima and minima as computed by Wolfer. A study of tables V and VI shows that at the end of the 17th and the beginning of the 18th century, there was a prolonged period of minimum activity, and during this period the northern hemisphere was quiescent, with almost a total absence of spots. The dissymetry in the frequency of spots in the two hemispheres obtains in one direction or another throughout various cycles.

---

(1) Hand Buch Der Astrophysik IV 97
(2) M. N. 64 747 1904 and 82 534 1922
(3) M. N. 85 467--471 1925
(4) Hand Buch Der Astrophysic IV 99.

TABLE VI SUNSPOT MAXIMA AND MINIMA..

MINIMA			MAXIMA		
Epoch	Weight	Period	Epoch	Weight	Period
1610.8	5	8 .2	1615.5	2	10 .5
1619.0	1	15.0	1626.0	5	13.5
1634.0	2	11.0	1639.5	2	9.5
1645.0	5	10.0	1649.0	1	11.0
1655.0	1	11.0	1660.0	1	15.0
1666.0	2	13.5	1675.0	3	10.0
1679.5	2	10.0	1685.0	2	8.0
1689.5	2	8.5	1693.0	1	12.5
1698.0	1	14.0	1705.5	4	12.7
1712.0	3	11.5	1718.2	6	9.3
1723.5	2	10.5	1727.5	4	11.2
1734.0	2	11.0	1738.7	2	11.6
1745.0	2	10.2	1750.3	7	11.2
1755.2	9	11.3	1761.5	7	8.2
1766.5	5	9.0	1769.7	8	8.7
1775.5	7	9.2	1778.4	5	9.7
1784.7	4	13.6	1788.1	4	17.1
1798.3	9	12.3	1805.2	5	11.2
1810.6	8	12.7	1816.4	8	13.5
1823.3	10	10.6	1829.9	10	7.3
1833.9	10	9.6	1837.2	10	10.9
1843.5	10	12.5	1848.1	10	12.0
1856.0	10	11.2	1860.1	10	10.5
1867.2	10	11.7	1870.6	10	13.3
1878.9	10	10.7	1883.9	10	10.2
1889.6	10	12.1	1894.1	10	12.3
1901.7	10	11.9	1906.4	10	11.2
1913.6	10		1917.6	10	

TABLE VI SUNSPOTS MAXIMA AND MINIMA..

MINIMA			MAXIMA		
Epoch	Weight	Period	Epoch	Weight	Period
1610.8	5	8 .2	1615.5	2	10 .5
1619.0	1	15.0	1626.0	5	13.5
1634.0	2	11.0	1639.5	2	9.5
1645.0	5	10.0	1649.0	1	11.0
1655.0	1	11.0	1660.0	1	15.0
1666.0	2	13.5	1675.0	2	10.0
1679.5	2	10.0	1685.0	2	8.0
1689.5	2	8.5	1693.0	1	12.5
1698.0	1	14.0	1705.5	4	12.7
1712.0	3	11.5	1718.2	6	9.3
1723.5	2	10.5	1727.5	4	11.2
1734.0	2	11.0	1738.7	2	11.6
1745.0	2	10.2	1750.3	7	11.2
1755.2	9	11.3	1761.5	7	8.2
1766.5	5	9.0	1769.7	8	8.7
1775.5	7	9.2	1778.4	5	9.7
1784.7	4	13.6	1788.1	4	17.1
1798.3	9	12.3	1805.2	5	11.2
1810.6	8	12.7	1816.4	8	13.5
1823.3	10	10.6	1829.9	10	7.3
1833.9	10	9.6	1837.2	10	10.9
1843.5	10	12.5	1848.1	10	12.0
1856.0	10	11.2	1860.1	10	10.5
1867.2	10	11.7	1870.6	10	13.3
1878.9	10	10.7	1883.9	10	10.2
1889.6	10	12.1	1894.1	10	12.3
1901.7	10	11.9	1906.4	10	11.2
1913.6	10		1917.6	10	

III Motion of Sunspots:

(A) Proper Motions: The general motion of the spot zones in latitude was discussed above. The proper motion of the spots themselves may be in both latitude and longitude or in either of them. This individual motion of the spots may be regular or nonuniform. The systematic proper motion of the long-lived spots has been investigated at Greewich. The results are summarized as follows: (1)

(1) The average random motion in latitude of single spots is very nearly one degree per sidereal rotation; for the centres of groups of spots it is 1°.25.

(2) As regards a systematic latitude drift, single spots, which offer perhaps the most reliable material do not support the existence of any such drift of as much as one-tenth of a degree per rotation.

(3) It is interesting to notice that the general progression towards the equator of the spot zones during the course of a solar cycle (which on the average would amount to 0°.14 per siderial rotation cannot be traced in the motion of individual spots.

(4) The most remarkable feature is the rapid forward movement in longitude of a leader spot at its formation, accompanied by an increase in its area. For the first two days the diurnal motion is + 1°.0 greater than the average movement of a spot in the same latitude. At 15 from the equator this amounts to a motion of 15 .2 in longitude per day. If these differences be correlated with variation in level, a leader spot at its formation is high and rapidly descends, at the same time its area increases. (2)

Thus the motion of spots in latitude is found to be very small and even vanishing, while the systematic motion in longitude is remarkable. The motion of the follower spot is constantly slower than the mean velocity corresponding to its latitude, especially at the times of its formation when the mean motion is 0.3 towards the east. The change in the area of the follower shows a rapid increase to a maximum on the 3rd or fourth day followed by a slow decrease.

Another characteristic of the sunspot groups formed by a pair of separated nuclei (which often resolve themselves into several components and which possess well defined magnetic properties) is that the axis of the group is inclined to the solar equator in a lesser or greater degree depending upon the latitude. The angle of the inclination is found to vary slightly during the life of a group, and that there is a well defined connection between this angle and the latitude of the group. The follower spot appears to be further from the equator than its leader, and the higher the

---

(1) M. N. 85 185 1925 and 85 553 1926

(2) In the 4th result, the explanation given appears to be unsatisfactory. There is no reason why the follower spot should not experience the same change in level. For a probably more satisfactory explanation, depending on Bjerknes theory, we may refer in the following chapters.

latitude, the greater the inclination of the axis to the equator. This holds for both hemispheres. Generally the angle of inclination seems to depend entirely on the latitude of the group, and is independent of the epoch in the cycle, in low latitudes the axes are nearly parallel to the equator; and the inclination increases with latitude up to a maximum of about 11°.

(B) Radial motions in Sunspots:

The hydrogen flocculi overlying the spots usually show a vortical structure. The discovery of the magnetic fields in sunspots lead to the supposition that the incandescent gases which consitute the spots are also endowed with vortical motion. Evershed [1] investigated the radial motion of sunspots in 1909, and his conclusions were as follows:

(1) All spots examined show line-shifts of about the same order of magnitude when at the same distance from the centre of the sun's disc.

(2) The displacements disappear when the spot is within 10° of the centre.

(3) The displacements are most marked when the spot is between 30 and 50 east or west of the central meridian

(4) The displacements are of opposite sign on opposite sides of the central meridian.

(5) The displacements are invariably towards the violet on the preceding side of a spot and towards the red on the following side, when the spot is east of the central meridian, the reverse when west of it.

(6) Southern spots show the same direction of movement as the northern spots.

(7) No displacements are obtained when the slit bisects a spot in a direction at right angles to a line joining the spot and the centre of the sun's surface.

A hypothesis which seemed to Evershed to be in harmony with the facts stated above is that the displacements are due to a radial movement outwards from the spot centre. The motion must be essentially horizontal, or parallel to the sun's surface; this is shown by the total disappearance of the line shifts when the disturbance is at the centre of the disc. Thus the displacements are a Doppler effect. Evershed remarks also that a hypothesis of a vortex, or rotation of any kind, about an axis perpendicular to the sun's surface is untenable because of the conclusion No. 7 above; for it is evident that for a circular movement a nodal point should be found when the slit bisects the spot in a direction at right angles to this. This position of the node, however, differs from that actually found by about 90°.

When the slit centrally bisects a symmetrical spot in a direction approximating that giving the greatest shift, the displaced lines appear quite straight and inclined to the undisturbed lines, the greatest shift occuring at the outer limits of the penumbra. This inclination of lines of metallic vapours in spots suggests an accelerated motion outwards from the umbra. which in Evershed's observations attains a maximum of about

---

(1) M. N. 69 454 1909

2 km/sec at the extreme edge of the penumbra. Yet at the limits of the penumbra the motion apparently ceases abruptly, which cannot occur just after a maximum of motion.

In further researches (1) Evershed found that the high elevations show, when they are well defined, similar displacements but in an opposite direction, that is towards the violet in the region of the spots towards the limb, indicating an inward movement of calcium vapour from the high chromosphere. The amount of displacement is of the same order as that of the outward motion of the underlying gases. Evershed also notes that while the lines of the reversing layer appear perfectly straight, but inclined where they cross a spot, there being an appreciable break or jolt in the lines at the points where they pass from the penumbra to the surrounding photosphere, this is not the case with H and K lines, which generally form regular sinuous curves over spots when the slit lies in the direction of the centre of the sun's surface. With the slit in the opposite direction they, together with the other absorption lines, remain perfectly straight and undisplaced provided that no eruptive disturbance is in progress. An explanation of this behaviour of the Ca and Fe lines is suggested by professor Ritchie Smith. It is that the low-lying metallic vapours in their motion outwards penetrate into or even perhaps beneath the banked-up faculae surrounding the spot, so that there is an apparently sudden stoppage of the motion at the limits of the penumbra. The calcium vapour of the higher region, on the other hand, in its movement inwards towards the umbra, meets no such obstruction, and the motion can therefore be followed from a considerable distance outside the penumbra.

The inclination of the calcium lines crossing the spot would seem to indicate a diminishing velocity twards the umbra, just as the opposite inclination of the other lines indicate an accelerated movement. However this cannot be ascertained until it is proved that there is no vertical motion in the calcium vapour, a phenomenon which will be discussed in a moment.

Evershed also finds that the outflowing gases from the spots' centres sometimes move both radially and with some curvature, that is the motion is spiral. Summarizing his results on this phenominon, it may be said that these vapours have a mean rotational movement of 0.35 km per second, and a radial motion of about 2 km per sec. Combining these two motions we get a spiral movement opposite in the two hemispheres. Yet the apparent radial structure without curvature, often seen in the penumbral filements of symmetrical spots, suggests a purely radial horizontal movement.

ST. John undertook subsequent observations to that of Evershed at Mount Wilson in 1910 and 1911. His conclusions were as follows: (2

(1) The observations are in an entire agreement with Evershed's hypothesis that the displacements considered are due to a movement of the solar vapors tangential to the solar surface and radial to the axis of the spot "vortex".

---

(1) M. N. 70 218 1910
(2) ST. John A. P. J. 38 389 1913
and A. P. J. 37 351 1913

(2) The proportionality between the displacements and wavelengths shows that the phenomenon is due to the Doppler effect; and that what is observed is an actual flow of the material of the reversing layer out of the spots and of the chromospheric material into the spots.

(3) The increase of displacements indicating an outward flow with the decrease in the intensity of the lines of the reversing layer, and the increase of the displacements indicating an inward flow with increasing intensity of the high level lines of the chromosphere find their explanation in differences in level. The outward velocities increase with distance below a neutral level, and the inward velocities increase with distance above this neutral level, or the level of inversion of velocity.

(4) The vertical distribution of the velocities shows at high levels an inflow into spots and at low levels an outflow. A consideration of the quality and the quantity of the material and of the amount of energy involved in the two movements indicates that they are not in themselves a vortex system. This is in agreement with Evershed's note that vortex assumption will not explain his observations. But ST. John went further and concluded that.

(5) The type of vortex indicated is that of the terrestial tornado or hurricane. That is a whirling upward rush of material from the interior of the sun, which spreads out radially with rapid decreasing velocity, tangential to the solar surface, and entrains with it the gases of the reversing layer. The actual vortex is deep-seated, the outflow into the reversing layer being a portion of the upper part of it, and the inflow from the chromosphere being a secondary effect, a superficial indication of the underlying vortex in which the magnetic field originates. This takes away the probable disagreement between Evershed's hypothesis and Hale's vortex theory of the spots and his discovery of the magnetic field in 1908.

(6) The secondary phenomenon shows sometimes vortex motion of the high-level calcium vapour, visible evidences of vorter movements, and stream-line structure of the H_α flocculi, depending upon the rotational energy of the underlying vortex and the strength of the magnetic field of the spot.

(7) The absence of stream-line structure in the H_2 flocculi follows from the quiescent state of the calcium vapour producing these flocculi. This shows no vertical motion, and the flocculi are near the level of zero velocity along the solar surface.

(8) The flux of gases in and below the lower portion of the reversing layer would cause a piling up of material, the temperature of which would be raised by the rapid transformation of mechanical energy into heat, and the increased emission of the calcium vapour would then be a temperature effect.

(9) When the displacements of lines in the red and violet of equal solar intensities are compared, the lines in the violet show the smaller displacements and therefore originate at higher levels, a consequence of the scattering of light.

(10) The displacements of the iron lines arranged in the order of intensities from 00 to 10 form a regular descending scale for determining the relative levels in the solar atmosphere at which

the lines of like intensities of other elements originate.

(11) Assuming as a standard the series of displacements shown by the Fe lines, which decrease regularly from 0.034 Å for intensity 00 to 0.004 Å for intensity 10, the relative levels of the lines of 26 other elements of the reversing layer and chromosphere have been determined and plotted in a chart of distribution.

(12) The enhanced lines show smaller radial displacements than unenhanced lines of the same solar intensities and would appear to originate at higher levels in and near sunspots. This fact is in agreement with the ionisation theory discussed in Chapter I. Ionization is more complete at higher levels because of reduced pressure.

(13) The displacements of the Fraunhofer lines in the penumbrae of sunspots give a means of sounding the solar atmosphere and of assigning relative levels to the sources of the lines.

From the above conclusions we notice that the vapours of the various elements rise to varying elevations in visible quantities, that the lines of any given element have their origin at depths which increase with increasing intensity; that the enhanced lines are at higher levels than the unenhanced lines of the same intensity, and that we can reach greater depths in the sun with the red than with the violet portion of the spectrum.

However during the sunspot maximum between 1926 and 1930, Abetti[1] carried out a fresh series of observations at Arcetri on the motion of the metallic vapours in sunspots. The displacements of the lines, due to the Evershed effect, were referred to the positions of the lines at the centre of the solar disc. Somewhat different results from the previous observations were obtained. This is mainly because the Evershed effect is not constant and regular for all spots. The outward velocities of the metallic vapours are very variable and lie between zero and 6 km per second for radial components. Although tangential components show considerably smaller velocities up to 3 km/sec, yet they can be often measured. Moreover the maximum velocities are not at the outer edge of the penumbra because maximum displacements of the lines from their normal position occurs just between the umbra and penumbra.

So far we did not speak about the magnetic properties of sunspots nor about the vortical motion in the spots themselves. Most of our description hitherto has been of regions overlying spots or surrounding them and of extemnal properties of the sunspots. Now we shall consider other properties of spots concerning their constitution, vortical motion, and magnetic phenomenon.

Magnetic Properties of Sunspots:

It was mentioned in the previous chapter that spectroheliograms in H_α are characterised by the vortical structure of the hydrogen flocculi. The simpler vortices around spots indicate that their rotation is similar to terrestial cyclones, that is counterclockwise in the northern hemisphere and clockwise in the southern.

---

(1) Hand Buch Der Astrophysik VII 371

the lines of like intensities of other elements originate.

(11) Assuming as a standard the series of displacements shown by the Fe lines, which decrease regularly from 0.034 Å for intensity 00 to 0.004 Å for intensity 10, the relative levels of the lines of 86 other elements of the reversing layer and chromosphere have been determined and plotted in a chart of distribution.

(12) The enhanced lines show smaller radial displacements than unenhanced lines of the same solar intensities and would appear to originate at higher levels in and near sunspots. This fact is in agreement with the ionization theory discussed in Chapter I. Ionization is more complete at higher levels because of reduced pressure.

(13) The displacements of the Fraunhofer lines in the penumbrae of sunspots give a means of sounding the solar atmosphere and of assigning relative levels to the sources of the lines.

From the above conclusions we notice that the vapours of the various elements rise to varying elevations in visible quantities, that the lines of any given element have their origin at depths which increase with increasing intensity; that the enhanced lines are at higher levels than the unenhanced lines of the same intensity, and that we can reach greater depths in the sun with the red than with the violet portion of the spectrum.

However during the sunspot maximum between 1926 and 1930, Abetti[1] carried out a fresh series of observations at Arcetri on the motion of the metallic vapours in sunspots. The displacements of the lines, due to the Evershed effect, were referred to the positions of the lines at the centre of the solar disc. Somewhat different results from the previous observations were obtained. This is mainly because the Evershed effect is not constant and regular for all spots. The outward velocities of the metallic vapours are very variable and lie between zero and 6 km per second for radial components. Although tangential components show considerably smaller velocities up to 3 km/sec, yet they can be often measured. Moreover the maximum velocities are not at the outer edge of the penumbra because maximum displacements of the lines from their normal position occurs just between the umbra and penumbra.

So far we did not speak about the magnetic properties of sunspots nor about the vortical motion in the spots themselves. Most of our description hitherto has been of regions overlying spots or surrounding them and of extemnal properties of the sunspots. Now we shall consider other properties of spots concerning their constitution, vortical motion, and magnetic phenomenon.

Magnetic Properties of Sunspots:

It was mentioned in the previous chapter that spectroheliograms in Hα are characterised by the vortical structure of the hydrogen flocculi. The simpler vortices around spots indicate that their rotation is similar to terrestial cyclones, that is counterclockwise in the northern hemisphere and clockwise in the southern.

---

(1) Hand Buch Der Astrophysik VII 371

It is natural, therefore to assume that the rotation of the sun is the determining factor in the direction of motion of the solar vortices. But before going any further one should remember that the conditions in the solar atmosphere are quite different from those of the terrestial atmosphere; the extremely high temperature of the solar surface, the force of radiation pressure, and the direction from which the solar atmosphere receives its radiation, together with other factors, make the solar phenomena much more complex than those of the earth, and one should be very careful in his analogies, For example, under certain conditions the direction of motion of a vortex in the northern hemisphere or in the southern one may be quite different from that which is expected. Some cases have been recorded of spots in the same hemisphere very near to each other, with vortices revolving in different directions.

Solar vortices exert attraction on the surrounding gases; this is evident from the motion of dark flocculi which have been observed to be drawn occasionally towards the centre of a spot. But as a general rule, there is almost always vortical motion in the hydrogen flocculi above the spots. This suggested to Hale'(1) in 1908, the hypothesis that a sunspot consists of a vortex whose particles, electrified by ionization, are whirled at a high velocity. If we assume a preponderance of positive and negative charges ~~in~~ in the rapidly whirling vapours, we must admit a resulting magnetic field above the spots considered as electrical vortices.(2) Therefore here we are confronted by two whirls: a high-level hydrogen vortex, centering in sunspots, where the motion is spirally inwards and downwards; and a low-level electric vortex, formed in the photosphere which constitute the sunspot itself, here the motion is spirally upward and outward. It is easy to show by laboratory experiments that a primary vortex formed in water may give rise to a secondary vortex in a gaseous atmosphere above it, closely analogous to the hydrogen vortices above sunspots.

If the spots are electric vortices and if they should have, as a consequence, a magnetic field, then as in the case of the sun's general megnetic field, the Zeeman effect will be a great help in detecting it. This was done again by Hale and in 1919 Hale was able to make certain that each sunspot possesses a magnetic field.(3)

The tendency of spots formation toward the bipolar structure is so strongly marked that hardly more than 10 percent of all spots observed are wholly free from it. Thus the two major spots, the preceding and the follower are found to be of opposite polarity. Minor spots of opposite polarity sometimes occur in between or very near to these two major spots. In the case of spots which are apparently single, some traces of asymmetry, suggesting a bipolar structure of the spot, can usually be detected. Sometimes such evidence of asymmetry is afforded by faculae following or preceding the spot. It is usually found that a single spot, or a group of small spots all having the same magnetic polarity, is near the preceding end of a mass of calcium flocculi elongated in a direction not greatly inclined to the solar equator. In rare cases the spot occur near the following end of such a group of flocculi. In about 10 per cent of all cases investigated before 1919, the distribution of the flocculi is fairly symmetrical to the east and west of single spots. Thus the information given by the distribution of flocculi helps in the magnetic classification of sunspots. The scheme followed by Hale and his collaborators in classifying sunspots is based primarily upon the determination of their

of their magnetic polarities. Supplementary evidence is frequently needed, and this is supplied by the calcium and hydrogen spectroheliograms which show the distribution of flocculi near sunspots. In this magnetic classification of spots[4] three classes of spots are included in the scheme: (α) unipolar spots, (β) bipolar, and (γ) multipolar spots. These classes may be subdivided as follows:

(α) Unipolar spots:-- Single spots, or groups of small spots having the same magnetic polarity. In this class the distribution of flocculi may vary and thus we have the following subdivisions:

(α) Those in which the distribution of the calcium flocculi is fairly symmetrical preceding and following the centre of the group.

(α_p) those in which the centre of the spot-group precedes the centre of the surrounding calcium flocculi.

(α_f) Those in which the centre of the spot-group follows the centre of the surrounding calcium flocculi.

(β) Bipolar Spots: The simplest and most characteristics bipolar spots consists of two spots of opposite polarity. The line joining the two spots generally makes only a small angle with the solar equator. Each member of the group may be accompanied or replaced by many small spots, but the great majority of the spots constituting the preceding and following members of the group are of opposite polarity. Bipolar spots may be divided into four subdivisions as follows:

(β) Those in which the leading and following members, whether single or multiple are approximately equal in area.

(β_p) Those in which the leading member is the principal member of the group

(β_f) Those in which the following member is the principal member of the group.

($\beta\gamma$) Those in which the preceding or following members are accompanied by minor companions of opposite polarity.

(γ) Multipolar Spots: Groups of this character, comprising hardly more than 1 per cent of the total number of spots observed, contain spots of both polarities so irregularly distributed as to prevent classification as bipolar groups:

The bipolar class is the predominating type of the different classes of sunspots. Previous to the minimum of 1912 (cycle 1901--1912) the polarity of the preceding spots in the northern hemisphere was south or negative, and that of the following spots north or positive. In the southern hemisphere the polarity was the reverse of that of the northern hemisphere during the same cycle. We have seen before that the last spots of the preceding cycle appear in low latitudes, while the first spots of a new cycle appear in high latitudes and as the cycle progresses their mean latitude decreases. The spots of the new cycle (1912--1923 were of opposite polarity to those in low latitudes belonging to the preceding cycle. As the new cycle advanced, the spots became more numerous and the polarity did not change (except about 4 per cent) . The mean latitude of the spots gradually decreased, as expected, and towards the end of the cycle in 1922--1923 spots of another new cycle (1922--1933) with the polarity of the new spots

---

(1) Hale A. P. J. 28 100 and 315 1918
(2) To this hypothesis we shall refer later.
(3) Hale, Ellerman, Joy, and Nicholson A. P. J. 49 153 1919
(4) A. P. J. 49 170 1919

reversed. Thus the existence, at the time of sunspot minimum, of two temporary belts in each hemisphere, containing spots of opposite polarity is certain; and also the reversal of these polarities, after one cycle is over, is well established. Thus a definite law of sunspot polarity is empirically found. This law may be put in a different expression as follows:

The preceding spots, in a cycle of 11.5 years in one hemisphere, have the same polarity while the following spots have opposite polarity to the preceding spots. In the other hemisphere the same holds with a reversal of polarity. After 11.5 years the spots in the two hemispheres interchange polarity so that the magnetic period of sunspots is 23 years approximately.

V Temperature of Sunspots: It is necessary to add here a very important property of sunspots without which they perhaps would have been still undiscovered. This property which was hinted at before is the drop in temperature in spots, a drop which causes them to be seen as apparently dark spots in comparison with the bright photosphere. Most of the determinations of the temperature of sunspots agree that the difference between the temperature of the photosphere and the spots is about 2000 degrees centigrade. The temperature found from the sunspots spectral energy curve by Pettit and Nicholson[1] was 4860°K.

Assuming that the low temperatures of sunspots are due to the cooling of gases by expansion in the upper part of the vortex, Russell[2] estimates the degree of expansion necessary to produce a difference in temperature of the order 2000°—2500°K. Russell finds that the base of a spot vortex must be a region of very high temperature probably over 20,000°K. The increase of volume must be large, probably thirty times. [3]

Thus we notice that the sunspots phenomenon is rather complicated. The strange behaviour of the spots in their distribution, their magnetic field, their periodicity, and their temperature, demands an explanation. Different assumptions and theories were presented for the purpose, and the aim of the next part is to discuss the most important theories in brief and then show how the hydrodynamic theory of Bjerknes stands as the best explanation given until today.

---

(1) Mt. Wilson Contribution No. 397 (1930)
(2) A. P. J. 54 293-5 (1921)
(3) Russell's calculation will be given in the next part of this report.

PART III

THE HYDRODYNAMIC THEORY

OF

SUNSPOTS

CHAPTER VI

Introduction to the Bjerknes

Theory

The aim of this introduction is two-fold; first to show how the hydrodynamic theory became more acceptable than the electromagnetic hypothesis, and second to give an elementary discussion of certain hydrodynamic equations necessary for the theory of Bjerknes

I. The Electromagnetic Theory and Its Test:

We have seen in the previous chapters the large amount of accumulated data about spots. The empirical results concerning the periodicity, distribution and magnetic properties of the spots were well established inductively as a result of a long period of observation and research. These data required an explanation. But the difficulties involved in such an explanation are tremendous: our knowledge of the constitution of the sun is very meagre, and we can only investigate a small thickness of the solar globe. However, two different hypothesis were advanced for the explanation, which proved to be more practical than the other explanations. These were the electromagnetic and the hydrodynamic theories. The Bjerknes theorem will be discussed in more detail later.

It was mentioned before that the spots are an electromagnetic phenomenon due to an electric vortex in which charged particles moving in the solar atmosphere are constrained by the magnetic fields in the spots to follow their lines of force. The principles of the electromagnetic theory were applied to the explanation of sunspots by Störmer.(1)

Störmer begins from the hypothesis that the magnetic field in a sunspot is due to electrically charged particles moving along a logarithmic spiral at low levels around the centre of the spot. After a short mathematical calculation of the magnetic field produced by such a motion Störmer finds that the lines of force of the magnetic field due to the spiral whirl are curves in space with their projections on the plane of the whirl as logarithmic spirals cutting the current-lines of the whirl at right angles everywhere. Thus Stormer suggests that the visible whirls concluded from the structure of the flocculi are, like auroras, not real current-lines but lines of magnetic force due to a magnetic field at a lower level. The low level vortex tends to be more and more circular rather than spiral as the structure of the overlying flocculi tends to become more and more radial. Therefore, the direction of the apparent whirls shown by the hydrogen flocculi, that is the curvature of the projected lines of force, is dependent upon the sign, motion, and

---

(1) A. P. J. 43 347 1916

direction of the invisible under-lying electric whirl which surrounds the spot. By applying his theory to the observed solar phenomena, Stormer concludes the same as professor Hale that the Zeeman effect discovered in sunspot spectra is due to a whirl of negatively charged electric particles.

Hale[1] then tests this theory further and concludes that it does not esplain the facts as they are and that this hypothesis is a working one but does not agree with some magnetic properties of the spots. Hales test is summarized below.

On the assumption that the Zeeman effect is due to a negatively charged particles in a whirl we have five possible cases:

(1) If the flow is circular around an axis passing through the centre of the sun, the lines of force will be plane curves with their planes intersecting in the axis of rotation. The projections of these lines on the solar surface will be straight lines all radiating from the centre of the spot outwards and the hydrogen flocculi should have therefore a radial structure if the spot is seen near the centre of the sun's disc.

(2) If the flow is spirally outward in the clockwise direction, the projected lines of force will show a clockwise curvature and the north pole of the spot will be upward. The clockwise curvature used here is such that if a point moves inward along the lines of force, the rotational component of its motion is in a clockwise direction.

(3) If the flow is spirally outward and counter-clockwise, then the curvature of the projected lines of force will be counter-clockwise and the south pole of the spot will be upward. The counter-clockwise curvature used here is such that if a point moves inward along the lines of force, the rotational component of the inward motion is counter-clockwise.

(4) If the flow is spirally inward and counter-clockwise, the curvature of the projected lines of force produced by such a whirl will be clockwise

(5) If the flow is spirally inward and clockwise, the curvature of the projected lines of force will be counter-clockwise and the north pole of the spot upward.

The study of the Zeeman effect in sunspots carried out by Hale and collaborators shows that the angle between the lines of force and the solar surface, which is about 20° at the centre of the spot, decreases gradually to about Zero degrees near the outer edge of the penumbra. Beyond the penumbra the field becomes too weak to be detected by the method used by Hale. Thus the observed field is confined to the region within the penumbra and an extremely narrow zone surrounding it, while the field calculated by Störmer's theory is much larger in extent. Also the photographs of the hydrogen vortices taken by Hale show a practically radial structure near the spot, even when the curvature is very marked at greater distances. In the shallower layer of the chromosphere, to which the spiral structure studies in that investigation are confined, the projections of the lines of force due to a circular current would be practically straight radial lines. Even if the curved lines in vertical planes

---

(1) Proceedings of the National Academy of Science of the U. S. A. Volume 11 November 1925 page 690

extending to higher levels were in question, it is diffecult to see how their projections could simulate the observed vortex structure or avoid betrayal by their changes in form as they are carried across the disk by the solar rotaion.

If, in spite of the limited extent of the magnetic field and the radial structure of the hydrogen flocculi near spots, the electromagnetic theory could be modified so as to explain the observed vortices, the direction of curvature of the lines of force must nevertheless conform with the polarity of the spots in question if two assumptions are permissible:

(1) The sign of the charge of the particles causing the magnetic field is invariable in all single spots and in the preceding spot of bipolar groups

(2) The direction of any radial component, whether inward or outward, is invariable also in thesame way as the sign of the charge is.

The data from which the sunspot polarity law was deduced indicate this invariable character of the charge and that the direction of the low-level whirls is reversed at every sunspot minimum. Assuming these two assumptions, which are legitimate, Hale begins to apply the polarity criterion making use of the records of polarities of spots at Mt. Wilson. For this test, Hale selected at randum 51 of the best hydrogen spectroheliograms made on various dates from 1908 to 1924, that is including spots belonging to three successive 11 1/2-year cycles covering two complete reversals of magnetic polarity.

The test showed that there is no relation between polarity and direction of whirl and that after the reversal of the magnetic polarity of the spots at two successive minima there was no corrosponding reversal in the direction of whirls in the associated hydrogen vortices This therefore, does not support the electromagnetic theory. Hale found also that 81 per cent of the northern vortices and 84 per cent of the southern vortices irrespective of the 11 1/2-year cycle in which they occur, agree in their direction of whirl with the analogous terrestial phenomenon of tornados. This agreements suggests that the hydrogen vortices are hydrodynamical phenomena and that their direction of whirl is generally determined, not by the direction of whirl of the sunspot vortices below them, but by the eastward and westward deflection, resulting from the solar rotation of currents flowing northward and southward in the solar atmosphere toward centres of attraction above sunspots. This hypothesis is strengthened by the above mentioned failure of the electromagnetic theory and by the motions in the solar atmosphere above spots observed by Evershed and St. John. However, certain exceptions were observed in Hale's test, but these could be explained by the following causes acting singly or in conjunntion .

(1) The small relative drift resulting from the slow change with latitude of the linear velocity of the atmosphere caused by the sun's rotation.

(2) The turbulence of the solar atmosphere, especially near active spots, which may prevent the formation of an induced vortex, mask its structure or determine its direction of whirl.

(3) The fact that about three-quarters of all sunspots, at the time of their formation, are either unipolar or bipolar groups in which the preceding member is larger than the following member. In such cases, according to the hydrodynamical hypothesis, the direction of whirl

of the hydrogen vortex induced over unipolar spots or over the preceding member of bipolar groups should, in general, follow the law of terrestial cyclones. However, in cases where the following spot is larger than the preceding spot at the time of formation, the direction of the hydrogen whirl above the following spot should, in general, conform with the terrestial law, while the preceding whirl should be of opposite sign if the influence of the following whirl outweighs that of the rotational drifts. In bipolar groups where both spots are equal at the time of formation, each spot has an equal chance of starting a whirl following the terrestial law, and local conditions may determine the outcome.

The above test was made by Hale one year before Bjerknes published his theory of sunspots. Bjerknes theory investigates the possibility of such a hydrodynamical hypothesis that would explain the observed properties of spots. Before entering into a full discussion of this famous theory, it seems necessary to give some preliminary derivations of hydrodynamical equations which are important in the understanding of Bjerknes theory.

II Elements of the Hydrodynamics of Circulation. (1)

In the following derivations, the most elementary possible methods were used and in certain cases we begin from the general dynamic principles instead of the hydrodynamic equations of motion; but in certain cases vector notation may be used.

A. Circulation;

Consider a continuous chain of fluid particles forming a closed curve. Each of these particles has a definite velocity $\vec{u}$, and the component of this velocity, tangent to the curve is u_t. By the summation of these tangential velocities along the curve we get

$$C = \int u_t \, ds = \int \vec{u} \cdot ds \qquad 127$$

The quantity C is called the circulation of the fluid along the curve s, where ds is a line element of it. A simple property of the integral of the form (127) will be required in the derivation of other theorems and in all practical applications. Consider a series of curves 1,2,3,---n adjacent to each other, as in fig (10), and let c_1, c_2, c_3 ---c_m, be the corresponding values of the line integrals of equation (127). Take the sum of all these line integrals assuming the same positive direction of circulation for each of the curves, then, as it can be seen from the figure, the line integrals along every part of the curve that has two paths in common, cancel, for in the summation the corresponding line integrals enter once positively and once negatively. The result of the summation is therefore simply equal to the line integral, C, along the outer boundary, that is.

$$C = c_1 + c_2 + c_3 + \cdots + c_m \qquad 128$$

or in other words "the sum of the line integrals along a series of adjacent curves is equal to the line integral along the common exterior boundary."

---

(1) "The Dynamic Principles of Circulatory Movements in The atmosphere" by V. Bjerknes.
and "Hydrodynamics" by Lamb
and different texts on "Vector Analysis"

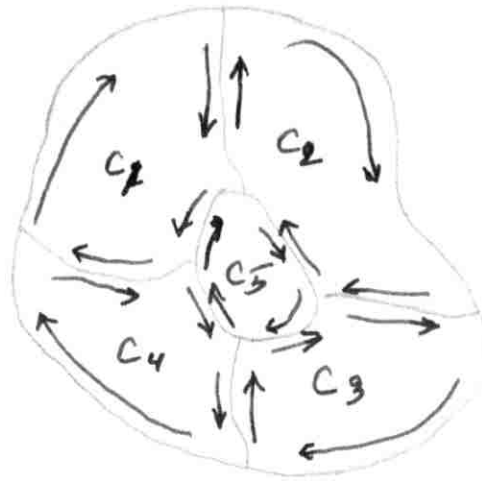

Figure 10

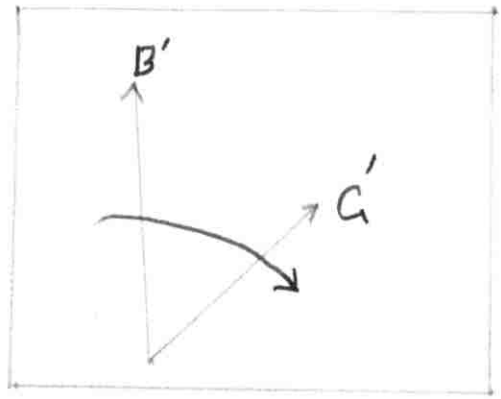

Figure 11

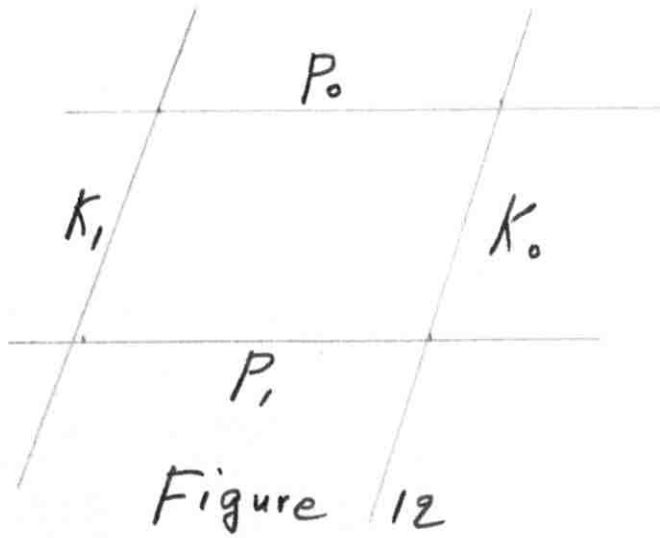

Figure 12

Now we shall consider the law according to which the circulation changes with the time under any given dynamic conditions. Before we deal with this problem it is necessary to investigate the mathematical expression for the change of circulation with time.

Let dx, dy, dz be the projections upon the x, y, z axes of of the linear element of the curve, ds, and let $\vec{U}$x, Uy, Uz be the projections upon the same axes of the velocity $\vec{U}$ of the point on the curve represented by (x, y, z), then the line integral becomes

$$c = \int (u_x dx + u_y dy + u_z dz) \qquad (129)$$

The curve is in motion, and, therefore, not only the velocities change but also the projections dx, dy, and dz of ds, with the time. Therefore differentiating (129) with respect to the time we get

$$\frac{dc}{dt} = \left\{ \int \left(\frac{du_x}{dt} dx + \frac{du_y}{dt} dy + \frac{du_z}{dt} dz \right) + \int \left(u_x \frac{d}{dt} dx + u_y \frac{d}{dt} dy + u_z \frac{d}{dt} dz \right) \right\} \qquad (130)$$

Since the differentiation with respect to the time and the division of the curve into linear elements are independent operations we can, therefore, interchange the order of these two operations in the second integral of (130) and we get

$$\frac{dc}{dt} = \text{the first integral} + \int \left(u_x d\frac{dx}{dt} + u_y d\frac{dy}{dt} + u_z d\frac{dz}{dt} \right)$$

but $\frac{dx}{dt} = u_x$, etc. and therefore we have

$$\frac{dc}{dt} = \text{the first integral} + \int (u_x du_x + u_y du_y + u_z du_z)$$

or $$\frac{dc}{dt} = \text{the first integral} + \int 1/2\, d(u^2_x + u'_y + u'_z)$$

but $d(u^2_x + u'_y + u'_z)$ is the integral of a total differential and vanishes when taken along a closed curve. Therefore we have for $\frac{dc}{dt}$ the first integral of (130) only. Putting Ax for $\frac{du_x}{dt}$, etc. we get therefore

$$\frac{dc}{dt} = \int (A_x dx + A_y dy + A_z dz) \qquad (131)$$

that is to say if At is the acceleration along the tangent of the curve we have

$$\frac{dc}{dt} = \int A_t ds \qquad (132)$$

$$= \int \vec{A} \cdot \vec{ds}$$

or in other words "the increase of the circulation of a closed curve

in a unit of time is equal to the integral, taken along the curve, of that component of accelaration that is tangential to the curve." Thus in finding the dynamic law of the change of circulation with the time we need only to integrate the component accelerations due to the individual active forces in the direction tangential to the curve. Therefore, all accelerating forces that have a line integral equal to zero along closed curves are important. This leads to a very important simplification of our problem, for it is well known that all accelerating forces of a conservative nature have this property. Therefore the force of gravity which is conservative need not be taken into consideration in calculating the circulation along closed curves. If also, we omit the consideration of friction and the deflecting force of the earth's rotation, then we shall only have to consider the accelerating force resulting from the pressure of the fluid. The line integral of this force will be easily determined after we consider a geometrical representation of the dynamic conditions in the interior of gaseous or fluid media.

(B) Geometric representation of the Dynamic conditions in Liquid or Gaseous media:--

Draw isobaaic surfaces for unit differences in pressurethen then the pressure gradient is given by

$$\vec{G} = -\frac{d\vec{p}}{dn} \qquad (133)$$

where $\vec{n}$ is the normal to the isobasic surface, P is the pressure, and $\vec{G}$ is a vector quantity pointing towards the decreasing pressure while $\vec{n}$ points towards increasing pressure. The acceleration that the pressure gradient gives to a particle of fluid depends on the inertia, that is to say, on the density of the particle; it is equal to the pressure gradient divided by the density, or still simpler, it is equal to the gradient multiplied by the specific volume K of the fluid particle. Therefore to be able to express the distribution of pressure and at the same time that of the specific volume throughout the fluid we draw isostesic surfaces for unit differences of the specific volume. Then we have

$$\vec{B} = \frac{d\vec{k}}{dn} \qquad (134)$$

where k is the specific volume and n the normal to an isostesic surface taken positively in the direction of increasing specific volume. Therefore $\vec{B}$, the "reciprocal" inertia gradient is a vector quantity that points in the direction of increasing specific volumes and since the mobility of the fluid increases with the specific volume, Bjerknes calls $\vec{B}$ the vector of motion. It is noted that in equation (134) we used the positive sign, whereas in (133) the negative sign occurs. A vector quantity, $-\vec{B}$, defined in complete analogy with equation (133) would in general have a direction almost exactly opposite to the direction of the pressure gradient, since with diminishing pressure an increasing specific volume usually follows. On the other hand, the vector of motion B, has

approximately the same direction as that of the pressure gradient, $\vec{G}$, and is therefore to be preferred to $(-\vec{B})$ in the application. A few general remarks as to the properties of the isobaric and isosteric surfaces are important:

(1) An isobaric surface can never come to an end in the interior of a fluid, it must either reenter into itself or else end at the boundary surfaces of the fluid.

(2) Two neighbouring surfaces, representing different values of the pressure P, can never intersect each other throughout their whole course they must be separated from each other by an isobaric layer, which on its part has the same fundamental property as the surfaces, namely either returning into itself or ending in the boundary surfaces of the fluid. Similarly the successive isosteric surfaces, are separated from each other by corresponding isosteric layers.

These two sets of surfaces together divide the whole space into tubular or prismatic portions, which are designated by Bjerknes as isobar0-isosteric tubes. From the properties of the isobaric and the isosteric layers that belong to these tubes, it follows that also each of these tubes runs into itself or terminates at the boundary surfaces of the fluid. If the surfaces are drawn for unit differences of pressure and specific volume then they are called unit tubes. If we assume that we use the units just mentioned of proper dimensions, then we may consider the corresponding unit tubes as infinitesimal solenoids. The cross-section of the larger isobaro-isosteric tubes have the form of curved quadrilaterals, the cross sections of the solenoids are rectilinear parallelograms . Since the solenoids have this property that they either return into themselves or terminate at the boundary surfaces, therefore, every closed curve in the fluid incloses a definite bundle of solenoids; the number, N, of solenoids in this bundle becomes a simple definite number as soon as the units of specific volume and of pressure have been chosen.

) Deduction of the Fundamental Dynamic Theorem Relative to the Circulation

Consider a portion of a fluid so small that within it we may consider the specific volume and the pressure as linear variable quantities; In this portion of the fluid the isobaric surfaces extend as a set of parallel equidistant planes, and the isosteric surfaces are also another set of parallel equidistant planes. The solenoids then are tubes whose cross-sections are a set of parallelograms congruent to each other. The vector of motion, as well as the pressure gradient, will have throughout this part of the fluid an invariable magnitude and direction. If the specific volume is constant for all the particles of the fluid under consideration, then the pressure gradient would give equal accelerations to all points and the result of the effect of the pressure gradient during an element of time would remain a simple pure motion of translation superposed upon the previous velocity of this part of the fluid. But since the specific volume varies from one point to another in this fluid except for points on the same isosteric surface, then the different points will take up different accelerations in such a way that the lighter portions will move faster than the heavier portions. Therefore in this case the pressure gradient produces not only a translatory but also a rotatory motion, which causes the fluid masses to turn around the intersections of the isobaric and isosteric surfaces as axes, and in the direction from the vector of motion, $\vec{B}$, by the shortest way to the pressure gradient $\vec{G}$. By reason of this rotation of the fluid masses, there results a

circulation of all closed curves consisting of particles of fluid. We need consider only plane curves within the small portion of the fluid under consideration. The following rule will determine the direction of the acceleration of circulation that one of these curves experiences.

"Project the pressure gradient and the vector of motion on the plane of the curve; then the acceleration of circulation is directed by the shortest route from the projection, B' of the vector of motion toward the projection, G', of the gradient." This is apparent in figure (11).

In order to find the qualitative law for the resulting acceleration of circulation use is made of equation (132) according to which the increase per unit of time in the circulation is proportional to the line integral of the component of the acceleration that is tangential to the curve. We shall first determine the value of this line integral of the acceleration for the curve produced by the intersection of an isobaro-isosteric tube with any arbitrary plane. This curve has a form of a parallelogram (fig. 12) two of whose parallel sides, P_0 and P_1, lie in an isobaric plane, and the other two, K_0 and K_1, lie in an isosteric plane. If h is the distance of the two isobaric planes from each other, then the gradient has the numerical value

$$G = \frac{P_0 - P_1}{h} \qquad (135)$$

Since the pressure gradient is perpendicular to the two isobaric sides of the parallelogram, it can cause no acceleration in a direction tangential to these lines. But the pressure gradient forms an angle, θ with the isosteric sides of the parallelogram and consequently produces, in a direction parallel to these lines, the component accelerations $K_1 G \cos\theta$ and $K_0 G \cos\theta$. If we refer both of these to the same direction of circulation around the curve $P_0 K_0 P_1 K_1$, then they become

$$K_1 G \cos\theta \quad \text{and} \quad -K_0 G \cos\theta .$$

In order to find the value of the line integral we have to multiply these quantities by the length of the corresponding line elements and add the products thus formed. But both sides of the parallelograms have the same length $h \sec\theta$, so that we find $(K_1 - K_0)Gh$ as the value of the line integral. If we introduce the value of G from (135) in this line integral it becomes equal to

$$(K_1 - K_0)(P_1 - P_0)$$

Finally we may specialize by the assumption that the isobaro-isosteric tube under consideration is a solenoid. According to the definition of the solenoid we have $K_1 - K_0 = 1$, and $P_1 - P_0 = 1$, and hence the line integral is equal to one. This simple result is therefore stated thus: the increase per unit time in the circulation of a curve which is the section of a solenoid by any given plane has the numerical value of unity. This increase of circulation will be considered $+1$ when its direction agrees with the direction chosen as positive for the movement along the curve and by -1 in the opposite case. This result as mentioned above is for the special case of a curve that is the intersection of a plane with

a solenoid, and we shall now pass to the general theorem of any curve Through the given arbitrary curve we draw a surface which intersects all the solenoids inclosed within the curve. On this surface the solenoids determine a system of parallelogramatic curves, each of which receives in the unit of time an increase of circulation of either +1 or -1. But according to the summation theorem expressed in equation (128) for line integrals, the sum of the line integral along all individual contours is equal to the line integral along the exterior contour, and, therefore the latter is simply equal to the number of the included solenoids if all turn in the same directions, otherwise it is equal to the excess, N, of the number of solenoids turning positively over the number turning negatively. Since this line integral is equal to the increase per unit of time in the circulattion C, of the curve under consideration, we can, therefore express the result by the formula

$$N = \frac{dc}{dt} \qquad (136)$$

We can consider the number, N, with its algebraic sign, simply as the number of solenoids inclosed within the curve and can express the result by the following theorem:
"The increase in a unit of time in the circulation of any given closed curve is equal to the number of solenoid, inclosed within the curve."
With the help of this theorem we can follow the variation with time of the varying value of the circulation of a closed chain of fluid particles provided that we know at every moment the courses of the isobaric and isosteric surfaces. The number, N, will vary continually for two reasons:
(1) The curve is in motion
(2) The isobaric and the isosteric surfaces vary in consequence of the varying form and location of the conditions as to density and pressure, so that the curve incloses a bundle of solenoids that is continually varying.

III Hydrodynamical Considerations:

In the previous section of this chapter, the equation of the change of circulation of the fluid with respect to time was derived in a very elementary method depending on general dynamic principles. But to have a better understanding of Bjerknes theorem we need to consider some of the elements of hydrodynamics which concern vortex motion. For this purpose the follow ng derivations of hydrodynamical equation will be given for the case of frictionless liquids. In these derivations the density, ρ, of the fluid will be assumed to be a function of the pressure, P;

$$\rho = f(P) \qquad 137$$

(A) Equation of Continuity:
Let $\vec{V}(u,v,w)$ be the velocity of the fluid and $\vec{F}(x,y,z)$ be the force per unit mass acting on the fluid. Consider a fixed surface S in

the fluid. We assume that there is neither a sink nor a source in this surfaces and we measure the rate of increase of matter in it by the surface integral of the flux of fluid taken through the surface along the inward drawn normal, that is along $-\vec{n}$. Therefore we have

$$\frac{\partial m}{\partial t} = -\iint_S \rho \vec{V} \cdot \vec{n}\, ds \qquad (138)$$

where m is the mass of the fluid and S is the fixed surface. But from the divergence theorem

$$\iint_S \vec{w} \cdot \vec{n}\, ds = \iiint \nabla \cdot \vec{w}\, dv \qquad (139)$$

by applying it to the ~~left~~ right hand side of 138 we get

$$\frac{\partial m}{\partial t} = -\iint_s \rho \vec{v} \cdot \vec{n}\, ds = -\iiint \nabla \cdot (\rho \vec{v})\, dv \qquad (140)$$

where the volume integral is taken over the volume bounded by the surface S.
But since the mass is given by

$$m = \iiint \rho\, dv$$

therefore

$$\frac{\partial m}{\partial t} = \frac{\partial}{\partial t} \iiint \rho\, dv = \iiint \frac{\partial \rho}{\partial t}\, dv \qquad (141)$$

Since S is any surface from (141) and (140) we get

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \vec{v})$$

or

$$\frac{\partial \rho}{\partial t} \pm \nabla \cdot (\rho \vec{V}) = 0 \qquad (142)$$

Expressed in Cartesian coordinates equation 142 becomes

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\partial}{\partial z}(\rho w) = 0 \qquad (143)$$

Equation 142, or its cartesian equivalent, is called the equation of continuity. It states that matter is neither created nor destroyed at any point in the fluid.

In the previous derivation of the equation of continuity we used the Eulerian method, that is, we watched the changes that occured at a certain selected region in the space occupied by the fluid. However another method is used and known by the Lagrangian method; here we follow a selected particle or portion of the fluid throughout its motion and watch what changes it experiences.

To take into consideration the rate of change of a function due to both a fixed position and the motion of a point in the space characterized by this function we make use of the definition of the gradient. Let $\phi(x,y,z,t)$ be any function. According to the definition of the gradient of a function we have

$$\frac{\partial \phi}{\partial s} ds = \vec{ds} \cdot \nabla \phi \qquad (144)$$

where s is any curve along which ϕ changes but $\vec{ds} = \vec{V}dt$, substituting this in (144) we get

$$\frac{\partial \phi}{\partial s} ds = \vec{V} \cdot \nabla \phi \, dt \qquad (145)$$

but

$$\frac{d\phi}{dt} = \frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial s} \frac{ds}{dt} \qquad (146)$$

substituting in for 146 for $\frac{\partial \phi}{\partial s}$ its value in (145) we get

$$\frac{d\phi}{dt} = \frac{\partial \phi}{\partial t} + \vec{V} \cdot \nabla \phi \qquad (147)$$

This last equation gives the rate of change of a function ϕ as it is followed in its motion. And we can apply equation 147 for finding both the acceleration of the portion of fluid under consideration and to find the rate of change of density of a definite portion of the fluid.

For the acceleration of a definite portion of a fluid we have only to substitue for ϕ in (147) the velocity V and thus we get

$$\vec{A} = \frac{d\vec{v}}{dt} = \frac{\partial \vec{v}}{\partial t} + \vec{V} \cdot \nabla \vec{V} \qquad (148)$$

And for the rate of change of density we substitute for ϕ in (147) the density (ρ) and we get

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \vec{V} \cdot \nabla \rho \qquad (149)$$

It is easy to see that (149) is equivalent to the cartesian equation:

$$\left.\begin{aligned} \frac{d\rho}{dt} &= \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{dx}\frac{dx}{dt} + \frac{\partial \rho}{dy}\frac{dy}{dt} + \frac{\partial \rho}{\partial z}\frac{dz}{dt} \\ \text{or } \frac{d\rho}{dt} &= \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{dx}u + \frac{\partial \rho}{\partial y}v + \frac{\partial \rho}{\partial t}w \end{aligned}\right\} \qquad (150)$$

Now expanding $\nabla \cdot (\rho \vec{V})$ in the equation of continuity No, 142 by

using the vectorial equation

$$\Delta \cdot (\rho \vec{\omega}) = \vec{\omega} \cdot \nabla \rho + \rho \nabla \cdot \vec{\omega}$$

we transform it to the form

$$\frac{\partial \rho}{\partial t} + \vec{V} \cdot \nabla \rho + \rho \nabla \cdot \vec{V} = 0 \qquad (151)$$

Substituting (149) in 151 we get

$$\frac{d\rho}{dt} + \rho \nabla \cdot \vec{v} = 0 \qquad (152)$$

If the fluid is incompressible, then ρ does not vary neither with position nor with time. Therefore we have from 152 equating $\frac{d\rho}{dt}$ to zero.

$$\rho \nabla \cdot \vec{V} = 0$$

or

$$\nabla \cdot V = 0 \qquad (153)$$

which shows that $\vec{V}$ is then a solenoidal vector with its stream.lines forming closed curves or ending at infinity. If we consider a definite element of fluid consisting of the same particles, then as it moves its mass remains constant or

$$\rho v = \text{constant where v = volume}$$

differentiating with respect to time we get

$$v \frac{d\rho}{dt} + \rho \frac{dv}{dt} = 0$$

therefore $\frac{1}{v} \frac{dv}{dt} = -\frac{1}{\rho} \frac{d\rho}{dt}$ (154)

On substituting for $\frac{d\rho}{dt}$ from (152) we get

$$\frac{1}{v} \frac{dv}{dt} = \frac{1}{\rho} \frac{d\rho}{dt} = \nabla \cdot \vec{V} \qquad (155)$$

Thus it is apparent from (155) that $\nabla \cdot \vec{V}$ represents the fractional decrease of density per unit of time, or the rate of increase of volume per unit of volume, or the time rate of dilatation which is the divergence. Equation (153) follows also from (155) when ρ is constant.

(B) Euler's Equations of Fluid Motion:

Consider a definite mass of fluid enclosed in a surface S. Let $\vec{F}$ per unit mass or $\rho \vec{F}$ per unit volume be the external force acting on this mass, and let p be the pressure function acting normally over the enclosing surface and along the inwardly drawn normal. From Newtons law the rate of increase of movementum, $\Sigma \rho \vec{V} dv$ of the fluid is equal to the sum of applied forces $\vec{F}$ acting directly on the mass of the fluid and the forces $\Sigma p\, ds$

resulting from the pressures acting on the surrounding surface, or

$$\frac{d}{dt}\iiint \rho \vec{V}\, d\upsilon = \iiint \rho \vec{F}\, d\upsilon + \iint_S p\, \vec{n}\, ds$$

Where the volume integral is taken over the volume bounded by S.

But $\iint_S p\vec{n}\, ds = \iiint \nabla p\, d\upsilon$ from the divergence theorem.

Therefore we have

$$\iiint \frac{d}{dt}(\rho \vec{V} d\upsilon) = \iiint (\rho \vec{F} - \nabla p)\, d\upsilon$$

But $\frac{d}{dt}(\rho \vec{V} d\upsilon) = \rho \frac{d\vec{V}}{dt} d\upsilon + \vec{V} \frac{d}{dt}(\rho d\upsilon)$

and the last term vanishes because the mass of a definite portion remains constant; and therefore we have finally

$$\iiint \left(\frac{d\vec{V}}{dt}\rho\right) d\upsilon = \iiint (\rho \vec{F} - \nabla p)\, d\upsilon \qquad (156)$$

Since this was proved for any arbitrarily chosen volume we have then

$$\rho \frac{d\vec{V}}{dt} = \rho \vec{F} - \nabla p$$

Using (148) we get

$$\rho \frac{d\vec{V}}{dt} = \rho \frac{\partial \vec{V}}{\partial t} + \rho\, \vec{V} \cdot \nabla \vec{V} = \rho \vec{F} - \nabla p \qquad (157)$$

This is known as Euler's equation of motion in vector form and its cartesian equivalent is

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} &= X - \frac{1}{\rho}\frac{\partial p}{\partial x} \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} &= Y - \frac{1}{\rho}\frac{\partial p}{\partial y} \qquad (158) \\
\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} &= Z - \frac{1}{\rho}\frac{\partial p}{\partial z}
\end{aligned}$$

where u, v, and w are the components of $\vec{V}$ along x, y, z axis respectively and X Y, and Z are the components of $\vec{F}$.

If we divide equation 157 by ρ and employ the vectorial identity

$$\vec{V} \cdot \nabla \vec{V} = \frac{1}{2} \nabla \vec{V}^2 - \vec{V} \times (\nabla \times \vec{V})$$

we get a transform of 157 in the form

$$\frac{\partial \vec{V}}{\partial t} - \vec{V} (\Delta + \vec{V}) = \vec{F} - \frac{\nabla P}{\rho} - \frac{1}{2} \nabla \vec{V}^2 \qquad (159)$$

If the Force $\vec{F}$ is of a conservative character, that is if it is derived from a potential, then

$$\vec{F} = -\nabla \phi$$

Also if the pressure P at any point depends only upon the density ρ, we may define a quantity P such that

$$\frac{\nabla P}{\rho} = \nabla P \text{ or } P = \int \frac{dP}{\rho}$$

so that the equation 159 becomes

$$\frac{\partial \vec{V}}{\partial t} - \vec{V} \times \text{curl } \vec{V} = -\nabla \left(\phi + P + \frac{1}{2} \vec{V}^2 \right) \qquad (160)$$

or if we put $U = -(\phi + P + \frac{1}{2} \vec{V}^2)$

then

$$\frac{\partial \vec{V}}{\partial t} - \vec{V} \times \text{curl } \vec{V} = \nabla U \qquad (161)$$

But the curl of the velocity is equal to twice the angular velocity that is

$$\text{curl } \vec{V} = 2 \vec{\omega}$$

where $\vec{\omega}$ is the angular velocity of rotation of the fluid at the point considered. Then (160) becomes

$$\frac{\partial \vec{V}}{\partial t} - 2 \vec{V} \times \vec{\omega} = -\nabla \left(\phi + P + \frac{1}{2} \vec{V}^2 \right) \qquad (162)$$

If $\vec{F}$, $\vec{V}$, p, and are independent of the time then the motion is called steady. If we have a non-vortical steady motion, then $\vec{\omega} = 0$ and equation 162 becomes

$$-\nabla U - \nabla \left(\phi + P + \frac{1}{2} \vec{V}^2 \right) = 0$$

or integrating, we get

$$\phi + P + \frac{1}{2} \vec{V}^2 = \text{constant}$$

or

$$\phi + \frac{P}{\rho} + \frac{1}{2} \vec{V}^2 = \text{constant} \qquad (163)$$

which is the equation of motion for a steady nonvortical state.

If there are no external forces then $\vec{F} = 0$ and therefore ϕ constant and then equation 163 becomes

$$\frac{p}{\rho} + \frac{\vec{V}^2}{2} = \text{constant} \qquad (164)$$

Equation (164) implies that where the pressure is great the velocity must be small, and where the velocity is great, the pressure should be small. Thus in a constricted pipe the pressure is least at the constriction where the velocity of the incompressible fluid necessarily is the greatest.

(C) Vortex Motion:

If we apply the operation $\nabla\times$ to equation (162) and noticing that $\nabla \times \nabla \phi = 0$ we get

$$\text{Curl}\ \frac{\partial \vec{V}}{\partial t} + 2\,\nabla\times(\vec{\omega}\times V) = 0$$

or since t is independent of x, y, and z we get after expanding

$$2\,\nabla\times(\vec{\omega}\times\vec{V})$$

$$\frac{\partial}{\partial t}\,\text{curl}\,\vec{V} + 2\left(\vec{\omega}\,\nabla\cdot\vec{V} + \vec{V}\,\nabla\cdot\vec{\omega} - \vec{V}\Delta\cdot\vec{\omega} - \vec{\omega}\cdot A\vec{V}\right) = 0$$

since $\text{curl}\,\vec{V} = 2\vec{\omega}$

therefore $\nabla\cdot\omega = 1/2\ \nabla\cdot\nabla\times\vec{V} = 0$

and since also from 147

$$\frac{d\vec{\omega}}{dt} = \frac{\partial\vec{\omega}}{\partial t} + \vec{v}\cdot\nabla\vec{\omega}$$

we get therefore

$$\frac{d\vec{\omega}}{dt} + \vec{\omega}\,\nabla\cdot\vec{V} - \vec{\omega}\cdot\nabla\vec{V} = 0$$

which transforms into

$$\frac{d}{dt}\left(\frac{\vec{\omega}}{\rho}\right) = \frac{\vec{\omega}}{\rho}\cdot\nabla\vec{V} \qquad (165)$$

because from (155)

$$\nabla\cdot\vec{V} = -\frac{1}{\rho}\frac{d\rho}{dt}$$

and

$$\frac{d\vec{\omega}}{dt} + \vec{\omega}\,\nabla\cdot\vec{V} = \frac{d\vec{\omega}}{dt} - \frac{\vec{\omega}}{\rho}\frac{d\rho}{dt} = \rho\frac{d}{dt}\left(\frac{\vec{\omega}}{\rho}\right)$$

Differentiating (165) again we get

$$\frac{d^2}{dt^2}\left(\frac{\vec{\omega}}{\rho}\right) = \left[\frac{d}{dt}\left(\frac{\vec{\omega}}{\rho}\right)\right]\cdot\nabla\vec{V} + \frac{\vec{\omega}}{\rho}\cdot\frac{d}{dt}\left(\nabla\vec{V}\right) \qquad (166)$$

It is apparent from (165) and (166) that when $\vec{\omega}$ vanishes these equations vanish, and similarly all the successive derivatives may be shown to vanish. If $\vec{\omega}$ is ever zero, therefore, it will

always remain zero, because of Taylors theorem and of the vanishing of all the derivatives of $\vec{w}$.

This is known by the name "Helmoltz theorem" and it says that if no vorticity exists in any incompressible, frictionless fluid at any time it is impossible to produce any by means of a conservative system of forces, and therefore the motion will remain for ever non-vortical. Another proof of this theorem is given on the next page.

We have already defined the circulation along any path as the line integral of the velocity along that path. If the path is a closed one, we may express the circulation around it as a surface integral over any surface bounded by it by means of Stoke's theorem. Thus

$$\oint \vec{V} \cdot d\vec{V} = \iint_S \vec{n} \cdot \nabla \times \vec{V}\, ds \qquad (167)$$

Since $2\vec{w} = \nabla \times \vec{V}$ then 167 becomes

$$\oint \vec{V} \cdot d\vec{v} = \iint_S \vec{n} \cdot \nabla \times \vec{V} ds = 2 \iint_S \vec{n} \cdot \vec{\omega}\, ds \qquad (168)$$

Equation 168 implies that if the circulation of a fluid along a closed curve is not zero then $\nabla \cdot \vec{V}$ or $2\vec{w}$ is not zero and therefore the fluid must have vortical motion.

Lines drawn in the fluid so as at every point to coincide with the instantaneous axis of rotation of the corresponding fluid element are called vortex lines. Portions of the fluid bounded by vortex lines drawn through every point of an infinitely small closed curve are called vortex filaments, or simply vortices, and the boundary of a vortex filament is called a vortex tube.

Consider such a tube bounded by two surfaces S_1 and S_2. Applying the divergence theorem to the closed surface S_1, S_2 and the sides of the tube, and remembering that since $\vec{w}$ is solenoidal then $\nabla \cdot \vec{\omega} = 0$ we get

$$\iint_S \vec{n} \cdot \vec{\omega}\, ds = \iiint \nabla \cdot \vec{\omega}\, dv = 0 \qquad (169)$$

Since, in the surface integral, the sides contribute nothing to it, then there must be as much flux of $\vec{w}$ inward at S as there is outward at S_2, that is the flux is c nstantant throughout the tube. If the cross section of the vortex filament of such a tube is denoted by S, then the constancy of the flux of $\vec{w}$ is expressed by

$$S\, \vec{n} \cdot \vec{w} = \text{constant} \qquad (170)$$

where $\vec{n}$ is the normal to the cross-section. The product in (170) is called the strength of the filament. Thus if $\vec{w}$ is finite, or, in other words, if there exists a vorticity in a fluid, then S cannot vanish. Therefore a filament cannot end anywhere in the fluid and such filaments should either form closed curves or end at the surface of the fluid or at infinity. Then, all vortices form closed curves in the fluid or end in the surface. This same property of vortices is easily seen from the fact that the divergence of $\vec{w}$ is zero, that is $\vec{\omega}$ is a solenoidal vector.

We have thus proved that if $\vec{W}$ is zero in a frictionless incompressible fluid it remains so, and that if it is not zero it cannot vanish unless the vortex tube cuts the surface. We shall now prove that it is impossible to produce a vortex in a frictionless fluid. Defferentiating the expression for the circulation along any path, namely

$$C = \int_A^B \vec{V}.d\vec{r}$$

we get

$$\frac{dC}{dt} = \frac{d}{dt}\int_A^B \vec{V}.d\vec{r} = \int_A^B \frac{d\vec{V}}{dt}.d\vec{r} + \int_A^B \vec{V}.\frac{d}{dt}d\vec{r} \qquad (171)$$

If we assume that curl $\vec{V} = 2\vec{w} = 0$, that is there is no vorticity, and in this case equation (157) becomes, if $W = (-\phi + P)$

$$\frac{d\vec{V}}{dt} = \nabla W \qquad (172)$$

so that

$$\frac{d\vec{V}}{dt}.d\vec{r} = d\vec{r}.\nabla W = dW$$

also

$$\vec{V}.\frac{d}{dt}d\vec{r} = \vec{V}.d\frac{d\vec{r}}{dt} = \vec{V}.d\vec{V} = d\left(\frac{\vec{V}^2}{2}\right)$$

and therefore on substitution in (171) we get

$$\frac{dC}{dt} = \int_A^B d\left(W + \frac{\vec{V}^2}{2}\right) = \left[W + \frac{\vec{V}^2}{2}\right]_A^B \qquad (173)$$

If the path is a closed one, then

$$\left[W + \frac{\vec{V}^2}{2}\right]_A^{B=A} = 0$$

because W and $V^2/2$ are scalar point- functions of position in space and have identical values at the limits, so finally we have in the case of a frictionless fluid

$$\frac{dc}{dt} = 0 \qquad (174)$$

Equation (174) inplies that the circulation around any closed curve, formed of a chain of particles of the fluid, cannot change as these particles are carried about by the liquid. Since we assumed that there is no circulation because $\vec{\omega}$ is zero at the beginning then it remains so for ever, or in other words, it is impossible to create vorticity in a frictionless fluid by means of a

conservative system of forces. Also, since it is impossible to conceive of any system of forces that acts in a non-conservative manner on a frictionless fluid, it follows that it is impossible to create vorticity in any manner in a frictionless medium.

Now, having attained a rather sufficient understanding of the fundamental equations of hydrodynamics, and we are in a position to take up the Bjerknes theorem of sunspots.

CHAPTER VII

The Hydrodynamical Explanation of Sunspots

I Introduction:

We shall consider now the work of V. Bjerknes on the application of the hydrodynamical principles to the sun[1] in such a way that an explanation to the sunspot phenomenon is obtained.

We have seen in the previous chapter that the internal dynamics of a perfect fluid are governed by the field of mass and the field of pressure. The pressure gradient points in the direction of decreasing pressure and represents force per unit volume originating from pressure. The inertia gradient points in the direction of decreasing inertia or increasing specific volume (the inertia-gradient is equivalent to the vector of motion in the previous chapter). If the two vectors representing the two gradients do not coincide, then rotation about an axis normal to the two vectors occurs, tending to bring the densest masses in the rear; and the rotation, therefore, will be directed from the inertia gradient towards the pressure gradient. We have seen also that if the field of pressure be represented by isobaric surfaces drawn for unit differences of pressure, and the filed of mass by isosteric surfaces drawn for unit differences of specific volume, then space is divided into a set of unit tubes characterizing the asymmetry which the field of mass presents to that of pressure. The properties of such tubes were also discussed, and the fundamental dynamic equation of the variation of the circulation with time was derived. If barotropic conditions are assumed then the field of pressure governs that of inertia and the inertia gradient coincides with the pressure gradient thus reducing the number of unit tubes, N, in equation 136, to zero and we have therefore

$$\frac{dc}{dt} = 0$$

or $$C = \text{constant} \qquad (175)$$

Equation (175) shows the invariability of circulation under barotropic conditions in a closed fluid. It is known by the name of the "Helmholtz-kelvin principle of the conservation of circulation and vortices."

Circulations in fluids or gases must be in accordance with equation 136, namely,

$$\frac{dc}{dt} = N \qquad (136)$$

---

(1) "Solar Hydrodynamics" V. Bjerknes A. P. J. 64 93 1920

and may be calculated from it when we can determine the isosteric surfaces or the equivalent isothermal surfaces as something different from the isobaric surfaces; or conversly when the variations of circulation are known the relative distribution of the "iso-surfaces" of mass and pressure could be determined.

Let us now consider a horizontal curved current. For simplicity assume that the motion is steady or approximately steady. If the current has an upward increasing intensity, the upper masses will have an excess of centrifugal force. This produces a centrifugal pumping effect which lifts heavy masses of adiabatically cooled gas on the inner side of the current. Or if the current is of an upward decreasing intensity, the lower masses will have an excess of centrifugal force, which causes a centrifugal sucking effect that brings light masses of adiabatically heated gases down on the inner concave side of the current. Thus, a vortex with an upward-increasing velocity has a core of heavy masses or cold gas; a vortex with an upward-decreasing velocity, a core of light masses of hot gas.

Equation (136) gives completely the dynamics of the pumping and sucking effects and leads to a convenient formula characterizing the equilibrium conditions ultimately reached. In deriving such a formula, use of the gas equation.

$$\frac{1}{T}\frac{dT}{ds} = \frac{2\omega \sin\varphi}{g}\frac{dv}{dz} \qquad (176)$$

where

T = absolute temperature, introduced in the plan of the specific volume

dv = increase in velocity for an increase in height, dz,

g = acceleration of gravity

dT = the increase in temperature for an increase in length, ds, measured along a curve S, which is contained in the isobaric surface, runs normal to the lines of flow, and is counted positive outward, that is, from the concave to the convex side of the current.

ω = angular velocity of the moving particles around an axis making an angle φ with the isobaric surface.

$\omega \sin\varphi$, therefore, represents the component of angular velocity round the normal to the isobaric surface. Equation (176) holds as long as the vertical accelerations are zero or moderate and and the motion is approximately steady.

This principle of pumping and sucking effects of curved currents takes a slightly different form when we pass from absolute motion to the motion relative to a rotating body. A current which is straight, relative to the rotating rigid body, will be curved from the view-point of absolute motion. Instead of referring to the convex and concave sides of the current we refer to the left and right side of the apparently straight current and find the following rule.

"In the northern hemisphere of the earth or sun a current will carry cold masses on its left side, if it has an upward-increasing intensity, and on its right side, if it has upward decreasing intensity"(1

(1) A. P. J. 64 96 1926

For example the general atmospheric motion of the morthern hemisphere of the earth is an eastward drift with upward-increasing intensity and this explains why such a motion carries the cold polar air masses on its left side. In the case of such relative motions formula 176 may be applied with the change of φ into the latitude and ω into the angular velocity of the earth or the sun. A more complex formula may be derived for the case of a current which appears curved in its relative motion, but we shall be satisfied now with equation 176, if we interpret ω as the greater of the two angular velocities: that of the vortex, and that of the solid rotating body the sun or earth.

The production of circulations is best shown from thermodynamical considerations. For this purpose, consider a closed tube of uniform cross-section filled with fluid or gas in adiabatic equilibrium. We place in one branch a source of heat at a certain height and at the other branch a source of cold, then the fluid or gas will begin t circulate as in ("A" figure 13) . The circulation is maintained by a force equal to the difference in weight between two fluid columns of different temperatures but of the same height, which is equal to the difference in elevation of the two sources of heat and cold. This engine is reversible, that is, circulation could be reversed, as in B (figure 13), by means of a certain initial retrograde circulation; and the efficiency of the engine remains the same if the two heights of the sources of heat and cold are kept constant. The only difference between the two cases A and B is that the engine is self-sparting in A, while an initial impulse is required to start the retrograde motion in B. The efficiency will be greatest when the source of heat is at the lowest point and the source of cold at the heighest point of the tube as in C and D (figure 13). There is no difference, here, between direct and retrograde circulation, and the engine is not a self-starting one, but needs an initial impulse to start it in any direction. When the source of cold is at a lower level than the source of heat the efficiency is negative, and kinetic energy of circulation will be transferred into heat energy given to the source of cold. However, the circulation in the sun is bound to be of a stratified nature. This is a result of nonbarotropic conditions.

Thus, if the fluid or gas is not in barotropic equilibrium, that is if it is hetrogeneous, with heavier masses below and lighter masses above, a heated element will then ascend only to the level of fluid having its own density. This gives rise to stratified circulation.

Mendenhall and Mason[1], after a series of experiments, found that if we have a tube full of a hetrogeneous fluid, with different strata, then after a certain epoch, under constant temperature, these strata disappear; and that tubes of freshly suspended material would not develop stratifications in the constant temperature gradient across the liquid. Also they found that a density gradient in addition to the temperature gradient, is necessary to produce stratified circulation.

To see how this kind of circulation is produced, let us consider an intenesting experiment due to Alf Sinding-Larsen.[2] A test tube

---

(1) Proceedings of the National Acad. of Science of the U. S. A. 9, 199, 1923.
(2) Annalen der Physik 9, 1190, 1902

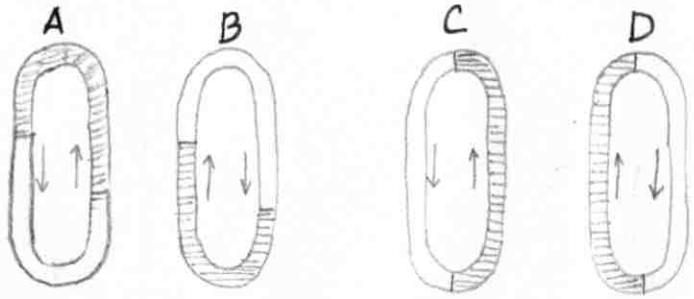

Fig 13

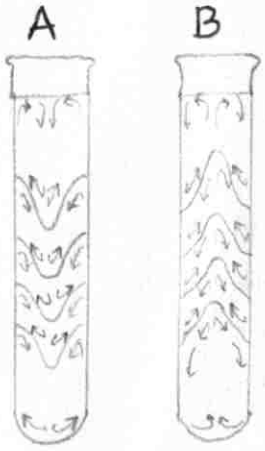

Fig 14

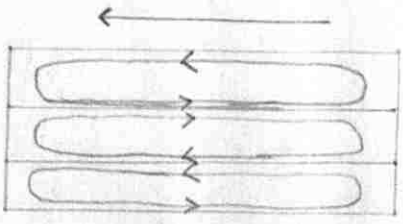

Fig 15

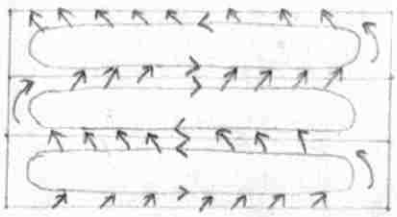

Fig 16

containinga hetrogenous fluid is put in hot water. The salt solution in the tube, then divides into overlying strata each having its independent circulation (as in figure 14 A). If then the tube is removed from the hot water into a vessel containing cold water, the reverse circulations are observed (as in figure 14 B). Circulation of this type is symmetrical with respect to the axis of the tube. The temperature gradient between the sides and the axis of the tube gave rise to this circulation in the hetrogeneous fluid of the tube. Circulation across the tube may be obtained by heating one side and cooling the other side of the tube.

In this experiment we have a counteracting force due to friction. But under favourable conditions friction may produce stratified circulation. The following consideration may illustrate this.

Consider a series of superimposed strata as in figure 15. A unidirectional wind on the surface of the top stratum will set it in circulation. By the force of friction the next stratum is put in opposite circulation and so on, with the result that all strata circulate like-toothed wheels with decreasing intensity as we proceed from stratum to stratum. In this case friction is the factor that produces circulation in a fluid which is supposed to be already stratified. If the solution, also, has initially continuous varying concentration, then the wind will produce an upper circulating and homogeneous layer, which will tend to produce a lower, second circulating layer and so on.

Certain conditions may also produce circulations where both friction and heat cooperate in producing and maintaining this stratified circulation. To see how this could happen, consider a fluid which is heated from below and cooled at the surface. As in figure 13 C and D, conditions are unstable initially, and to begin the circulation an impulse is necessary. However, this impulse may be avoided by inclining the heated bottom so that the flow of heat (small inclined arrows in figure 16) is inclined to the sections parallel to the bottom surface of the fluid, this inclination will cause circulation. It is apparent that if the fluid is hetrogeneous stratified circulation will be observed, and, as it is apparent from figure 16, the frictional force is cooperating with the temperature gradient in producing and maintaining the circulation.

Having acquired an elementary description of how stratified cirdulation may be produced, we shall now summarize the data concerning the sunspot phenomenon and give some further assumptions necessary for the explanation.

II Empirical Data Concerning the Sun:

The various aspects of the solar phenomena were discussed in part II of this report, but for the sake of definiteness the following summary of the Solar phenomena necessary for the theory will be given.

1. The sun rotates on its axis in about 25.5 days with a velocity decreasing from equator to both poles. The different layers also have velocities increasing from photosphere to the surface.
2. Sunspots seem to be depressions in the photosphere, and they occur in lower latitudes between 5° and 45° S. or N. of the equator.
3. Sunspots of each cycle appear within two zones, 10° to 15° symmetrically situated on both sides of equator. These zones begin from latitude about ±40° and gradually come to within ±5° from equator in a period of about 11.5 years. Before the disappearance of one cycle a new cycle appears in the higher latitudes again.

4. Spots usually appear in pairs with an axis inclined to the equator in such a way that the angle of inclination varies with the latitude. The maximum of this inclination is about 11° in middle latitudes and it decreases as the cycle comes to an end until it is approximately zero near the equator.
5. Sunspots show magnetic polarity, members of a binary system show opposite polarity.
6. The preceding spots of one cycle have the same polarity in one hemisphere, opposite polarities in the two hemispheres, and the whole polarity is reversed in each new cycle.
7. The preceding spot is larger in area and its increase is more rapid than that of the following spot.
8. The following spot disappears before the leading one.
9. The motion of the preceding spot in longitude is more than the following spot.
10. In very rare cases some spots were seen very near to the equator.
11. The ascent from minimum to maximum is steeper than the descent from maximum to minimum in a sunspot activity curve of cycle.
12. The equator-wards progression of spot zones is not found in the spots themselves after they are formed.
13. The minimum of prominence frequency is always noticed at a spots minimum, and in general the solar activity is more at spots maximum than at spots minimum.
14. Hydrogen spectroheliograms in the region of sunspots often show spiral structure. Curvature is opposite in the two hemispheres and does not change from cycle to cycle, if this structure represents lines of flow, motion is both radial and circular. Hydrogen clouds are seen drawn inward into a spot. Motion outward from the spot in the lower ~~layer~~ layers of the solar atmosphere has also been observed.

To interpret the observed phenomena, the following assumptions concerning the sun were made by Bjerknes.

(1) Since the sun consists of gases of different specific weights, it will be near a state of stable internal equilibrium and not an adiabatic one. Thus a mass of cooled gas at sun's surface will not descend to the centre of the sun, but it will descend to a certain level from which when heated it will ascend again. Thus, conditions are favourable for stratified circulation.
(2) Solar radiation is responsible for the internal dynamics of the sun.
(3) The heat lost by the photosphere is restored from deeper layers by internal radiation, convection, and conduction. Radiation is predominant, conduction is negligible, but convection though negligible at lower levels becomes important near the surface where the temperature and, therefore, the radiation becomes less.
(4) The rotation of the sun tends to produce a general zonal symmetry around the axis of rotation.
(5) Sunspots are a kind of vortex. The magnetic fields are reversed in polarity with a reversal of vortex motion.

Starting from these assumptions Bjerknes develops his theory of solar hydrodynamics which gives the explanation to most of the observed phenomena.

III Bjerknes' ~~Interputation~~ Explanation of the Solar Phenomena;

(1) The Vortex Theory of Sun-spots:--

If we assume that the production of sunspots is similar to the terrestial phenomenon of tropical cyclones that is, due to vortical

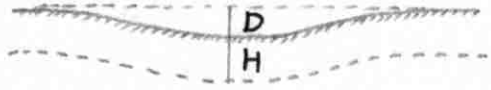

Fig 17

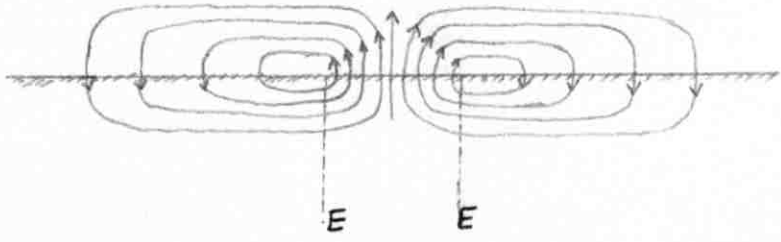

Fig 18

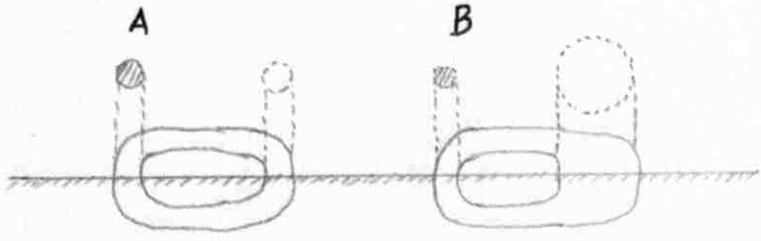

Fig 19

motion in the sun, then it seems interesting to compare the relative size of these two phenomena. The diameter of a small tropical cyclone is about 125 km. If we enlarge the earth 100 times then it approximates the dimensions of the sun. The cyclone then will be of the dimensions of the equatorial plane of the earth, and thus represents the area of an average sunspot. The height of a tropical cyclone is between 5 and 10 km. and on this ratio the depth of a sunspot will become 500 to 1000 km.

Our first problem to be discussed is how the vortex theory accounts for the visibility of sunspots. For this purpose, use is made of equation (176). This equation gives the temperature in a balanced vortex in which the cooled heavy masses are just carried by the centrifugal pumping effect, and no ascending motion is present. Sunspot vortices do not have this balanced character. The lifted cooled masses are subject to continuous radiation from below, and the pumping must go on continuously to lift the cooled masses at a sufficient rate to keep them sufficiently cool in spite of the radiation. However, formula 176 , for a balanced vortex will give a first approximation. If for ds sin ϕ we subsitute the element dr of a radius vector from the axis of the vortex then equation 176 becomes

$$\frac{dT}{T} = \frac{2\omega}{g}\frac{dv}{dz}\,dr$$

If we assume the velocity, v, to be zero at a certain depth H below the surface of the photosphere and increase upward linearly until it reaches the value V or ωr at the surface, then,

$$\frac{dv}{dz} = \frac{V}{H} = \frac{\omega r}{H}$$

and therefore

$$\frac{dT}{T} = \frac{2\omega^2 r}{Hg}\,dr$$

But the ratio of centrifugal force $m\omega^2 r$ to mg is the tangent of the angle of inclination of an element ds of the photosphere, and this multiplied by the horizontal line element dr gives the vertical projection dz of a line element ds contained in the inclined surface of the photosphere. Thus we have

$$\frac{dT}{T} = \frac{dz}{H}$$

Integrating this equation along the surface of the photosphere from a great distance to the centre of the vortex, and assuming that H is constant and also the change in temperature ΔT is small as compared with T, we get the simple relation between the drop in temperature ΔT and the corresponding dip (D figure 17) below the normal level of the surface of the photosphere, that is

$$\Delta T = \frac{2TD}{H} \qquad (177)$$

And for large values of the drop in temperature we must use the more exact result

$$\log_e \frac{T_0}{T} = \frac{2D}{H} \qquad (178)$$

motion in the sun, then it seems interesting to compare the relative size of these two phenomena. The diameter of a small tropical cyclone is about 125 km. If we enlarge the earth 100 dimensions times then it approximates the dimensions of the sun. The cyclone then will be of the dimensions of the equatorial plane of the earth, and thus represents the area of an average sunspot. The height of a tropical cyclone is between 5 and 10 km. and on this ration the depth of a sunspot will become 500 to 1000 km.

Our first problem to be discussed is how the vortex theory accounts for the visibility of sunspots. For this purpose, use is made of equation (176). This equation gives the temperature in a balanced vortex in which the cooled heavy masses are just carried by the centrifugal pumping effect, and no ascending motion is present. Sunspot vortices do not have this balanced character. The lifted cooled masses are subject to continuous radiation from below, and the pumping must go on continuously to lift the cooled masses at a sufficient rate to keep them sufficiently cool in spite of the radiation. However, formula 176 , for a balanced vortex will give a first approximation. If for ds sin θ we subsitute the element dr of a radius vector from the axis of the vortex then equation 176 becomes

$$\frac{dT}{T} = \frac{2\omega}{g}\frac{dv}{dz}\,dr$$

If we assume the velocity, v, to be zero at a certain depth H below the surface of the photosphere and increase upward linearly until it reaches the value V or ωr at the surface, then,

$$\frac{dv}{dz} = \frac{v}{H} = \frac{\omega r}{H}$$

and therefore

$$\frac{dT}{T} = \frac{2\omega^2 r}{Hg}\,dr$$

But the ratio of centrifugal force $m\omega^2 r$ to mg is the tangent of the angle of inclination of an element ds of the photosphere, and this multiplied by the horizontal line element dr gives the vertical projection dz of a line element ds contained in the inclined surface of the photosphere. Thus we have

$$\frac{dT}{T} = \frac{dz}{H}$$

Integrating this equation along the surface of the photosphere from a great distance to the centre of the vortex, and assuming that H is constant and also the change in temperature ΔT is small as compared with T, we get the simple relation between the drop in temperature ΔT and the corresponding dip (D figure 17) below the normal level of the surface of the photosphere, that is

$$\Delta T = \frac{2TD}{H} \qquad (177)$$

And for large values of the drop in temperature we must use the more exact result

$$\log_e \frac{T_0}{T} = \frac{2D}{H} \qquad (178)$$

Formula 177 means that the drop in temperature is approximately proportional to the dip in the surface of the photosphere. The dip is due to the excess weight of the relatively cool gases forming the core of the vortex; these cooled gases are prevented from leaving the surface by the centrifugal pumping effect arising from the excess of circulatory motion in the higher strata. Also for the same value of the dip, the drop in temperature is inversely proportional to the thickness, H, of the vortex. From 177 it can be shown that a dip in the photosphere equal to 1/10 of the thickness of the vortex gives a drop of 1100° C for T = 6000° C. If the diameter of a vortex is ten times the thickness then the average inclination required to produce a dip of 1/10 of the thickness would be 1/100. The velocities required to produce this inclination at different distances r from the axis of rotation are:

r =	10	100	1000	10000	100000	km
v =	0.2	0.6	1.7	5.5	17.5	km/sec.

For a four-fold inclination which gives temperature drop of 3300C the velocities will be doubled, thus remaining very small ascompared with observation. This is accounted for by Bjerknes by the fact that the vortex is an unbalanced one and the results obtained above are only approximations.

Another important question, in addition to the visibility of spots, is the depth from which the cooled gases should be pumped up in order to give the required drop in temperature, for instance 1100°. The ediabatic decrease of temperature in the earth's atmosphere is 1° C per hundred meters of height. On the sun with gravity 27 times stronger, the decrease would be 27° per hundred meters for an atmosphere like that of the earth, and Bjerknes assumes for the solar gases a value which is not very different from this. When these gases are lifted 10,000 meters or lo km, they should cool down 2700. How much this would reduce their temperature below that at the new level dpends upon the solar temperature gradient. If, as in the earth's atmosphere, this is half the adiabatic gredient, the result would be 1300 . However, these assumptions may be misleading, because of the differences between the heating of the earth's and the solar atmospheres.

A calculation due to Russell[1], for the degree of expansion which gives such a drop in temperature is given below.

The relatively low temperatures of sunspots are due to the cooling of the gases by expansion in the upper part of the vortex. The temperature of sunspots may be taken to be 2000° less than that of the photosphere. To estimate the degree of expansion necessary to produce such a cooling, let us consider a mass of gas rising in the vortex. It cannot cool faster than the rate given by adiabatic expansion. Meanwhile, the temperature of the surrounding solar gases through which the vortex column rises, must diminish upward, and this temperature gradient cannot be less than that of radiative equilibrium, for otherwise the observed outward flow of heat from the sun's interior could not be maintained. We have seen in our discussion of radiative equilibrium that

---

$T \propto p^{\frac{1}{4}} \propto \rho^{\frac{1}{3}}$

---

(1) A. P. J. 54 293-5 1921

Formula 177 means that the drop in temperature is approximately proportional to the dip in the surface of the photosphere. The dip is due to the excess weight of the relatively cool gases forming the core of the vortex; these cooled gases are prevented from leaving the surface by the centrifugal pumping effect arising from the excess of circulatory motion in the higher strata. Also for the same value of the dip, the drop in temperature is inversely proportional to the thickness, H, of the vortex. From 177 it can be shown that a dip in the photosphere equal to 1/10 of the thickness of the vortex gives a drop of 1100° C for T = 6000° C. If the diameter of a vortex is ten times the thickness then the average inclination required to produce a dip of 1/10 of the thickness would be 1/100. The velocities required to produce this inclination at different distances r from the axis of rotation are:

r =	10	100	1000	10000	100000	km
v =	0.2	0.6	1.7	5.5	17.5	km/sec.

For a four-fold inclination which gives temperature drop of 3300C the velocities will be doubled, thus remaining very small ascompared with observation. This is accounted for by Bjerknes by the fact that the vortex is an unbalanced one and the results obtained above are only approximations.

Another important question, in addition to the visibility of spots, is the depth from which the cooled gases should be pumped up in order to give the required drop in temperature, for instance 1100°. The adiabatic decrease of temperature in the earth's atmosphere is 1° C per hundred meters of height. On the sun with gravity 27 times stronger, the decrease would be 27° per hundred meters for an atmosphere like that of the earth, and Bjerknes assumes for the solar gases a value which is not very different from this. When these gases are lifted 10,000 meters or 10 km, they should cool down 2700. How much this would reduce their temperature below that at the new level dpends upon the solar temperature gradient. If, as in the earth's atmosphere, this is half the adiabatic gradient, the result would be 1300°. However, these assumptions may be misleading, because of the differences between the heating of the earth's and the solar atmospheres.

A calculation due to Russell[1], for the degree of expansion which gives such a drop in temperature is given below.

The relatively low temperatures of sunspots are due to the cooling of the gases by expansion in the upper part of the vortex. The temperature of sunspots may be taken to be 2000° less than that of the photosphere. To estimate the degree of expansion necessary to produce such a cooling, let us consider a mass of gas rising in the vortex. It cannot cool faster than the rate given by adiabatic expansion. Meanwhile, the temperature of the surrounding solar gases through which the vortex column rises, must diminish upward, and this temperature gradient cannot be less than that of radiative equilibrium, for otherwise the observed outward flow of heat from the sun's interior could not be maintained. We have seen in our discussion of radiative equilibrium that

$$T \propto P^{\frac{1}{4}} \propto \rho^{\frac{1}{3}}$$

(1) A. P. J. 54 293-5 1921

In adiabatic expansion

$$T \propto \rho^{\gamma-1} \propto P^{\frac{\gamma-1}{\gamma}}$$

γ being the ratio of specific heats.
Suppose that the vortex starts at a depth where the temperature, pressure, and density are T_0, P_0, ρ_0, and rises to the surface where, outside the spots, the quantities are T_1, P_1, ρ_0. In the spots, at the visible surface, the temperature will have a lower value T_2. The pressure will be about the same as P_1. and also the solar gravity is the same.
Hence

$$P_2 = P_1$$

and

$$\rho_2 = \frac{T_1}{T_2}\rho$$

For radiative equilibrium we have therefore

$$\frac{T_1}{T_0} = \left(\frac{P_1}{P_0}\right)^{\frac{1}{4}}$$

and for adiabatic equilibrium

$$\frac{T_2}{T_0} = \left(\frac{P_1}{P_0}\right)^{\frac{\gamma-1}{\gamma}}$$

therefore

$$\frac{T_2}{T_1} = \left(\frac{P_1}{P_0}\right)^{\frac{3\gamma-4}{4\gamma}} = \left(\frac{T_1}{T_0}\right)^{\frac{3\gamma-4}{\gamma}}$$

or

$$\frac{T_0}{T_1} = \left(\frac{T_1}{T_2}\right)^{\frac{\gamma}{3\gamma-4}}$$

and

$$\frac{P_0}{P_1} = \left(\frac{T_0}{T_1}\right)^4$$

and

$$\frac{\rho_0}{\rho_1} = \left(\frac{T_0}{T_1}\right)^3$$

Now in the actual case T_1 is about 6000°, T_2 = 3500° to 4000° and $\frac{T_1}{T_2} = 1.5$ or more. The greatest possible value of γ is 5/3 for a monatomic gas. This gives

$$T_0 = (2.0) \times T_1 = 12000^\circ C$$

$$P_0 = 16\, P_1$$

$$\rho_0 = 8\, \rho_1 = 5.3\, \rho_2$$

But this must greatly underestimate the amount of expansion. ~~But~~

~~this must greatly underestimate the amount of expansion~~. But if $\gamma = 7/2$ (diatomic gas) then

$T_0 = 17\ T_1 = 100000^\circ$

$P_0 = 80000\ P_1$

$\rho_0 = 5000\ \rho_1 = 3000\ \rho_2$

for $\gamma = 3/2$ we have

$T_0 = 3.7\ T_1 = 22000^\circ$

$P_0 = 190\ P_1$

$\rho_0 = 51\ \rho_1 = 34\ \rho_2$

From these few cases it is evident that the base of the spot vortex must be a region of very high temperature probably over 20000° C. The increase of volume on ascending must be larger, probably 30 times. These results show that the height which such gases should rise to should be much more than what Bjerknes estimated by the analogy of the earth's atmosphere.

However, the Bjerknes theorem explains very well the cause of the visibility of spots, but Bjerknes does not explain why the drop in temperature should be of a certain amount in most of the observed cases.

Under the intensive radiation from below, the cooled masses forming at any moment the core of the vortex cannot long retain their low temperature. As they become heated they will be replaced by new masses pumped from below and float out radially over the surface of the photosphere. This outward radial motion at the bottom of the solar atmosphere must give rise to a sink and a corresponding radial inflow in the upper strata. This inflow will assume a spiral character because of the rotation of the sun. Since this spiral motion is due only to the rotation of the sun, therefore, it should not change with the reversal of the polarity at the beginning of a new cycle.

For further discussion of the structure of sunspot vortices use is made of vortex lines. Each element of a vortex line, as was mentioned in the previous chapter, is an axis of local rotation, the direction of which is given by the positive screw rule. By geometric necessity, vortex lines can never stop, they run back into themselves, just like magnetic lines. Where vortex and magnetic lines are concentrated we have strong vorticity and strong magnetic field. Where they are dispersed the vorticity and the magnetic field are both weak.

The simplest hypothesis that can be made is that each sunspot is an independent vortex, Figure (18) shows the vortex lines representation of such a vortex. At the surface of the photosphere between E and E' we have a limited area of intensive positive vorticity and strong magnetic field; outside we have a weak negative vorticity and a weak field opposite to that in the middle area.

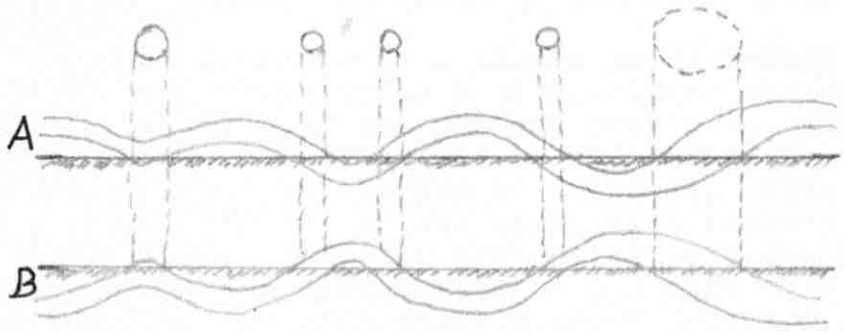

Fig 20

Fig 21

But such a simple hypothesis does not account for the usual binary spot phenomenon, and is replaced for this purpose, by a vortex ring, a suggestion due to Hale. The two spots of the binary are the intersection of the vortex ring with the photosphere (figure 19 A). When the ring cuts the surface so diffusely that nothing could be observed then one spot instead of two will be observed (figure 19 B). This second hypothesis also fails to account for the properties of spots of one cycle. For this a complete vortical connection between all spots of the same cycle is postulated (figure 20)

Thus all spots may belong to a single zonal vortex which surrounds the sun approximately as a parallel and gives rise to a spot, wherever it passes from the photosphere to the atmosphere or vice versa. This zonal vortex may belong mainly to the atmosphere or mainly to the photosphere (figure 20 A and B). This zonal vortex theory gives a simple picture of the collective properties of sunspots of the same cycle. Thus the properties of spots as that of the binary appearance, the parallelism of the axis of a binary system to the equator, the same succession of polarity in the same cycle, the bipolar character of sunspot groups at their appearance the single spot occurrence, and that a companion to a single spot appears in the rear if the original spot has the leading polarity for the cycle, and ahead if the opposite state of affairs predominates all these find their immediate explanation in this hypothesis. In addition the present theory reduces the progression of sunspot phenomena from higher to lower latitudes to a motion of the single zonal vortex, while the reversed succession of polarities from cycle to cycle is explained as a reversal of the rotation of the zonal vortex. All these properties and even the irregularities observed in certain cases follow from this hypothesis in its both subphotospheric or over-photospheric position.

But there is a sufficient reason to decide that this zonal vortex is sub-photospheric, and that is because just below the radiating stratum of the photo-sphere is found the strongest tendency to form convective circualtions of thermal origin, thus favouring the formation of a vortex which acts as in C or D of figure 13, to carry hot gases from below upward and cool gases from the surface downward. It remains now to explain why such a sub-photospheric vortex should occasionally ascend to the surface. This may be attributed to the attraction or repulsion existing between two parallel vortices; or to the attraction of the surface of the photosphere (considered as a rigid body) to the vortex which has a sinuous shape. A more important reason is that the circulation below the surface of the sun can take only a stratified form with limited vertical excursions, whereas no special forces limit the horizontal excursions, therefore the vortex must be very flat, the circulating curves having very flat elliptic forms with horizontal major axis and vertical minor axis. This strained type of circulation may have a strong tendency to turn over into a horizontal plane where it will be very stable. This tendency to realize the most stable motion may be the main reason why the vortex rises and cuts the photosphere, and why afterward new lengths of the photospheric vortex are drawn up to the already-formed horizontal sunspot vortex, thus

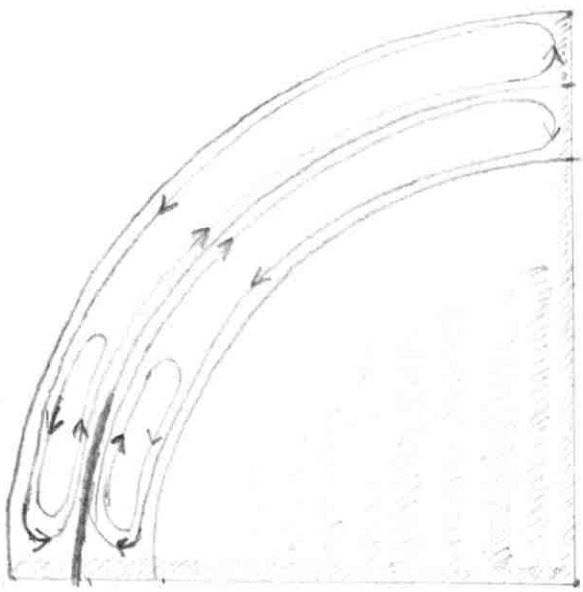

Fig 22

giving it a continuous supply of sub-photospheric energy. When the vortex tube is thus bent up, the strained elliptic circulation in the vertical plane changes into a circular circulation in a horizontal plane (figure 21). For the same area of cross-section this gives much shorter circulating curves, and therefore an intensified circulation. This circulation in a horizontal plane lifts by the pumping effect, gases from below and makes the spot visible by the lowering of the temperature as explained before. The kinetic energy of this vorticity is therefore changed into mechanical energy of pumping, but new energy is supplied by the new lengths of the sub-photospheric vortex that may be raised up to the surface. Bjerknes explains the maintainance of a spot for a certain period of time by this supply of energy.

Thus a zonal vortex assumption explains the properties of spots of a cycle. To explain the properties of different cycles and to show how such vortical motion is produced in the sun, we shall consider the general solar circulation.

(2) The General Solar Circulation:

Since there is a density gradient and a temperature gradient in the sun as we proceed outward from the sun's centre, and even since strata of different density are very likely pre-existing in the solar interor, then conditions are such that stratified circulation should be produced. Due to the rotation of the sun, this stratified circulation will be symmetrical with respect to the solar axis of rotation. An example of the simplest possible scheme of stratified circulation in the rotating sun is shown in figure (22). Here the highest stratum is supposed to have a circulation from the poles to the equator at the surface, and from equator to pole at the bottom. The next lower stratum will have naturally an opposite circulation and so on. The highest layer is different from the earth's atmosphere in that the circulation of the latter is from equator to pole on the outer surface because the main source of heat energy is at the equator and therefore the circulation is self starting in that direction, while in the sun the circulation may take both directions (C and D figure 15). The motive power of the circulation in the upper stratum is the cooling of the photosphere by radiation, and the corresponding heating of the lower part of this highest stratum by radiation from within and by contact with the next lower stratum which is at a higher temperature. This temperature gradient favours heat transport by convection. If such a circulation exists, the surface of the sun, as seen from the earth, will not rotate as a rigid body. Only those parts which do not participate with the general circulation rotate as a rigid body (Shaded area in figure 22). Those circulating parts are now subject to the terrestial law governing deflection of the winds, thus explaining what is called equatorial acceleration, which is nothing but retardation in the higher latitudes. Due to the slow motion of such a circulation, and to the high angular velocity of the rotating sun this motion will appear as a practically pure east-west motion with a slow displacement of the masses in the north-south direction. Observations indicate that the retarded masses in moderately high latitudes have had their angular velocities around the sun's axis reduced by 1/20. This would represent an

east wind relative to the quasi-rigid body of the sun which would traverse the solar circumference in about 20 solar days (which is equivalent to one terrestial year) with a velocity about 130 meters per second. The motion from the poles to the equator should therefore require several years, and the corresponding north wind would be very weak.

The sub-photospheric zonal vortex that causes sunspots must belong to the lower, slow-moving layers of this east-current. Then when the vortex rises, cuts the photospheric surface, and forms a binary spot system, the two separated parts of the tube must be affected differently by the stronger east wind near the surface. This might be an origine of the asymmetrical properties of binaries.

During their slow motion from the poles to the equator, the masses forming the surface of the photosphere must be gradually cooled causing a lower temperature at the equator than at the poles. Formula 176 may be adopted to the discussion of this temperature, interpreting φ as the latitude and ω as the angular velocity of the sun, then

$$ds = R d\varphi$$

where R is the radius of the sun, and

$$\frac{dv}{dz} = \frac{v}{h}$$

Where h is the thickness of the overlying current of the upper circulating stratum.

Also since the circumferential velocity V_E is equal to ωR then equation 176 becomes

$$\frac{dT}{T} = \frac{2 V_E V \sin\varphi}{g h} d\varphi \qquad (1)$$

Integrating this equation from pole to equator, that is, from $\varphi = \frac{\pi}{2}$ to $\varphi = 0$, we get for a first approximation

$$\frac{\Delta T}{T} = \frac{2 V_E V}{g h}$$

Bjerknes applies this approximate formula to the earth and finds that the difference in temperature between pole and equator of the earth demanded by this equation is between 30° and 40° corresponding to a velocity in the upper troposphere between 11 and 15 m/sec. ~~Now if we multiply this by velocity by 10 we obtain the~~ We found that the approximate value for the eastward current on the sun ~~nearly~~ [13] V = 130

(1) In Bjerknes article in the A.P.J. 64 117 1926 a slight mistake is found. The correction is made here. Instead of cos φ used by Bjerknes, sin φ should be used. And instead of integrating from equator to pole the reverse should be done. However the results are the same, becuase it happens that Bjerknes mistake in integrating compensates for his first mistake.

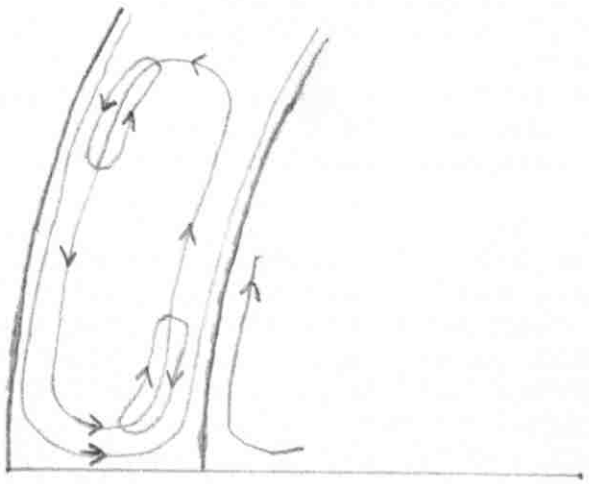

Fig 23

m/sec. Applying the previous formula to the sun where

V_E = 2000 m/se.

g = 275

V = 130 m/sec

T = 6000° c.

we get

$$\Delta T = \frac{12\,000\,000}{h}\,^{\circ}$$

This equation gives for h = 100 km a temperature difference of 120° and for h = 1000 km a temperature difference of 12° . The thickness of the upper stratum in the sun will be most probably between these two limits and therefore the difference in temperature between the pole and equator would be between 120 and 12 , a difference which is very difficult if not impossible to detect experimentally.

The occurrence of sunspots was accounted for by a zonal vortex, and the equatorial acceleration by the general circulation of the exterior part of the photosphere. Since the boundary layers of the photosphere cool while travelling from pole to equator, then they descend down and hotter gases rise upward, thus causing a concentration of circulation near the middle and lower latitudes, which is indicated by an inner circulating curve in figure 22. This is also favourable for the formation of the extra zonal vortices to which the sunspots were referred.

To explain the reversal of polarity, as well as all the previously explained phenomena, one has to admit the existence of two zonal vortices having opposite direction of rotation and are carried round by the general circulation between middle and lower latitudes (figure 23). From a thermal point of view this supposition is possible, because both directions of rotation are compatible with the thermal origine of vortex zones. Mechanically, vortices rotating like toothed wheels assist each other by their mutual frictional effect. Since the general circulation from pole to equator and back to the pole requires many years, then a period of 23 years for the revolution of the zonal vortices about each other does not seem unreasonable. Since we have sunspots in the two hemispheres having the same periodicity and showing opposite polarity, it may be assumed that the same assumptions for the hemisphere made above may be assumed for the second hemisphere because of frictional coupling of the two hemispheres.

Thus we see how Hale and Bjerknes have explained the relation of solar vortices to the physical properties of sunspots. They attributed the spot to a deep-seated vortex whose termini appear when cutting the photospheric surface and there they transfer angular momentum to the sun's atmosphere in such a manner as to produce "solar cyclones". The rotation of the ionized gas produces a separation of change giving rise to a magnetic field. The polarity of a spot, then depends upon the direction of rotation of the deep-seated vortex and therefore it was possible to correlate the magnetic sun-spot period with an internal circulation period in the manner proposed by Bjerknes and explained above.

however, this magnetic field will not account for the strong fields observed. The explanation of such fields was given by Ross Gunn whose investigations[1] of the motion of ions spiraling about an inhomogeneous magnetic field have shown that drift motions are imposed which are oppositely directed for the positive and negative ions. Under conditions of radial symmetry and a closed circuit, such as exists just outside a sunspot, a current flows which is in such a direction as to reduce the local inhomogeneity, but at the same time increases the total magnetic flux inclosed by the current circuit. The square of the computed fields are found to be propertional to a logarithmic function of the radius of the spot and the depth of the conducting layer. We shall not deal with such computations here but reference may be made to the articles by Gunn describing such phenomena.[2]

We have seen how the Bjerknes assumption of the vortex rings in the sun explains many of the points mentioned in the summary of the properites of sunspots in part II of this chapter. Before concluding this chapter few other properties which were not explained in our previous discussion will be explained now. The explanation is not necessarily final but it may be taken as a first trial.

It was mentioned before that the preceding spot has a larger motion in longitude than the following spot. This may be explained as a result of Newton's law of action and reaction. When the vortex cuts the surface of the photosphere the separation of the two termini will give the preceding spot a push along the direction of rotation of the sun, and the following along the opposite direction. It is interesting to compare this with the motion of the preceding spot in the exceptional case where there is a reversal in polarity due to a reversal of the sinuous direction of the part of the vortex which cuts the photosphere. Also it may be pointed out that if the necessary data be available, then a calculation of the dimensions of the vortex may be made using this explanation.

Also it was mentioned that the maximum frequency of spots is between the latitude $\pm$ 10° and $\pm$ 20°. This may be explained by the fact that the two vortex ring are radially situated with respect to the centre of the sun and they may repel each other causing the upper one to be nearer to the surface; also in its motion the vortex ring will be nearer to the surface at this latitude. However, another factor is worth mentioning here, and that is the pulsational character of the sun. It is observed, as was mentioned before, that the solar diameter is minimum when the solar activity is maximum. This may help the vortex ring to be nearer to the surface of the photosphere. Also this may be another reason why a zonal vortex should rise to the surface of the photosphere and cut it. But such a pulsation cannot be a cause to the production of spots as was suggested by Sussman.

---

(1) Physical Review 33, 832, 1929

(2) Physical review 33 614 1929
and A. P.J. 69 287 1929

Another interesting property is that of the sunspot activity curve. It was mentioned before that the ascent from minimum to maximum is steeper than the descent from maximum to minimum. This may be explained to be due to the equatorial accelerationwhich depends upon the distance from the equator.

Also it was mentioned above that the increase in area of the following spot is more rapid than that of the leading spot. This may be explained by the rotation of the ring around the axis of the sun.

The inclination of the axis of a sunspot group may also be explained by assuming a certain orientation of the zonal vortex ring, and that this orientation varies with the latituded as a result of the attraction and repulsion of the other vortex rings.

However, ther are certain phenomena which are not yet explanable by the Bjerknes theorem. For example it does not explain why the drop in temperature of a spot should be almost constant in all spots, and also the fact that the individual spots do not show a latitude displacement similar to the vortex rings. Thus Bjerknes explanation of the maintenance of a spot by energy of the other parts of the vortex that are pulled up to the surface is not in agreement with the above mentioned fact.

In conclusion it should be mentioned that the Bjerknes theorem, though appears to be only a working hypothesis and a forced explanation, yet it is the most acceptable explanation because it gives the explanation of a large diversity of phenomena that proved up to now to be unexplainable.

CONCLUDING REMARK:

On reading the previous chapters one feels a sense of disconnectedness. This is expected because the aim of this report is not to give an explanation to the sunspot phenomenon. The primary aim is to form a collection of the necessary subjects to form a basis for a comprehensive study of sunspots and other various solar phenomens. Thus a theory about the solar activity has to take into consideration the ionization, radiative equilibrium, and hydrodynamics of the sun. The general magnetic field is discussed because it may have a relation to the origin of the magnetic fields in sunspots.

A theory about sunspots should not only explain the sunspot phenomenon; but other solar phenomens are so closely related to sunspots that they demand a somewhat common origin. The need is, therefore, not for a "forced" explanation of a certain fact. The need is for a general solar theory or even to a new stellar model that takes into consideration the subjects mentioned above and gives an "unforced" explanation for the solar phenomens.

APPENDIX C. 1

Zonal Harmonics

Zonal Harmonics are a special case of the Legendre Polynomials. The Legendre's polynomials first arose from a consideration of the expansion of $(1 - 2xh + h^2)^{-\frac{1}{2}}$ in a series of ascending powers of h. Thus we have

$$(1 - 2xh + h^2)^{-\frac{1}{2}} = P_0(x) + h P_1(x) + h^2 P_2(x) + \cdots + h^n P_n(x) + \cdots \quad (1)$$

But From the binomial theorem if $(2|xh| - |h|^2) < 1$ we get

$$h(2x-h)]^{-\frac{1}{2}} = 1 + \tfrac{1}{2}h(2x-h) + \frac{1\cdot 3}{2\cdot 4}h^2(2x-h)^2 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}h^3(2x-h)^3 + \cdots$$

$$\cdots + \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}h^n(2x-h)^n + \cdots \quad (2)$$

Equating the coefficients of (1) and (2) we get for the nth function

$$P_n(x) = \frac{1\cdot 3\cdot 5\cdots 2n-1}{2\cdot 4\cdot 6\cdots 2n}\left[(2x)^n - \frac{2n}{2n-1}\frac{n-1}{1}(2x)^{n-2} + \frac{2n(2n-2)}{(2n-1)(2n-3)}\frac{(n-2)(n-3)}{1\cdot 2}(2x)^{n-4} - \cdots\right] \quad (3)$$

$$(x) = \frac{(2n-1)!!}{n!}\left[x^n - \frac{n(n-1)}{2(2n-1)}x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2\cdot 4\cdot(2n-1)(2n-3)}x^{n-4} - \cdots\right] \quad (3)$$

here

$(2n-1)!!$ stands for $(2n-1)(2n-3)(2n-5)\cdots 3\cdot 1$

$$(2n-1)!! = \frac{(2n)!}{2^n\, n!} \quad \therefore P_n(x) = \sum_{r=0}^{m}(-1)^r \frac{(2n-2r)!}{2^n\, r!(n-r)!(n-2r)!}x^{n-2r}$$

$m = \frac{1}{2}n$ if $n = 2l$; $m = \frac{1}{2}(n-1)$ if $n = 2l+1$

Introducing the factor into the parenthesis in (3) we get

$$(x) = \frac{(2n-1)!!}{n!}x^n - \frac{(2n-3)!!}{2\cdot(n-2)!}x^{n-2} + \frac{(2n-5)!!}{2\cdot 4\cdot(n-4)!}x^{n-4} - \cdots \quad (4)$$

If n is even the series will contain $\frac{1}{2}n+1$ terms while if n is odd the series contains $\frac{1}{2}(n+1)$ terms.

From equation (4) we can find the different functions of (x). An easy substitution for n will give

$P_0(x) = 1$

$P_1(x) = x$

$P_2(x) = \frac{1}{2}(3x^2 - 1)$

---

further reference the following two books from which the above is taken may be mentioned: "Analysis" by Phillips and "Modern Analysis" by Whittaker and Watson.

$P_3(x) = \frac{1}{2}(5x^3 - 3x)$

$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$

$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x)$

Therefore the Legendre polynomials take the form

$$+ hx + \frac{h^2}{2}(3x^2-1) + \frac{h^3}{2}(5x^3-3x) + \frac{h^4}{8}(35x^4-30x^2+3)$$
$$+ \frac{h^5}{8}(63x^5 - 70x^3 + 15x) + \cdots \qquad (5)$$

x may take any form of a variable; thus if we put for x, $\cos\theta$ we get what is called zonal harmonics. Then e polynomial will take the form

$$(\theta) = A P_0(\cos\theta) + B P_1(\cos\theta) + C P_2(\cos\theta) + D P_3(\cos\theta) + \cdots \qquad (6)$$

ut from (5) we set

$$\theta) = A + B\cos\theta + \frac{C}{2}(3\cos^2\theta - 1) + \frac{D}{2}(5\cos^3\theta - 3\cos\theta) + \cdots \qquad (7)$$

e constants $A, B, C, \ldots$ can be calculated in the same way that of the constants in the expansion of a fourier series. efore it remains to prove that any two Legendre polynomials $P_m(x)$ and $P_n(x)$ are orthogonal. To be able to prove this nother property of the polynomials will be considered.

drigue's formula for the Legendre Polynomials: -

It is clear that when n is an integer we have by the of Leibnitz theorem (1)

$$\frac{d^n}{dx^n}(x^2-1)^n = \frac{d^n}{dx^n}\left\{(x^2-1)(x^2-1)^{n-1}\right\}$$
$$= (x^2-1)\frac{d^n}{dx^n}(x^2-1)^{n-1} + 2nx\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^{n-1} + n(n-1)\frac{d^{n-2}}{dx^{n-2}}(x^2-1)^{n-2} \qquad (8)$$

we have

$$\frac{d^n}{dx^n}(x^2-1)^n = \frac{d^n}{dx^n}\left[\sum_{r=0}^{m}(-1)^r \frac{n!}{r!}(n-r)!\, x^{2n-2r}\right]$$

$$\frac{d^n}{dx^n}(x^2-1)^n = \sum_{r=0}^{m}(-1)^r \frac{n!}{r!(n-r)!}\frac{(2n-2r)!}{(n-2r)!}x^{n-2r}$$

} From binomial theorem (

here $m = \frac{1}{2}n$ if $n = 2p$ or $m = \frac{1}{2}(n-1)$ if $n = 2p+1$. From the general pression for $P_n(x)$ given by (4) we have

$$P_n(x) = \sum_{r=0}^{m}(-1)^r \frac{(2n-2r)!}{2^n r!(n-r)!(n-2r)!}x^{n-2r} \qquad (4')$$

eibnitz theorem is stated thus

$$\frac{d^n}{dx^n}[f(x)\cdot g(x)] = g(x)\frac{d^n}{dx^n}f(x) + C_n^{n-1}\frac{d}{dx}g(x)\frac{d^{n-1}}{dx^{n-1}}f(x) + C_n^{n-2}\frac{d^2}{dx^2}g(x)\frac{d^{n-2}}{dx^{n-2}}f(x) + \cdots$$
$$+ C_n^{n-r}\frac{d^{n-r}}{dx^{n-r}}g(x)\frac{d^{n-r}}{dx^{n-r}}f(x) + \cdots$$

re $C_n^m = \frac{n(n-1)(n-2)\cdots(n-m+1)}{m!} = \frac{n!}{m!(n-m)!}$

where $m = \frac{1}{2}n$ if n is even and $\frac{1}{2}(n-1)$ if n is odd.

Comparing (4') with (9) we easily notice that

$$P_n(x) = \frac{1}{2^n \cdot n!} \frac{d^n}{dx^n}(x^2-1)^n \quad *$$

This result is known as Ridrigue's formula.

The Orthogonality of Legendre's Polynomials (4) in the interval -1 to +1. We shall now show that

$$\int_{-1}^{+1} P_m(x)\, P_n(x)\, dx = \begin{cases} 0 & \text{when } m \neq n \\ \frac{2}{2n+1} & \text{when } m = n \end{cases} \qquad (10)$$

If $r \leq n$ then $\frac{d^r}{dx^r}(x^2-1)^n$ is divisible by $(x^2-1)^{n-r}$ and therefore if $r < n$ then $\frac{d^r}{dx^r}(x^2-1)^n$ vanishes when $x = -1$ or $+1$. Now of the two numbers m and n let m be equal to or more than n. Then integrating the integral (10) by parts continually we get by using the Ridrigue formula under the integration signs

$$\int_{-1}^{1} \frac{d^m}{dx^m}(x^2-1)^m \cdot \frac{d^n}{dx^n}(x^2-1)^n dx = \left[\frac{d^{m-1}}{dx^{m-1}}(x^2-1)^m \frac{d^n}{dx^n}(x^2-1)^n\right]_{-1}^{+1} - \int_{-1}^{+1} \frac{d^{m-1}}{dx^{m-1}}(x^2-1)^m \frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n \cdot dx$$

$$= \cdots \cdots \cdots \cdots$$

$$= (-1)^m \int_{-1}^{+1} (x^2-1)^m \frac{d^{n+m}}{dx^{n+m}}(x^2-1)^n \quad (11)$$

This is because $\frac{d^r}{dx^r}(x^2-1)^q$ vanishes at $x = \pm 1$

Now, also when $m > n$ then $\frac{d^{m+n}}{dx^{m+n}}(x^2-1)^n = 0$ because the differential coefficients of $(x^2-1)^n$ of order higher than $2n$ vanish, and therefore when $m > n$ or $m+n > 2n$ we have

$$\int_{-1}^{+1} P_m(x)\, P_n(x)\, dx = 0 \qquad \text{for } m > n$$

When $m = n$ by (11) we have

$$\frac{1}{2^n \cdot (n!)^2} \int_{-1}^{+1} \frac{d^n}{dx^n}(x^2-1)^n \frac{d^n}{dx^n}(x^2-1)^n dx = (-1)^n \int_{-1}^{+1} (x^2-1)^n \frac{d^{2n}}{dx^{2n}}(x^2-1)^n dx$$

changing the variable x into $\cos\theta$ we get $= (2n)! \int_{-1}^{+1} (1-x^2)^n dx = 2 \cdot (2n)! \int_0^{+1} (1-x^2)^n dx$

$$\int_{-1}^{+1} P_n^2(x)\, dx = \left[2 \cdot (2n)! \int_0^{\frac{\pi}{2}} \sin^{2n+1}\theta \, d\theta\right] \left[\frac{1}{2^n \cdot (n!)^2}\right]^2 = \frac{2 \cdot (2n)!}{(2^n \cdot n!)^2} \frac{(2^n \cdot n!)^2}{(2n+1)!} = \frac{2}{2n+1}$$

$\therefore$ Equations (10) are proved.

From (10) we see that the Legendre Polynomials are orthogonal and we can find the coefficients of any expansions in terms of them just by multiplying (5) or (7) by $P_n(x)$ where $n = 0, 1, 2, 3$, etc. . successively and then integrating in each case

APPENDIX C. 2

Appendix.

Zonal Harmonics[1]

Zonal Harmonics are a special case of the legendre Polynomials. The legendre's Polynmials first arose from a consideration of the expansion of $(1 - 2xh + h^2)^{-\frac{1}{2}}$ in a series of ascending powers of h. Thus we have:

$$(1 - 2xh + h^2)^{-\frac{1}{2}} = P_0(x) + hP_1(x) + h^2 P_2(x) + \cdots + h^m P_m(x) + \cdots \quad (1)$$

But from the binomial theorem if $(2|xh| - |h|^2) < 1$ we get

$$[1 - h(2x - h)]^{-\frac{1}{2}} = 1 + \frac{1}{2}h(2x-h) + \frac{1\cdot 3}{2\cdot 4}h^2(2x-h)^2 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}h^3(2x-h)^3 + \cdots$$

$$\cdots + \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}h^n(2x-h)^n + \cdots \quad (2)$$

Equating the coefficients of (1) and (2) we get for the n th function

$$P_n(x) = \frac{1\cdot 3\cdot 5\cdots 2n-1}{2\cdot 4\cdot 6\cdots 2n}\left[(2x)^n - \frac{2n}{2n-1}\cdot\frac{n-1}{1}(2x)^{n-2} + \frac{2n(2n-2)}{(2n-1)(2n-3)}\frac{(n-2)(n-3)}{1\cdot 2}(2x)^{n-4} \cdots\right] \quad (3)$$

$$P_n(x) = \frac{(2n-1)!!}{n!}\left[x^n - \frac{n(n-1)}{2(2n-1)}x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2\cdot 4\cdot(2n-1)(2n-3)}x^{n-4} - \cdots\right] \quad (3)$$

where $(2n-1)!!$ stands for $(2n-1)(2n-3)(2n-5)\cdots\cdot 3\cdot 1$

or $$(2n-1)!! = \frac{(2n)!}{2^n n!}$$

Introducing the factor into the parenthesis in (3) we get

$$P_n(x) = \frac{(2n-1)!!}{n!}x^n - \frac{(2n-3)!!}{2\cdot(n-2)!}x^{n-2} + \frac{(2n-5)!!}{2\cdot 4\cdot(n-4)!}x^{n-4} - \cdots \quad (4)$$

If n is even the series will contain $\frac{1}{2}n+1$ terms while if n is odd the series contains $\frac{1}{2}(n+1)$ terms.

From equation (4) we can find the different functions $P_n(x)$. An easy substitution for n will give:

---

(1) for further reference the following two books from which the above is taken may be mentioned: "Analysis" by (Philips) and "Modern Analysis" by (Whittaker and Watson.).

$P_0(x) = 1$

$P_1(x) = x$

$P_2(x) = \frac{1}{2}(3x^2-1)$

$P_3(x) = \frac{1}{2}(5x^3-3x)$

$P_4(x) = \frac{1}{8}(35x^4-30x^2+3)$

$P_5(x) = \frac{1}{8}(63x^5-70x^3+15x)$

. . . . . .

Therefore the Legendre series takes the form:-

$$1+hx+\frac{h^2}{2}(3x^2-1)+\frac{h^3}{2}(5x^3-3x)+\frac{h^4}{8}(35x^4-30x^2+3)+\cdots \quad (5)$$

x may be any variable or any function of a variable. A special case is that when $x = \cos\theta$. Then we obtain what is called Zonal Harmonics. The Polynomial will take now the form:

$$f(\theta) = AP_0(\cos\theta) + BP_1(\cos\theta) + CP_2(\cos\theta) + \cdots \quad (6)$$

But from (5) we get:

$$f(\theta) = A + B\cos\theta + \frac{C}{2}(3\cos^2\theta-1) + \frac{D}{2}(5\cos^3\theta-3\cos\theta) + \cdots \quad (7)$$

The Constants A, B, C, ... may be calculated in the same way as that of the Constants in the expansion in terms of a Fourier series. Therefore it remains to prove that any two legendre polynomials $P_m(x)$ and $P_n(x)$ are orthogonal. To prove this, another property of the polynomials will be mentioned below:

(#) Rodrigues' formula for the Legendre polynomials:-

It is clear that when n is an integer we have by using the Leibnitz theorem(1): $\frac{d^n}{dx^n}(x^2-1)^n = \frac{d^n}{dx^n}\left\{(x^2-1)(x^2-1)^{n-1}\right\}$

$$= (x^2-1)\frac{d^n}{dx^n}(x^2-1)^{n-1} + 2nx\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^{n-1} +$$

$$+ n(n-1)\frac{d^{n-2}}{dx^{n-2}}(x^2-1)^{n-1} \quad (8).$$

(1) Leibnitz theorem is stated in the following formula:

$$\frac{d^n}{dx^n}\left[f(x)\cdot g(x)\right] = g(x)\frac{d^n}{dx^n}f(x) + C_n^{n-1}\frac{d}{dx}g(x)\frac{d^{n-1}}{dx^{n-1}}f(x) + \frac{C_n^{n-2}}{n}\frac{d^2}{dx^2}\cdot$$

$$+ C_n^{n-3}\frac{d^2}{dx^2}g(x)\frac{d^{n-2}}{dx^{n-2}}f(x) + \cdots$$

where: $C_n^r = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$

Also we have

$$\frac{d^n}{dx^n}(x^2-1)^n = \frac{d^n}{dx^n}\left[\sum_{r=0}^{n}(-1)^r \frac{n!}{r!}(n-r)!\, x^{2n-2r}\right]$$

or by using the binomial theorem

$$\frac{d^n}{dx^n}(x^2-1)^n = \sum_{r=0}^{m}(-1)^r \frac{n!}{r!(n-r)!}\frac{(2n-2r)!}{(n-2r)!}x^{n-2r} \qquad (9)$$

where $m = \frac{1}{2}n$ if $n = 2p$ $\quad m = \frac{1}{2}(n-1)$ if $n = 2p+1$

From the general expression for $P_n(x)$ given by (4) we have:

$$P_n(x) = \sum_{r=0}^{m}(-1)^r \frac{(2n-2r)!\, x^{n-2r}}{2^n\, r!\,(n-r)!\,(n-2r)!} \qquad (4')$$

where also $m = \frac{1}{2}n$ or $\frac{1}{2}(n-1)$ as above.

Comparing (4') with (9) we easily notice that:

$$P_n(x) = \frac{1}{2^n . n!}\frac{d^n}{dx^n}(x^2-1)^n$$

This result is known as Rodrigues' formula.

i) The Orthogonality of Legendre's polynomials (4) in the interval -1 to $+1$. We shall now show that:

$$\int_{-1}^{+1} P_m(x)\,P_n(x)\,dx = \begin{cases} 0 \text{ when } m \neq n \\ \dfrac{2}{2n+1} \text{ when } m = n \end{cases} \qquad (10)$$

if $r \leq n$ then $\frac{d^r}{dx^r}(x^2-1)^n$ is divisible by $(x^2-1)^{n-r}$ and so if $r < n$, $\frac{d^r}{dx^r}(x^2-1)^n$ vanishes when $x = 1$ or $x = -1$

Now of the two numbers m and n let m be equal to or greater than n. Then integrating by parts continually, the above integral (10) with the help of the Rodrigues' formula becomes:

$$\int_{-1}^{+1} P_m(x)\,P_n(x)\,dx = \int_{-1}^{+1}\frac{d^m}{dx^m}(x^2-1)^m \cdot \frac{d^n}{dx^n}(x^2-1)^n\,dx$$

$$= \left[\frac{d^{m-1}}{dx^{m-1}}(x^2-1)^m \frac{d^n}{dx^n}(x^2-1)^n\right]_{-1}^{+1} - \int_{-1}^{+1}\frac{d^{m-1}}{dx^{m-1}}(x^2-1)^m \frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n\,dx$$

. . . . . .

$$= (-1)^m \int_{-1}^{+1}(x^2-1)^m \frac{d^{n+m}}{dx^{n+m}}(x^2-1)^n\,dx \qquad (11)$$

because $\frac{d^r}{dx^r}(x^2-1)^2$ vanishes at -1 and $+1$

Now when $m > n$ then $\frac{d^{m+n}}{dx^{m+n}}(x^2-1)^n = 0$ because the differential coefficients of $(x^2-1)^n$ of order higher than $2n$ vanish; and therefore when m is greater than n or $m+n > 2n$ we have.

$$\int_{-1}^{+1} P_m(x) P_n(x) = 0 \quad \text{for } m > n$$

when $m = n$ by (11) we have.

$$\left[\frac{1}{2^n \cdot (2n)!}\right]^2 \int_{-1}^{+1} \frac{d^n}{dx^n}(x^2-1)^n \frac{d^n}{dx^n}(x^2-1)^n dx = (-1)^n \int_{-1}^{+1} (x^2-1)^n \frac{d^{2n}}{dx^{2n}}(x^2-1)^n dx$$

$$= (2n)! \int_{-1}^{+1} (1-x^2)^n dx \quad (1)$$

Changing the variable so that $\cos\theta = x$ we get

$$\int_{-1}^{+1} P_n^2(x)\, dx = \left[2 \cdot (2n)! \int_0^{\frac{\pi}{2}} \sin^{2n+1}(\theta)\, d\theta\right]\left[\frac{1}{2^n \cdot (2n)!}\right]^2$$

by the use of Γ-functions or by integration by parts we get:

$$\left\{\frac{1}{2^n \cdot (2n)!}\right\}^2 \int_{-1}^{+1} P_n^2(x)\, dx = \left[2 \cdot (2n)! \frac{2 \cdot 4 \cdot 6 \cdots (2n)}{3 \cdot 5 \cdot 7 \cdots (2n+1)}\right]\left[\frac{1}{2^n \cdot (2n)!}\right]^2$$

$$= \frac{2 \cdot (2n)!}{(2^n \cdot n!)^2} \frac{(2^n n!)^2}{(2n+1)!} = \frac{2}{2n+1}$$

∴ Equations (10) are proved. From this we see that the Legendre Polynomials and the zonal Harmonics are orthogonal sets. We may expand certain functions by them and compute the coefficients of expansion by multiplying (5) or (7) by $P_n(x)$ or $P_n(\cos\theta)$ where $n = (0 \text{ or } 1 \text{ or } 2 \text{ or } 3) \ldots$ and then integrating.

(1) This is obtained by integration by parts