THE SUN'S MAGNETIC FIELD[1]

By H. W. Babcock

Mount Wilson and Palomar Observatories, California Institute of Technology, Carnegie Institution of Washington

The study of magnetic fields on the surface of the sun has been greatly advanced within the last decade by sensitive new instrumental techniques. The existence of a "general," or main poloidal magnetic field has been established. This field is limited to the sun's polar caps, and it reversed its polarity near the time of the last sunspot maximum. The development of transitory, low-latitude, localized bipolar magnetic fields that are responsible for "centers of activity" has been followed at some length. As a result of observation, it has become clear that not only sunspots, but other optically observable features such as calcium flocculi, faculae, flares, prominences, and coronal brightenings are fundamentally related to the sun's magnetic field and its variations. The idea is well developed that internal amplification of the main dipolar field, up to the stage of instability, results in the outbreak of relatively strong, localized flux loops and in the formation of bipolar magnetic regions on the surface in low latitudes. If this is correct, it follows that probably all of the observable magnetic phenomena are manifestations of the same "general" magnetic field. By developing the logical consequences of emergence of magnetic flux from the progressively expanding bipolar regions, and by considering the underlying systematics of solar activity, it has been possible to formulate descriptive and semiquantitative theories of the 22-year magnetic cycle.

Magnetohydrodynamics is evidently the governing discipline within which the solar cycle as well as most of the specialized physical phenomena occurring at or near the surface of the sun must be discussed. But because of the complexity of the phenomena, advances have depended heavily on observation, and no doubt will continue to do so. It is therefore appropriate to review the instrumental developments that make possible the measurement of solar magnetic (and velocity) fields, particularly as the limitations of present-day observations need to be appreciated.

Techniques and Instruments

All quantitative measurements of magnetic fields at the sun's surface depend on the Zeeman effect in the spectrum. Even when a sensitive line—one having a wide Zeeman pattern—is used, a field intensity of some 20 kilogauss would be required to yield a displacement of 1 Å between the oppositely polarized outlying σ components. As the normal half-width of a Fraunhofer line is about 0.1 Å, typical sunspot fields (2-3 kilogauss) give line-displacements that are readily visible. Therefore, with a telescope and spectrograph of adequate size and dispersion, and with a suitable analyzer

[1] The survey of literature pertaining to this review was concluded in October 1962.

for polarization, it is a straightforward matter to make visual determinations of sunspot polarity and to estimate the field intensity, as was first done by Hale (1) in 1908. Daily magnetic observations of sunspots for the interval 1917–1924 have been given by Hale & Nicholson (2), with a description of the apparatus.

The subject of the measurement and spatial analysis of magnetic fields in sunspots has recently been reviewed by Treanor (3), who has introduced a new method of determining the direction of the magnetic vector within the umbra.

Hale's pioneering work on the magnetic fields of sunspots stimulated interest in the far more difficult problem of detecting the "general" magnetic field of the sun, the existence of which had been inferred long before from the coronal streamers. Beginning with the construction of the 150-foot tower telescope on Mount Wilson in 1910, a great deal of effort was expended on this problem by Hale and his associates as well as by solar physicists in other countries. Various methods were tried: measurements on spectrograms and on interferograms made with a Lummer plate, visual measurements on the optical spectrum, and dc photoelectric measurements. Because of the weakness of the field and of technological inadequacy, these and other early attempts were marked by no clear success. They have been reviewed by Babcock & Cowling (4).

Most of the new data relating to weak solar magnetic fields since 1952 have been acquired by means of the solar magnetograph (5). This is a sensitive photoelectric instrument, built around a solar telescope and an efficient grating spectrograph, that is capable of measuring Zeeman shifts as small as 0.0005 of the half-width of a Fraunhofer line. Such a shift results from a field of a few tenths of one gauss. Important contributions to early magnetograph design were made by Kiepenheuer (6).

A complete magnetograph comprises two systems. The first is a detector, or "gaussmeter," that yields an output voltage which is the electrical analogue of the field strength of the part of the sun's image entering the slit. The second system of the magnetograph combines the scanning and recording functions, thus providing for presentation of the results in the form of a "magnetogram" that shows the distribution, intensity, and polarity of the field.

Magnetographs are generally designed to respond to the line-of-sight component of the field, and thus measure the minute shift of the line as the analyzer for circular polarization, in front of the spectrograph slit, oscillates between the right- and left-hand states. The fact that the line-shift, as seen through a circular analyzer, is proportional to the sight-line component of the field was shown by Seares (7). The sensitivity is achieved through design principles that provide a favorable signal-to-noise ratio and freedom from spurious effects. It is also necessary that the optical efficiency be good, so that the information rate is adequate to permit fairly rapid scanning with sufficient angular resolution (say from 20″ down to 5″ or less). This efficiency is provided in part by use of a large, blazed, plane grating that yields distortion-free line profiles, with linear dispersion of about 1 Å per cm.

Some of the other essential features of the solar magnetograph are:

(*a*) An electrically excited crystal of ammonium dihydrogen phosphate, which is the heart of the analyzer. The crystal becomes birefringent on the application of a voltage to its transparent electrodes (8). This device has the great advantage that it is motionless, and so is free of undesirable false modulation that almost invariably seems to accompany the use of rotating optical elements. Together with a fixed Nicol prism or Polaroid, the ammonium dihydrogen phosphate crystal constitutes an oscillating circular analyzer. For greatest efficiency, the exciting voltage has a square wave form.

(*b*) The analyzer-amplifier system is of the "lock-in" type, which has proved so useful in many other applications where a weak signal must be measured in the presence of noise. In this case the noise arises from the shot effect of the photocurrent. The circular polarization in the line profile—the component of interest in the incoming light—is modulated by the optical analyzer that oscillates at a fixed frequency. The resulting oscillation in the position of the chosen spectrum line is then detected by two differentially connected multiplier phototubes that receive light from exit slits on opposite sides of the line profile where the intensity gradient is steepest. After amplification by a differential amplifier, the ac signal is synchronously demodulated and filtered. This system provides a very narrow band-width at the modulation frequency, which results in a favorable signal-to-noise ratio. The output is a varying dc voltage that is the analogue of the sight-line component of the magnetic field under observation.

(*c*) Means are provided for ensuring that the chosen spectrum line remains centered between the two exit slits during scanning, thus compensating for the Doppler shift that results from the sun's rotation. This can be accomplished by a tilting plate that is either coupled to the scanning drive, or that is servo-driven to maintain equality of the dc output of the two photomultipliers.

Additional features that, while not absolutely essential, contribute to the efficiency of operation include an "image slicer" according to the design of Bowen (9). The slicer is located just ahead of the entrance slit of the spectrograph. This optical device accepts light from a small rectangular aperture in the image plane and redistributes it by means of prisms along the 11-cm length of the slit, thus increasing the usable light by an order of magnitude. Further light efficiency is gained by accepting light from not one but the two magnetically sensitive spectrum lines. Double exit slits are used for lines λ 5247 (Cr I) and λ 5250 (Fe I). Field lenses in the form of off-axis strips cut from a thin plastic Fresnel lens are placed behind the slits to direct the light to the appropriate phototubes.

The magnetograph installed at the 150-foot tower telescope can be used to scan the sun's disk in a raster of about 80 parallel traces with a resolution of 23″. The complete scan is carried out automatically and requires about one hour. Alternatively, a fine-scan arrangement, as used by Howard & Babcock (10), permits scanning an area of about $5' \times 6'$ in about 15 minutes, with a scanning aperture of 5″ or 10″. The recording is accomplished by a camera that makes a time exposure of a cathode-ray tube which is intensity-modulated in response to the varying magnetic signal while the recording spot is deflected in a manner that conforms to the scanning of the sun's optical image. Magnetic polarity is indicated by the slant to right or left of the elongated recording spot on the cathode-ray tube, while abrupt changes in the intensity or form of the trace, occurring through the action of a signal-level discriminator, are made to occur at several preselected levels ranging from 1 to 60 gauss. The result is equivalent to a magnetic contour map of seven levels. An example of such a magnetogram is shown in Figure 1.

Magnetographs are sometimes used in the "Doppler" rather than the "Zeeman mode" for recording velocity fields. This requires only the insertion of a fixed circular polarizer ahead of the analyzer. Very high sensitivity, of the order of 50 m/sec, is thus obtained for differential velocities (5).

The original magnetograph that was developed at the Hale Solar Laboratory in Pasadena has been superseded by the improved instrument on Mount Wilson. Other magnetographs, similar in principle but differing in recording capability, are in use in Cambridge, England; at the Crimean Astrophysical Observatory, USSR; at the Institute of Terrestrial Magnetism, Ionosphere, and Radio Propagation of the Academy of Sciences of the USSR (11); and at Freiburg, Germany (12). Additional instruments are reported to be under development by Kiepenheuer at Capri; by Zirin at the High Altitude Observatory, Colorado; by Gaizuskas at Ottawa, Canada; and by Cimino at Rome.

The Crimean magnetograph, as developed by Nikulin, Severny & Stepanov (13), has been used mostly for fine-scans of selected regions. Recording is done largely on strip-charts, from which the data are transcribed manually to produce magnetic contour maps. The instrument has also been adapted to measurement of the transverse Zeeman effect. It has an automatic device for centering the spectrum line on the double exit slit.

Solar magnetic fields can be mapped with superior resolution although with less sensitivity by means of a photographic technique developed by Leighton (14) at Mount Wilson. For this, a spectroheliograph is fitted with an optical beam splitter and fixed analyzers for right- and left-hand circular polarization. Two monochromatic photographic images are obtained by means of exit slits placed on the side of a line profile where the intensity gradient is steep. By image superposition and a photographic subtraction process, a gray-tone picture is obtained, with parts darker and lighter than the average representing areas of one polarity or the other. The method is limited to fields stronger than about 30 gauss, and the zero point may be somewhat indefinite. It has proved especially valuable in showing the detailed correspondence between local, small-scale enhancements in the magnetic pattern and brightening in Ca II plages, as seen in Figure 2.

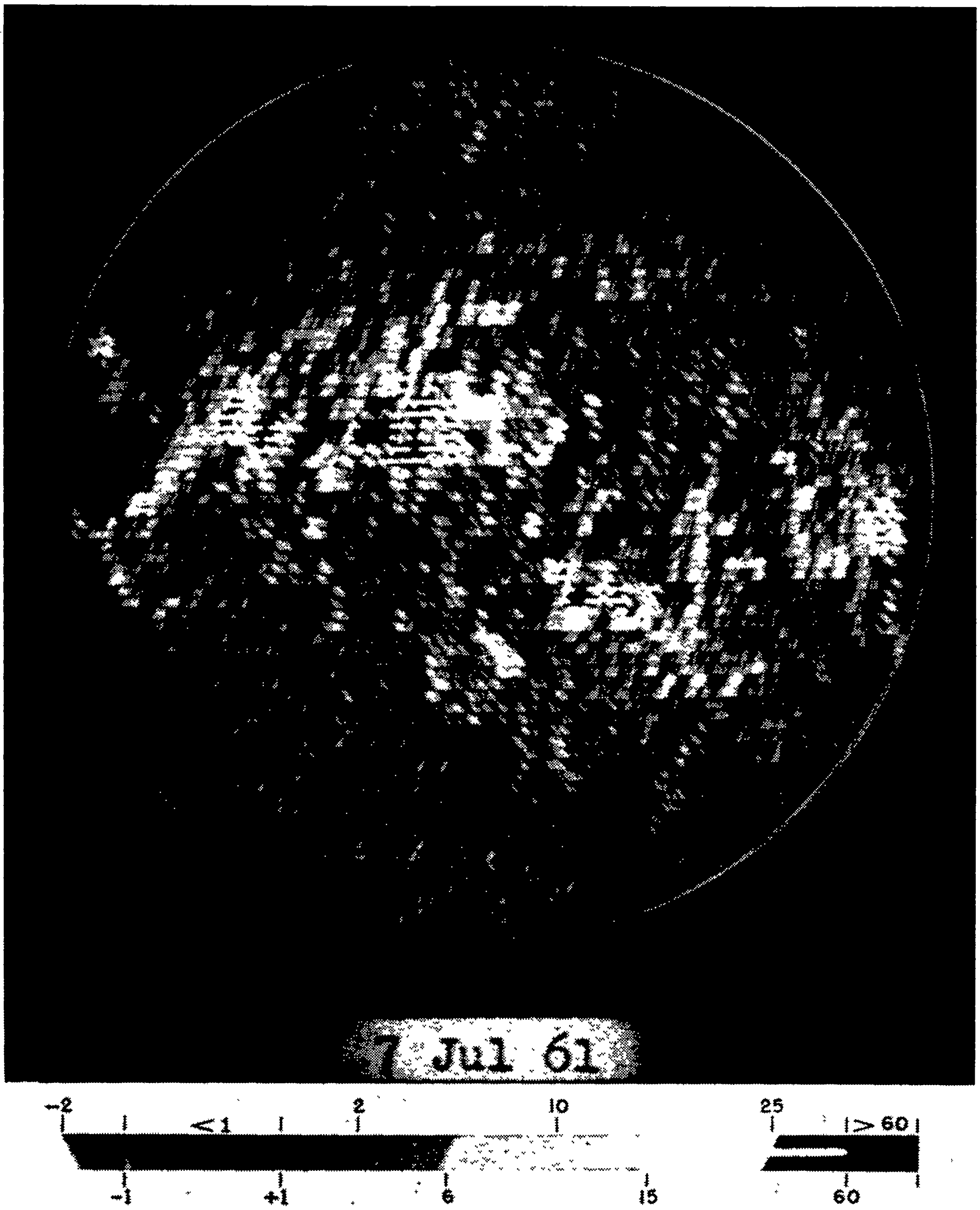

FIG. 1. Magnetogram of the sun obtained at Mount Wilson. Magnetic polarity is indicated by slant of the elongated spot to right (+) or left (−), while field strength at seven levels (as shown by calibration strip below) is given by changes in the intensity or form of the spot. Note the evidence for weak fields near the sun's poles, as well as for strong bipolar regions in low latitudes.

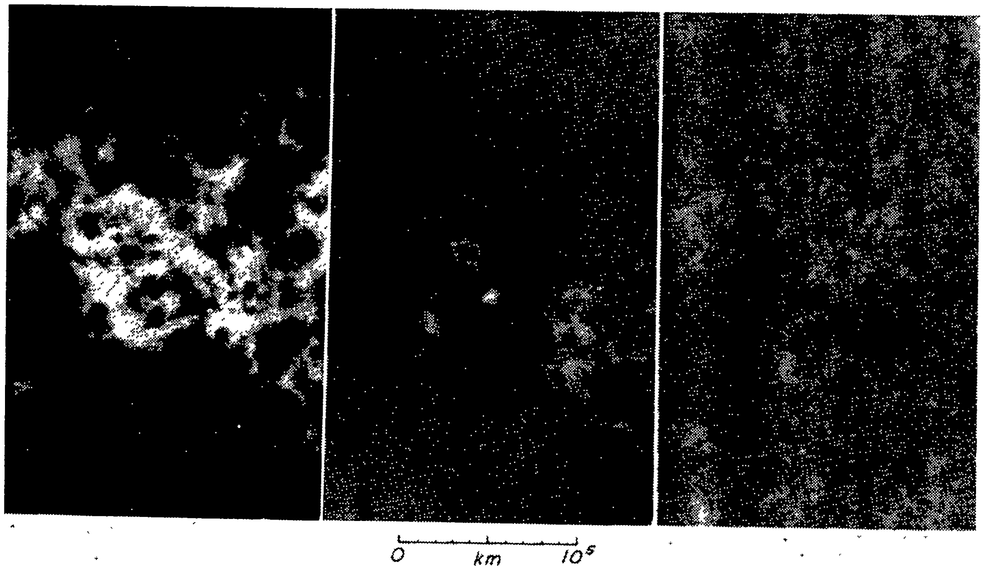

FIG. 2. Plates by R. B. Leighton showing Ca II emission (left), magnetic fields (center), and direct photograph (right) of a spot group on September 16, 1958.

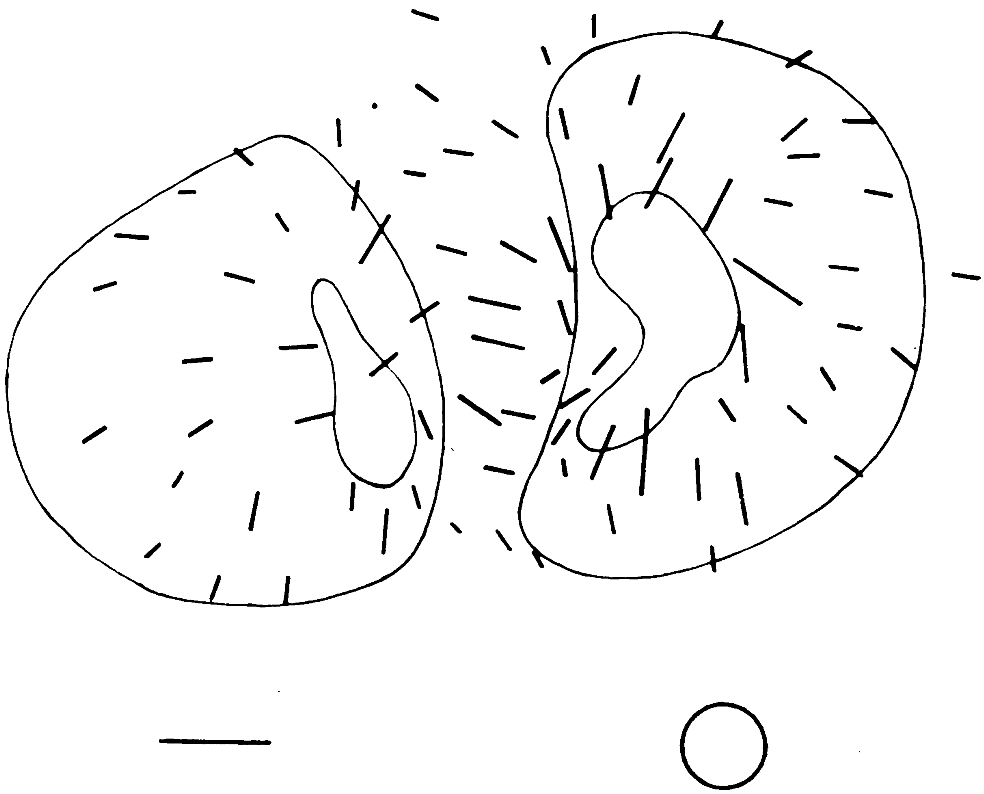

FIG. 3. Transverse component of the magnetic field around a pair of sunspots, as observed by A. Dollfus (Observatoire de Meudon). The length of the calibration line to the lower left corresponds to 1000 gauss. The circle has a diameter of 5″.

Weak transverse magnetic fields in the photosphere, in the near vicinity of sunspots, have been mapped by Dollfus (15) and by Leroy (16). An example is shown in Figure 3. In most cases the transverse component was in the range 100–1000 gauss. The method depends only indirectly on the Zeeman effect, and no spectrograph is used. A very sensitive detector for plane polarization is employed on the sun's image at the focus of a telescope. Both azimuth and amplitude of polarization are measured within a broad spectral region containing many rather strong absorption lines. The polarization arises from the tendency toward saturation of all components in a Zeeman pattern, where, if the line were very weak or in emission, the simple intensity rules for the components of the pattern would apply. The saturation of some or all of the components in absorption in a transverse field results in nonconservation of polarization for the whole pattern, as was pointed out by Lorentz (17). The position angle of the plane of polarization (electric vector) that results is parallel to the magnetic field H. Leroy (16) has derived a relationship between field intensity and degree of polarization, showing that, if $H=1000$ gauss, the predicted polarization approaches 1 per cent in the solar spectrum for the region between λ 4400 and λ 4900.

Results

Poloidal field.—Definite and reproducible evidence for the sun's main poloidal field was obtained with the Pasadena magnetograph in 1952–53 by Babcock & Babcock (18, 19). The north polar cap ($\phi > 55°$) showed consistently a weak field of positive polarity; the opposite polarity was found near the sun's south pole. (By definition a field of positive polarity is taken to be one for which the magnetic vector points out of the sun, or, in general, toward the observer.) The boundary in heliographic latitude was irregular and variable, while the field distribution showed considerable fine structure that changed from day to day. There was no evidence of any persistent obliquity between the magnetic and the rotational axes.

As it is the sight-line component of the field that is measured, and as there is generally a rather high obliquity γ of the line of sight to the lines of force emanating from the sun's polar caps, it is difficult to estimate the actual field intensity with precision. Variations in γ at the photospheric level can cause changes in the sight-line component. This effect probably is responsible for some of the day-to-day changes in the magnetic pattern. Furthermore, the obliquity of the sun's axis to the ecliptic (7°) results in a seasonal effect that generally favors the field of one polar cap and suppresses the other. The cap that is tilted toward the earth not only presents a larger area, but the inclination of the lines of force is more favorable for it.

Because of the weakness of the sun's polar fields, their study may require an unusually slow rate of scan and an exceptionally narrow band-width. These features enhance the difficulty of the measurements, but the existence of the "general" field and the order of magnitude of its mean intensity—about 1 gauss—have been established by a long series of magnetograms (or gaussmeter readings) extending over many years. An order-of-magnitude estimate of 8×10^{21} maxwells was made for the total magnetic flux associated with the polar caps, based on numerical integration of the field distribution on several magnetograms. After the report of these results in 1955, the observations were continued by Harold D. Babcock until about 1959.

The mean intensity of the polar fields appeared to decline somewhat after 1954, although no close control of the intensity calibration was possible. Then, in 1957, the field of the south polar cap reversed its polarity. A similar reversal occurred in the north, but not until 1958. For nearly 18 months, the same polarity (positive) was shown at both poles. This striking reversal of the main dipolar field, discovered by H. D. Babcock (20), took place near the phase of maximum sunspot activity in the solar cycle. As Waldmeier (21) has pointed out, if the northern and southern hemispheres are considered separately, the sunspot numbers reached a maximum in the south about one year earlier than in the north, and this suggests a physical connection with the earlier reversal of the south polar field.

Since 1959, the poloidal field has been under observation less frequently and with less consistent calibration. Generally, the same polarity has prevailed since the reversal of 1957–1958, but on some occasions it has been impossible to obtain sure evidence for the very weak field at one pole or the other. It is possible, although not firmly established, that the field is weaker now than in the corresponding years of the preceding cycle before it underwent reversal. The possibility is not excluded that the sun's poloidal field has a persistent, nonreversing component, smaller than the varying component.

The reversal of polarity at sunspot maximum strongly suggests that a close connection exists between the poloidal field and the sunspot cycle, with its 22-year magnetic period. It is a plausible supposition, awaiting confirmation, that a reversal of the main field occurs at about the time of maximum sunspot activity in every cycle. Presumably the reversal comes about through a pattern of circulation that shifts material, with its associated magnetic field, near the sun's surface, for it was well established by Cowling (22) that an impossibly long time would be required for diffusion of the lines of force on a large scale through the highly conducting plasma in the main body of the sun in such a way as to effect a major change.

Confirmatory measurements of the existence, order of magnitude, and polarity of the sun's main poloidal field have been made by von Klüber & Beggs (23) with the Cambridge magnetograph. They found, in 1959, that the field strength was around one gauss near each pole of the sun, and that the polarity was as reported by H. D. Babcock for the same year. These results, like those obtained in Pasadena and on Mount Wilson, show that the persistent fields are generally found only in latitudes greater than about 60°. Von Klüber reported that, during 1961, " . . . the field near the sun's north pole continued to be of south (negative) polarity, while at the sun's south pole the polarity has been variable."

Additional information on the structure and strength of the sun's magnetic field high in the corona has been obtained by Högbom (24), who analyzed the scattering of radio waves from Taurus A when it was nearly occulted by the sun. On plausible assumptions, he estimated the surface field strength at the sun's poles to be about 1.5 gauss. Indications were that the field lines in the high corona were nearly radial.

Low-latitude fields.—The number and strength of localized magnetic regions in low heliographic latitudes varies greatly with the phase of the solar cycle, but some are almost always observable with the magnetograph, even when no sunspots are to be seen. On only a very few occasions in 1955 (at the minimum of the solar cycle) were no such regions found. A summary of observations of low-latitude magnetic regions and of their relationship to optically observable activity has been given for the years 1952–1954 by Babcock & Babcock (18).

Typically, these transitory magnetic regions are bipolar, and their main field strength, integrated over the area of the scanning aperture, may vary from a fraction of a gauss upward to the fields of 2 or 3 kilogauss found in sunspot umbrae. There is much diversity in size, duration, field intensity, and total magnetic flux, but typically the latter is of the order of 10^{21} maxwells. The bipolar magnetic regions are initially compact, and produce most of their "activity" while they are young. The regions almost invariably expand progressively as they become older. The total magnetic flux of the bipolar region reaches a maximum while the region is yet compact; the lines of force, as indicated by coronal features, arch higher into the corona, but remain rooted in the respective positive and negative parts of the bipolar magnetic region as long as it is identifiable. Within the rather crude observational limits, there appears to be equality of positive and negative flux, as demanded by theory (div $\mathbf{H}=0$). Calcium flocculi, faculae, spots, flares, and active prominences appear in the early stages of the bipolar magnetic region as the field strength grows and usually while it is rather compact. Then, with expansion of the magnetic region and consequent reduction of field intensity, evidences of activity disappear in reverse order. Stable prominences may persist until the field strength has fallen to quite low values, and, in general, the weak field remains after direct optical evidence for the active region has disappeared. A comparison of the magnetograms with other observations of solar activity emphasizes the fundamental nature of the magnetic field.

As bipolar magnetic regions grow in age and in area, their boundaries usually become ill defined and they frequently merge with adjacent magnetic regions; for these reasons, quantitative measures of total magnetic flux and of area become quite difficult. The difficulty is enhanced by foreshortening of the areas and changing obliquity of the lines of force as the sun rotates.

A few of the bipolar magnetic regions have large areas and weak fields throughout their entire existence and do not seem to produce any pronounced optical activity. It would seem, then, that sunspot groups, active regions, and centers of activity are subordinate to, and refer to phenomena of shorter duration than the magnetic regions themselves. For this reason, the terms bipolar magnetic region, multipolar magnetic region, and unipolar magnetic region are sometimes useful.

Far more infrequent are unipolar magnetic regions. The best example, which persisted for at least six solar rotations, was observed in 1953. It had a total flux of 1.8×10^{21} maxwells and an area of about one sixth of the sun's disk. Such unipolar magnetic regions may be the relics of bipolar regions, of which the preceding or the following part has expanded past the point of recognition. The lines of force above the unipolar magnetic region must have a great radial extent from the sun. A good correlation was noted between the central meridian passage of the best unipolar region and a 27-day sequence of geomagnetic disturbances (with a phase lag of 2–3 days). There was also a correlation with times of maximum primary cosmic ray intensity (25). Thus, there is reason to believe that unipolar magnetic regions can be identified with the hypothetical "M" regions. These are not to be confused with the sources of much more intense and sporadic geomagnetic storms, which are usually chromospheric flares occurring close to spots in bipolar or multipolar magnetic regions.

A virtual one-to-one correspondence between bright calcium flocculi or plages and strong bipolar or multipolar magnetic regions was found on the early magnetograms of 1952–1954 (18). At that time a long slit (45″ in length, or about 40,000 km on the sun) was in use, and plages were noted wherever the average field was 2 gauss or more. Because the plages have a more or less filamentary fine structure, probably conforming to local condensations or groupings of the lines of force, it is likely that the actual field required to produce the bright features of the plages is significantly greater than 2 gauss. The brightness has been found to increase with field strength up to $H = 70$ gauss, above which it decreases again, according to Stepanov & Petrova (26). More detailed comparisons of the field distribution with the structural details of plages have been made by Leighton (14) and by Howard (27). The latter, using a 10″ scanning aperture, found that calcium plages were outlined very nearly by a 10-gauss contour line.

Sunspots.—The subject of the magnetic fields of sunspots is a large one, really beyond the scope of this review, so that only a few recent results will be mentioned here. Nevertheless, sunspot fields can probably be regarded as actually parts of the general magnetic field of the sun. Spots, even when single and isolated as observed in white light, are invariably shown by the magnetograph to occur within bipolar or multipolar magnetic regions, where they represent local concentrations of the field. According to the theory developed by Biermann (28), the nearly vertical lines of force of the sunspot field inhibit convection, and in this way reduce the outflow of energy so that the spot is dark compared to the surrounding photosphere. The bipolar nature of sunspot groups, with conservation of positive and negative flux, was the basis of Cowling's (29) suggestion that such groups are formed by the emergence of flux loops from a submerged toroidal belt encirling the sun, it being presumed that there are two belts of opposite polarity, one on each side of the sun's equator.

Kiepenheuer (30) has provided a detailed summary of sunspots and their magnetic fields as of 1953. According to Treanor's (3) more recent investigations, already mentioned, the field-strength variation within a spot is in agreement with Broxon's parabolic law, and 40 per cent of the flux is found to be confined within a cone of unit solid angle. Comparisons of field strength at λ 6173 (Fe I) and λ 6149 (Fe II) suggested that the gradient in depth does not exceed 1 gauss/km. From analyses of line profiles, Baranovsky & Stepanov (31) derived gradients of 1.0 to 1.8 gauss/km.

The subject of the magnetic field in sunspot umbrae has been investigated by Bumba (32). Attention was centered on the unexplained strength of the π component of the Zeeman pattern in the spectra of the umbrae. As was noted by Hale and co-workers (33), "In many spots the π component is strong even at the center of the umbra when the spot is near the center of the sun." Bumba attempts to explain this effect by the form of the lines of force of the magnetic field within the umbra. According to him, peculiarities of the Zeeman patterns within the umbra can be accounted for on the supposition that the lines of force are twisted into helices, the pitch increasing with height. These lines of force form a nearly horizontal ring in the lower layers of the umbra, while they become approximately radial in the higher levels. As we shall see later, this is in conformity with the idea that twisting of the submerged magnetic flux ropes comes about as a result of differential rotation.

Flares.—The modern study of solar flares (or bright chromospheric eruptions as they were once called) was initiated by Hale (34). Having provided a number of spectrohelioscopes to some 20 different observatories, he organized a world-wide surveillance of the sun, one aim of which was to accumulate new data on flares and their relationship to geomagnetic storms. It was recognized that flares almost invariably occur in close relationship to sunspots. In particular, it was found that flares occur more often near complex groups (i.e., in multipolar magnetic regions), and when the spots are changing rapidly. Kiepenheuer (30) gave a review of the subject up to 1953, and showed that the magnetic energy of a spot field is sufficient to account for all the emissions of a flare. Giovanelli (35) suggested that flares result from electric discharges occurring at neutral points of the magnetic field. Further work on the neutral-point theory of flares has been done by Dungey (36, 37) and by Sweet (38), and has been criticized by Cowling (29). Severny has advanced a magnetic shock-wave theory of flares (39).

Much of the recent work on flares has been directed to the observation, with magnetographs, of the fields associated with them. This is particularly difficult because of the high resolution on the disk that is required for the adequate portrayal of the details of the complex magnetic patterns in the presence of irregularities due to guiding and solar seeing. Also, because of the abruptness of the phenomena, it is seldom that a sequence of fine-scan magnetograms can be started early enough in the development of a flare. Severny (40) has obtained magnetic records of a considerable number of flares; generally the area mapped measured some 2′ by 3′ or larger. Severny believed that flares tend to occur in neutral points of magnetic fields when the field gradient is sufficiently large, and that the appearance of the flares destroys the surrounding magnetic field. In this connection it has not always been entirely clear whether a "neutral point" is a point at which the field $H=0$, or whether the term is used to mean a point at which the sight-line component of the field, as determined by the magnetograph, is zero. According to Bruzek's (41) analysis of the records of several flares, the great flares and most of the minor ones originated at the edges of umbrae and in regions of strong magnetic fields, but never in "neutral points" of zero field strength.

Evans (42) reported measurements of changes in the magnetic field associated with a flare of intensity 1+. According to him, the field of the sunspots and of the area outside the spots showed a sharp decrease of about 16 per cent during the rising phase of flare intensity, and an even sharper recovery immediately after the flare maximum.

Michard, Mouradian & Semel (43) observed magnetic fields related to flares. They were able to measure longitudinal fields stronger than about 30–40 gauss, and to make qualitative estimates about transverse fields. They concluded that if important field variations are associated with flares, they must occur before the beginning of the Hα emissions.

Howard & Babcock (10), observing the large flare of July 16, 1959, with a sequence of fine-scan magnetograms, found no reliable evidence for changes in the magnetic pattern of the photospheric fields associated with the flare, in the range 5–40 gauss.

Bappu & Punetha (44), noting the almost one-to-one correspondence between the details of the calcium flocculi and the details of the longitudinal magnetic field, have examined calcium spectroheliograms, taken during the occurrence of solar flares, from the 60-year Kodaikanal collection. Analyzing records before, during, and after four large flares, they concluded that no changes of the longitudinal component of the field in the flare region greater than 20 gauss take place during the occurrence of major flares.

Evidently the measurement of the magnetic fields related to solar flares, and their changes, is a matter of considerable technical difficulty that has not as yet been adequately solved. Increased resolution and precision of observation are required, over a wide range of field intensity, ranging from the spot umbrae to the outlying fields near the borders of the active region. The realization by observers of 30 years ago that flares were favored by regions in which the spots were changing rapidly is significant, as is the fact that flares often represent a brightening of pre-existing features with little change in position or form. It is not the attribution of the source of flare energy to the magnetic field which is debatable, but rather the details of the observations and the theory to explain them. Current opinion, as indicated by the recent work of Howard & Severny (45), favors the idea that the magnetic fields significant in flare development are those of the associated spots, and probably not to any marked degree those of the surroundings.

Stable filaments.—The disposition of stable filamentary prominences in relationship to the magnetic fields of bipolar magnetic regions was studied in a preliminary way in 1955. It was found that such stable filaments, occurring in older, inactive bipolar regions, frequently lay at right angles to the lines of force, dividing, at an elevated level, the positive and negative parts of the regions. Presumably the partly ionized material of the prominence was supported in a trough resulting from a sag in the lines of force near the top of the magnetic arches above the bipolar region. The theory of the support of filaments has been treated by Kippenhahn & Schlüter (46). A few stable filaments appeared to delineate the boundaries of magnetic regions, especially on the poleward side. Active prominences, on the other hand, occurring usually near sunspots, almost certainly are aligned with the magnetic field, although detailed observational studies are few.

Zirin & Severny (47) reported the measurement of fields of the order of 200 gauss in five active prominences and possible weak fields (25–50 gauss) in two quiescent prominences. Later, Yoshpa (11) measured magnetic fields in active prominences with the solar magnetograph of IZMIRAN. The Hβ line was used. These fields reached several hundred gauss. It was stated that the field in quiet prominences is probably less than the threshold of the instrument (50 gauss).

Granules.—Reliable measurements of the magnetic field of individual solar granules have not yet been achieved and confirmed. A granule is so small that seeing and guiding problems make it extremely difficult to hold such features on the entrance aperture for the interval required to accumulate the desired information. From the fact that large areas of the quiet sun (apart from the weakest magnetic regions) usually show a mean field not greater than 0.2 gauss (when a long slit is used), Babcock & Babcock (18) concluded that, if granules are assumed to have coherent fields, randomly oriented, the order of magnitiude of the individual fields is no greater than about 2 gauss. Such observations can be refined.

Steshenko (48) attempted to measure the fields of granules by photographing their spectra through circular analyzers. Imperfect seeing limited the capability of the method, and it could only be concluded that an upper limit to the field of separate granules was 40–60 gauss.

Semel (49), using spectrograms obtained by Michard at Sacramento Peak, claimed to have determined a relatively precise value of 24 gauss for the field of granules. There was no indication, however, that the measures were known to pertain to granules of the truly quiet photosphere, outside of weak bipolar magnetic regions. As Semel pointed out, the fine structure of weak local fields could vitiate the results, and the proof that such a weak field did not exist in the area measured would require a magnetograph of high sensitivity.

The detailed disposition of the field lines of solar magnetic regions, with respect to the granules and to the larger, quasi-stable convective features of the solar atmosphere, requires clarification. The present view is that the field lines assume a concentrated, columnar structure, nearly perpendicular to the surface, between the convective cells in which there is sustained horizontal outflow of plasma in all directions from a rising center. These convective cells, constituting a "supergranulation" pattern, have been described by Leighton, Noyes & Simon (50).

The Solar Magnetic Cycle

The large magnetohydrodynamical problem of the sun is the solar cycle itself. Obviously a great many phenomena are related in this cycle, and all are interlinked by the magnetic field. The 11-year variation in the number of sunspots is the most obvious aspect; next is the variation in the mean latitude of sunspots as embodied in Spörer's law and portrayed in the familiar "butterfly diagram" of Maunder (51). These and other related phenomena must be accounted for in any comprehensive theory, all within the framework of magnetohydrodynamics. Some of these related phenomena are: (*a*) the existence and (presumed cyclic) reversal of the poloidal field of the sun; (*b*) the sun's differential rotation (equatorial acceleration); (*c*) Hale's laws of sunspot polarity (which apply generally to bipolar magnetic regions and which show that, because of the alternating polarity, the magnetic cycle covers some 22 years; (*d*) the fact that surface magnetic regions disappear by expansion; (*e*) the dominance of "preceding" spots over "following" spots; (*f*) the cyclic variation in form and intensity of the coronal streamers; (*g*) the chromospheric "whirls" around many sunspots, which are independent of magnetic polarity.

Several recent attempts have been made to develop a theory or model of the solar magnetic cycle. These efforts generally owe much to the fundamental discussions and earlier theoretical work of Cowling (29). None is complete, but all have something in common. These proposals have been made independently by Parker (52), Waldmeier (53), Allen (54), Steenbeck (55), Alfvén (56), and Babcock (57). It is impossible, within the scope of this review, to intercompare all of these theories, or even to describe any one of them in detail. Perhaps the writer will be excused for presenting some limited aspects of the one with which he is most familiar.

The changes in the magnetic pattern at the surface of the sun could not occur as rapidly as they do unless they are largely dependent on motions or circulation of the conducting plasma, carrying with it the embedded magnetic-field lines. This is substantiated by the differential rotation and by the evidence for meridional currents. The latter comes from studies of the latitude drift of individual spots by Tuominen (58) and by Richardson & Schwarzschild (59). At the outset, we assume that the fields essential to the solar cycle are fairly shallow. The lines of force of the initial poloidal field lie nearly in meridian planes; their submerged parts, lying at a depth of the order of 0.1 R, join the north and south polar caps; externally, they loop out to great distances, with considerable irregularity. The differential rotation gradually draws out the submerged lines of force toward the west, beginning at the equator, until they have circled the sun several times. The close spiral wrapping thus formed constitutes a greatly amplified field that is nearly toroidal. At a given latitude, the amplification factor is sec ψ, where ψ is the angle between the meridian and the line of force. The relative strength of the spirally wrapped field can be computed as a function of heliographic latitude, ϕ, because the law of differential rotation is known. From the work of Newton & Nunn (60) on sunspot motions, it can be shown that the differential lag in longitude, with respect to the equator, is, in radians,

$$\Delta\theta = 17.6\,(n + 3)\sin^2\phi \qquad (\phi < 35°)$$

where n is the elapsed time in years measured from the beginning of the new sunspot cycle. It is implicitly assumed here that initially ($n = -3$) the field is purely poloidal, and that after three years of amplification the submerged toroidal field has reached, at $\phi = \pm 30°$, a critical intensity such that instabilities begin to occur, thus initiating a new sunspot cycle. At this time ($n = 0$) the equator will have gained in its rotation about 2.1 turns on the latitude circles at $\phi = \pm 30°$, and the field at this latitude will have been amplified by a factor of about 46. The instabilities in the submerged magnetic flux ropes near the 30° latitude circles will bring flux loops to the surface, with the formation of bipolar magnetic regions. This results in multiple fragmentation of the flux ropes, thus terminating amplification at this latitude. Amplification continues at lower latitudes, however, so that the field strength of the submerged flux ropes increases, reaching the critical value at progressively lower latitudes. Thus, migration in latitude of the flux ropes themselves is not required. Some ten years after the onset of activity, the latitude at which the critical value of the field is being reached will have decreased, at a diminishing rate, to about 7°. It has been shown that the simple equation

$$\sin\phi_c = \pm\, 1.5/(n + 3)$$

relates the age of the cycle in years n to the latitude ϕ_c at which the field is becoming critical. This equation, which is plotted in Figure 4, represents a derivation of Spörer's law. As turbulence of the plasma superposes considerable irregularity on the smooth relationship expressed by the foregoing equation, the width of the zones of solar activity is readily accounted for.

By tracing the polarity of the field lines, it is seen that the polarity of the spot groups produced by this model is consistent with the polarity of the initial poloidal field. Consideration of the details of the mechanism further shows that plausible arguments can be adduced for the preponderance of leading as compared to following spots, for the recurrence of activity in certain zones of longitude, and for the chromospheric whirls around spots. The total flux (8×10^{21} maxwells) of the initial poloidal field, upon amplification and conversion to a strong, spirally wrapped field, can readily provide sufficient magnetic flux for some thousands of bipolar magnetic regions, in con-

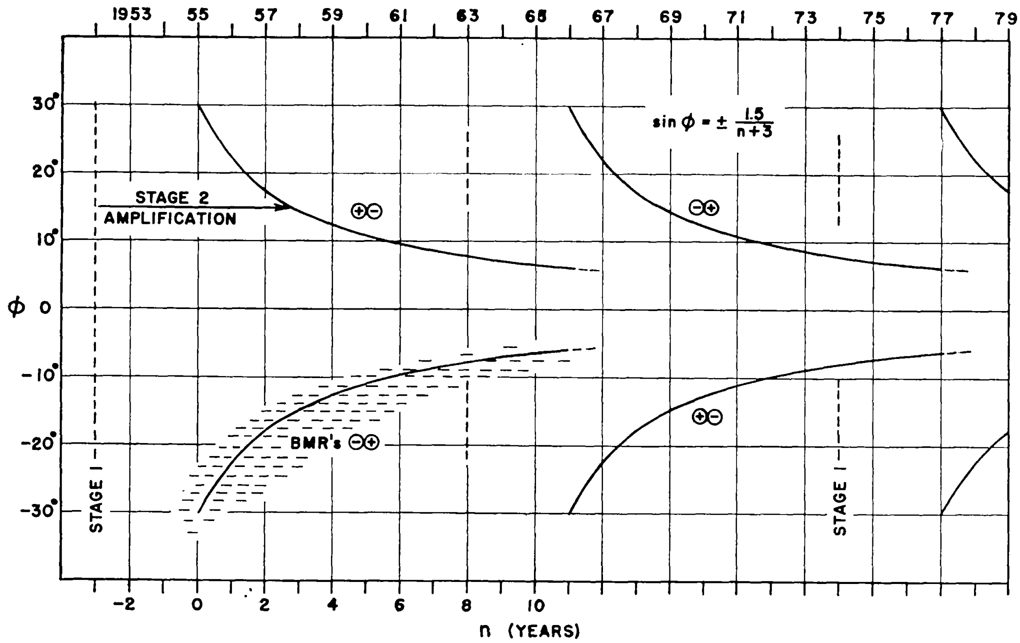

FIG. 4. The curved lines converging toward the equator represent, as a function of time, the latitude at which the submerged magnetic field is reaching the critical value required for instability and the outbreak of bipolar magnetic regions or sunspot groups.

formity with the number observed, and with the figure of 10^{21} maxwells for the order of magnitude of the flux of the average bipolar regions.

These bipolar magnetic regions, in their waning stages, generally expand until they are ill defined, or until they lose identity by merging with other such regions. The greater part of the merging is to east or west, but there is also a tendency for the preceding parts of the regions to migrate toward the equator, where they neutralize their counterparts from the opposite hemisphere; the following parts, meanwhile, migrate toward the poles, where they first neutralize the initial polar cap fields and later replace them with a poloidal field of reversed polarity. Only a small fraction of the available magnetic flux is required for this supplantation of the poloidal field of the first stage.

The arching lines of force in the corona that join the two parts of each bipolar magnetic region expand outward to great distances, and, according to this model, eventually break away, to be carried outward with the general flow of plasma away from the sun. This is compatible with the optical observation of coronal features. These far-reaching loops have sometimes been termed "magnetic bottles." Energetic particles, accelerated by sporadic occurrences such as flares, are constrained by coupling to such field lines. The balance of kinetic and magnetic forces is important in the dynamics of the interplanetary medium.

In the merging of bipolar magnetic regions with other such regions to the east and west, one visualizes the lines of force that formed the toroidal component of the submerged solar field as deviously slipping out through the photospheric layers, and (except for the small fraction that go to form the new reversed poloidal field) being lost to the sun. This release of magnetic flux by transverse leakage through the photosphere is facilitated by the reduced conductivity of this layer due to turbulence, a subject discussed by Sweet (61) and by Elsasser (62). The magnetic energy thus released must be largely reconverted to thermal energy of the high corona.

It has been estimated that the energy going into the formation of the toroidal component of the field in each 11-year sunspot cycle is 10^{36} ergs. The immediate source of this energy is the kinetic energy of differential rotation, which must be maintained in some way if the magnetic cycle is to persist for more than a few thousand years at its present level. Mestel (63) has discussed the theory of equatorial acceleration in a rotating magnetic star in a steady state, pointing out that the existence of the solar cycle requires a periodic rather than a steady solution.

LITERATURE CITED

1. Hale, G. E., *Astrophys. J.*, **28,** 315 (1908)
2. Hale, G. E., and Nicholson, S. B., *Magnetic Observations of Sunspots* (*Carnegie Inst. of Washington Publ. No. 498*, Parts I and II, Washington, D. C., 1938)
3. Treanor, P. J., *Mon. Notices Roy. Astron. Soc.*, **120,** 412 (1960)
4. Babcock, H. W., and Cowling, T. G., *Mon. Notices Roy. Astron. Soc.*, **113,** 357 (1953)
5. Babcock, H. W., *Astrophys. J.*, **118,** 387 (1953)
6. Kiepenheuer, K. O., *Astrophys. J.*, **117,** 447 (1953)
7. Seares, F. H., *Observatory*, **43,** 310 (1920)
8. Billings, B. H., *J. Opt. Soc. Am.*, **39,** 797 (1949)
9. Bowen, I. S., *Astrophys. J.*, **88,** 113 (1938)
10. Howard, R., and Babcock, H. W., *Astrophys. J.*, **132,** 218 (1960)
11. Yoshpa, D. A., *Geomagnetism and Aeronomy*, **2,** No. 1, 172 (1962)
12. Deubner, F. L., Kiepenheuer, K. O., and Liedler, R., *Z. Astrophys.*, **52,** 118 (1961)
13. Nikulin, N. S., Severny, A. B., and Stepanov, V. Ye, *Isv. Crimean Astrophys. Obs.*, **19,** 3 (1958)
14. Leighton, R. B., *Astrophys. J.*, **130,** 366 (1959)
15. Dollfus, A., *Compt. Rend.*, **246,** 2345; **246,** 3590 (1950)
16. Leroy, J. L., *Contributions a l'étude de la polarisation de la lumière solarie* (Doctoral thesis, Univ. of Paris, 1962)
17. Lorentz, H. A., quoted by Wood, R. W.. *Physical Optics*, 521 (Macmillan, New York, 1911)
18. Babcock, H. W., and Babcock, H. D., *Astrophys. J.*, **121,** 349 (1955)
19. Babcock, H. W., and Babcock, H. D., *Intern. Astron. Union, Symp., 6th, Stockholm, August 1956*, 161 (1958)
20. Babcock, H. D., *Astrophys. J.*, **130,** 364 (1959)
21. Waldmeier, M., *Z. Astrophys.*, **49,** 176 (1960)
22. Cowling, T. G., *Mon. Notices Roy. Astron. Soc.*, **105,** 166 (1945)
23. Klüber, H. von, and Beggs, D. W., *Quart. J. Roy. Astron. Soc.*, **1,** 92 (1960); **3,** 116 (1962)
24. Högbom, J. A., *Mon. Notices Roy. Astron. Soc.*, **120,** 530 (1960)
25. Simpson, J. A., Babcock, H. W., and Babcock, H. D., *Phys. Rev.*, **98,** 1402 (1955)
26. Stepanov, V. E., and Petrova, N., *Isv. Crimean Astrophys. Obs.*, **21,** 152 (1959)
27. Howard, R.. *Astrophys. J.*, **130,** 193 (1959)
28. Biermann, L., *Vierteljahrsschr. Astr. Ges.*, **76,** 194 (1941)
29. Cowling, T. G., *The Sun*, Chap. 8 (Kuiper, G. P., Ed., Univ. of Chicago Press, Chicago, 1953)
30. Kiepenheuer, K. O., *The Sun*, Chap. 6

(Univ. of Chicago Press, Chicago, 1953)
31. Baranovsky, E. A., and Stepanov, V. Ye., *Isv. Crimean Astrophys. Obs.*, **21,** 180 (1959)
32. Bumba, V., *Bull. Astron. Inst. Czechoslovakia*, **13,** 42; **13,** 48 (1962)
33. Hale, G. E., Ellerman, F., Nicholson, S. B., and Joy, A. H., *Astrophys. J.*, **49,** 153 (1919)
34. Hale, G. E., *Astrophys. J.*, **73,** 379 (1931)
35. Giovanelli, R. G., *Mon. Notices Roy. Astron. Soc.*, **107,** 338 (1947)
36. Dungey, J. W., *Phil. Mag.*, **44,** 725 (1953)
37. Dungey, J. W., *Phil. Mag.*, **44,** 725 (1953)
37. Dungey, J. W., *Intern. Astron. Union, Symp., 6th, Stockholm, August 1956*, 153 (1958)
38. Sweet, P. A., *Intern. Astron. Union, Symp., 6th, Stockholm, August 1956*, 123 (1958)
39. Severny, A. B., *Isv. Crimean Astrophys. Obs.*, **20,** 22 (1958)
40. Severny, A. B., *Isv. Crimean Astrophys. Obs.*, **22,** 12 (1960)
41. Bruzek, A., *Z. Astrophys.*, **50,** 110 (1960)
42. Evans, J. H., *Astron. J.*, **64,** 330 (1959)
43. Michard, R., Mouradian, Z., and Semel, M., *Ann. Astrophys.*, **24,** 54 (1961)
44. Bappu, M. K. V., and Punetha, L. M., *Observatory*, **82,** 170 (1962)
45. Howard, R., and Severny, A. B., *Astrophys. J.* (In press, 1963)
46. Kippenhahn, R., and Schlüter, A., *Z. Astrophys.*, **43,** 36 (1957)
47. Zirin, H., and Severny, A. B., *Observatory*, **81,** 155 (1961)
48. Steshenko, N. V., *Isv. Crimean Astrophys. Obs.*, **22,** 49 (1959)
49. Semel, M., *Compt. Rend.*, **254,** 3978 (1962)
50. Leighton, R. B., Noyes, R. W., and Simon, G. W., *Astrophys. J.*, **135,** 474 (1962)
51. Maunder, E. W. *Mon. Notices Roy. Astron. Soc.*, **82,** 534 (1922)
52. Parker, E., *Astrophys. J.*, **121,** 491 (1955)
53. Waldmeier, M., *Z. Astrophys.*, **49,** 176 (1960)
54. Allen, C. W., *Observatory*, **80,** 94 (1960)
55. Steenbeck, M., *Beitr. Plasmaphysik*, **3,** 153 (1961)
56. Alfvén, H., *Astrophys. J.*, **133,** 1049 (1961)
57. Babcock, H. W., *Astrophys. J.*, **133,** 572 (1961)
58. Tuominen, J., *Z. Astrophys.*, **21,** 96 (1941); **30,** 261 (1952); **37,** 45 (1955); **51, 91** (1961)
59. Richardson, R. S., and Schwarzschild, M., *Accad. Nazl. Lincei, Atti Convegno Volta*, **11,** 228 (1953)
60. Newton, H. W., and Nunn, M. L., *Mon. Notices Roy. Astron. Soc.*, **111,** 413 (1951)
61. Sweet, P. A., *Mon. Notices Roy. Astron. Soc.*, **110,** 69 (1950)
62. Elsasser, W. M., *Rev. Mod. Phys.*, **28,** 135 (1956)
63. Mestel, L., *Mon. Notices Roy. Astron. Soc.*, **122,** 473 (1961)