PART II

SOLAR ELECTRODYNAMICS

https://doi.org/10.1017/S0074180900237686 Published online by Cambridge University Press

PAPER 12

SOLAR ELECTRODYNAMICS

T. G. COWLING

The University, Leeds, England

ABSTRACT

A historical account of the subject's development is attempted. Prior to 1940, the most significant papers were those by Larmor (1919) and Cowling (1934) on dynamo theories of solar fields: by Kiepenheuer (1935) on the corona; and by Ferraro (1937) on isorotation. These indicated the importance of electromagnetic forces and were groping towards the idea of frozen-in fields. The latter idea was, however, not clearly stated before Alfvén's 1941–2 papers.

Theory since then is divided into sections concerned with mechanical effects of magnetic fields, theories of sunspots, and the nature and origin of solar magnetic fields. The first includes theories of magnetic control of support of coronal filaments and prominences (van de Hulst, Alfvén, Dungey) and theories of magnetic influence on sunspot equilibrium. The second includes Alfvén's and Walén's theories of the solar cycle, and Biermann's explanation of sunspot coolness in terms of magnetic inhibition of convection. Sunspot theories, being discussed more fully by Biermann, are considered only briefly.

Electromagnetic heating covers theories of coronal heating and flares, discharge phenomena, particle acceleration and radio emission. Many of the older theories (Alfvén's on coronal heating, Giovanelli's on flares, that of Bagge and Biermann on cosmic rays) are set aside because of their neglect of self-induction effects and inadequacy of the mechanism of conversion. The relative motion of charged particles and neutral atoms (Piddington, Cowling) is described as supplying a powerful heating effect.

As regards the magnitude of the general solar magnetic field, it is suggested that the observed value can be discarded only if decisive reasons are given. Other theories having so far proved inadequate, dynamo theories of the origin of solar fields are regarded as the most promising. These can be partial, as when a toroidal field capable of explaining spot fields is supposed to be generated from the general field (Walén and others), or when a turbulent field is supposed to be generated from a smaller regular field (Alfvén and others): or total, when a simultaneous explanation of all fields is attempted (e.g. Parker). A general appraisal is made of the different theories.

In what follows, a historical account of the development of solar electrodynamics will be attempted. Emphasis will be rather on theories advanced to explain the different phenomena than on the observations; the phenomena are well known, their explanation is difficult.

Solar electrodynamics may be said to have begun in 1889, when Bigelow inferred a solar magnetic field from the form of coronal streamers near the poles. However, real interest in the subject began only after Hale's measurements of sunspot magnetic fields and his announcement of the existence of a general solar magnetic field. His discoveries stimulated Larmor, in 1919, to suggest a dynamo theory of the origin of such fields. Even though the suggestion was shown by Cowling in 1934 to be untenable in the form in which he had advanced it, the general idea of dynamo maintenance of cosmic fields is one that is still fruitful.

The Mount Wilson workers originally suggested that the sun's general field, and probably sunspot fields, are limited to low layers in the sun's atmosphere. In 1928 Chapman advanced a theory of the radial limitation of the general field, based on a study of the drifts of charged particles in crossed electric, magnetic and gravitational fields; a second theory, based on the diamagnetic properties of free electrons, was advanced by Ross Gunn. Cowling showed in 1929 that such theories were untenable, and that any correct theory must be based rather on a study of the electric currents flowing. Later the phenomenon of rapid radial limitation was recognized as non-existent. Indeed, the very existence of a general solar field was for a time regarded as extremely doubtful; only in the last few years has the work of H. W. and H. D. Babcock established the existence of such a field, though they have found that its properties are far different from those earlier ascribed to it. Before 1945, most theoretical work was based on the assumption of a dipole-like field, with maximum strength about 50 gauss near the poles.

During the decade 1930–40 some progress was made in understanding the basic laws of motion of an ionized gas in a magnetic field. The developments were parallel to those made in theories of the ionosphere, put forward about 1930. The importance of mechanical interaction between material and field in a sunspot was clearly recognized by Cowling in his 1934 paper, and there were glimmerings of the idea that lines of force are frozen into the material. Cowling did not, however, fully realize the extent to which lines of force can be frozen into the material, because of a mistaken belief that a polarization electric field permits material to slip freely through a magnetic field.

Kiepenheuer, in a paper on the corona in 1935, made use of the idea of frozen-in fields; he supposed that masses of gas ejected into the corona break off, and carry with them, pieces of the photospheric field. The lines of force are probably too firmly frozen into the material to permit pieces of the field to be broken off like this, but even this partial recognition of frozen-in fields is interesting. The same principle provides the simplest explanation of Ferraro's law of isorotation (1937) that in a star in steady non-uniform rotation the angular velocity must be the same at all points of a line of force; but this explanation was not, in fact, given.

Just before the last war, moving pictures of prominence motions were first seen. These offered clear evidence of the influence of magnetic fields. Jets of material were seen to be continually thrown up along curved paths, returning along the same paths. Since charged particles are known to follow lines of force, the inference was obvious that these curved paths are along lines of force. It was also speculated, both then and later, that quiescent prominences might in part be supported by magnetic forces.

The subject developed more rapidly after 1941, stimulated by Alfvén's work. Alfvén clearly stated the principle of frozen-in fields; he showed that in consequence magneto-hydrodynamic waves could be propagated along the lines of force in a conducting fluid, and re-derived the law of isorotation; and he indicated again the importance of magnetic forces to sunspot equilibrium. The importance of his work was in his emphasis on the two-way interaction between magnetic fields and material motions. True, his results were all implicit in the theory of Maxwell's stresses; others had earlier attained some of his ideas; and sometimes magneto-hydrodynamic waves were invoked when clearly the actual interaction between field and motion was far more complicated than a wave motion. But these facts should not obscure the importance of the impact of Alfvén's ideas.

From this point on, the ramifications of the subject become too great for it to be treated as a whole. Developments in the theory of sunspots will first be considered. Alfvén advanced a theory of sunspots, based on the assumed production of whirl rings in the sun's interior, and their progression along lines of force of the general field to the sun's surface. This theory was developed and made more precise by Walén. Alfvén never wholly accepted Walén's account, and Walén has since repudiated it; but it still remains the most complete account of Alfvén's theory. It is unfortunate that Alfvén has never given more details, in particular of the way in which the surface magnetic field is created from a whirl ring; at present his theory is a theory of the solar cycle without any proof that it can explain the existence of individual spots. However, it would not be appropriate for me here to add to the criticisms of Alfvén's work which I have made elsewhere in more than one place.

Walén, after abandoning Alfvén's ideas, has himself put forward a theory of the solar cycle. This is based on the idea of torsional oscillations round the sun's axis, due to the stiffness introduced by the general magnetic field. The oscillations are excited by periodic convulsions in the sun's interior. Once again, since I have criticized Walén's work elsewhere I shall not go into details here.

In 1939, Biermann suggested to me in a letter that sunspots might be due to the inhibition of convection by a magnetic field, and the consequent reduction of heat transported to the surface. This suggestion appeared, almost as a chance remark, in a paper by him in 1941. The inhibition of small-scale convection by a magnetic field was inferred, on general physical arguments, by Walén in 1949; detailed mathematical work by Thompson and Chandrasekhar, while confirming that a magnetic field interferes with convection, showed that it does not make it impossible with a sunspot. I have tried myself to develop the theory that sunspot darkening is due to the reduction of convection in a spot; Hoyle has suggested a theory on rather different lines, convection being not so much reduced as restricted to motion along the lines of force.

Theories of electromagnetic heating will next be considered. In 1940, Alfvén suggested that some prominences might be the visible signs of electromagnetic discharges. Such a belief was possible when a prominence could be regarded as shining because hotter than the corona, but it is untenable now that the coronal temperature is known to be 10^6 °C. All that survives of it is the suggestion, which Dungey has elaborated, that electric currents flowing along prominence arches may lead to increased densities through a pinch effect. Even this is subject to strong objections, both because of the difficulty in making currents flow along such an arch, and because of the instability of such currents. Prominence arches appear rather to consist of material moving along lines of force; however, a real difficulty, not so far discussed, is to explain how coronal material condenses into prominences in spite of the resistance provided by a magnetic field.

In 1948, Giovanelli suggested that solar flares are due to electromagnetic discharges along magnetic lines of force. He supposed the discharges to arise because the electrical conductivity of an ionized gas increases rapidly with the temperature, so that the beginnings of a discharge produce a highly conducting channel along which the further discharge can readily proceed. This idea is attractive, but Piddington and Cowling have shown that it encounters overwhelming difficulties. The electric field available to drive the discharge was only guessed, and in view of the extent to which lines of force are frozen into the material it is doubtful whether fields of the magnitude required can exist. Moreover—and this is the fundamental objection—self-induction ensures that an increase in conductivity along a channel of the size of a flare does not lead to any appreciable increase in the current flowing, within a time of the order of a day. Thus increases in conductivity cannot lead to a sudden increase in radiation within a few minutes, as actually observed; a decreased resistance should rather be expected to lead to decreased heating.

In 1947, Alfvén suggested that the high temperature of the corona might be due to Joule heating by magneto-hydrodynamic waves. His suggestion was based on a formula due to Cowling, which indicated a reduction in conductivity transverse to the lines of force, this leading to enhanced heating. Schlüter later showed that in a fully ionized gas the reduction in conductivity is of such a nature as to produce no increase in the heating effect of electric currents; this is reasonable, since the magnetic field does not increase the number of collisions between ions and electrons, from which Joule heating arises. This disposes of the suggestion that Joule heating is important in the corona itself. However, Piddington and Cowling have recently shown that in a partially ionized gas the reduction in conductivity does correspond to a real increase in the Joule heating. Thus the production of fast coronal particles may be possible in the upper layers of the chromosphere, where a small number of neutral particles are still present.

Piddington suggests that a similar mechanism may be responsible for solar flares. Magneto-hydrodynamic waves can be supposed to be generated in sub-surface layers of a sunspot by convection which, though held in check by a magnetic field, is none the less present there on a reduced scale. Such waves in certain circumstances travel upward with relatively little loss in energy, provided that they are associated with mainly horizontal motions; when they reach a sufficient heat, the conversion of energy into Joule heat becomes important, and a flare is observed. On this theory, a flare is simply an enhanced form of an activity present all the time, presumably that responsible for plages. A different theory of flares has been put forward by Dungey, who suggested that an instability near a neutral point of the magnetic field might lead to a progressive increase in the electric currents flowing near such a point. This provides a discharge theory which is an advance on Giovanelli's but which nevertheless still appears a little artificial.

Since 1945, a large flare has on a number of occasions been observed to be accompanied by the emission of numerous soft cosmic-ray particles from the sun. Their mechanism of acceleration is almost certainly electromagnetic. Bagge and Biermann suggested that they might arise near a magnetic neutral point, accelerated by an electric field due to the relative motion of two magnetic fields. Such a suggestion appears untenable, in view of the closeness which with lines of force are frozen into the material; the necessary electric fields can hardly exist. One possibility appears to be that particles may on occasion travel with the phase-velocity of magnetohydrodynamic waves, speeding up with these waves as they travel into regions of less density. This is a more ambitious version of an idea advanced by the Babcocks, according to which constrictions in a bundle of lines of force may accelerate particles from lower levels into the corona. The mechanism has certain affinities with a Fermi mechanism of acceleration, and also with the mechanism suggested by Menzel and Salisbury, according to which particles are accelerated by riding on the crest of low-frequency electromagnetic waves.

Solar radio emission is in a different category from other phenomena considered above since, apart from certain outbursts, it seems to have little connexion with magnetic fields. The origin of the outbursts is generally believed to be some form of plasma oscillations. Any theoretical discussion of such oscillations is normally based on assumptions which preclude the possibility of escape of the radiation generated, but there is no obvious reason why such assumptions are essential.

Finally, one may comment on theories of the nature and origin of solar magnetic fields. Alfvén's theory of sunspots posited a general solar field of order 25 gauss, whereas the latest observations indicate a field of order 1 gauss. Alfvén has suggested that a field of order 25 gauss may in fact exist, overlaid by a turbulent field several times greater, and that the lower observed figure is due to lower observability of the turbulent + general fields when they reinforce each other than when they are opposed. A theorist has a reluctance, sometimes misplaced, to trying to explain away the observations; and in any case, whereas Alfvén's ideas may be reasonable, one cannot be satisfied with less than an argument which shows that the observed field must inevitably be smaller than the real one.

Attempts to explain solar fields, either the general field or that of sunspots, have so far met with limited success. Explanations in terms of thermo-electric effects have proved inadequate. Biermann has indeed shown that in the presence of non-uniform rotation such effects may produce toroidal fields some hundreds of gauss strong, but since such fields cannot be reversed from one sunspot cycle to the next, they can hardly explain the observed phenomena. The time of decay calculated by

https://doi.org/10.1017/S0074180900237686 Published online by Cambridge University Press

Cowling for the general field leaves open the possibility that this may be a relic of an interstellar field existing before the sun's formation, but only if turbulence does not materially reduce the time of decay. Apart from this, the most promising possibility of explaining the fields seems to be in terms of a dynamo theory.

Some partial theories attempt to explain one of the solar fields in terms of another. For example, Walén and others have attempted, by invoking torsional oscillations or other periodic changes in the angular velocity, to derive from the general field a toroidal field capable of explaining spot fields. This suggestion involves difficulties about orders of magnitude, unless one supposes the general field to be stronger below than at the surface. Some workers have preferred to reverse the argument, seeing in the sun's general field the survival of fields of previous sunspot cycles. Again, turbulence in the sub-surface layers is sometimes supposed to twist the lines of force of the general field to give a stronger but irregular turbulent field. On the other hand, Walén has suggested that turbulence may actually operate to prevent a magnetic field from penetrating the turbulent layer, and arguments by Sweet and Elsasser indicate that a field which does penetrate the turbulent layer will, at least, decay rapidly.

A complete dynamo theory must explain all solar fields, general, spot and the rest, by induction due to a motion which, though not necessarily completely regular, should at least possess certain regular characteristics. The most ambitious theory of this kind to date is that of Parker, who invokes as a regular characteristic the effect of Coriolis forces in twisting rising and falling convection currents. He succeeds in deducing a field largely confined to the sun's outer layers, which is largely poloidal at high latitudes, toroidal at low. The surface field steadily travels down towards the equator, and has a zonal structure, a belt with a toroidal field of one sign being followed at higher latitudes by one with a field of opposite sign. The theory is difficult to express in precise mathematical form, but some deeper investigation of it is really required.

This concludes my historical account. I have unfortunately had to omit any account of Russian work, and hope that others will repair the omission. I have tried to avoid undue bias; at the same time, one cannot present an account like this without voicing one's own personal opinions. I do not expect anyone to agree with everything that I have said.

Our subject is a young one, even though its beginnings were earlier than is sometimes realized. Workers too often think of the subject as having begun with their own work; indeed, because it is so young and because connected accounts of it are not numerous, the rediscovery of earlier results has not been rare. The law of isorotation, first found by Ferraro, was certainly found independently by Alfvén and Sweet. The fact that inequalities of angular velocity might produce from the general field a toroidal field of importance in the theory of sunspots was certainly realized by me before 1945; I have since seen it appear under the names of at least four authors. Certain basic mathematical equations have similarly been derived independently by more than one author, and interpreted differently by them. Such repetitions are inevitable in so young a subject as our own; so to speak, ideas are in the air, and different workers pluck them out of the air at different times and in diverse manners. For this reason, though I have tried to attribute priority where due, questions of priority do not seem to me to be those of prime importance. The main triumphs of our subject are still in the future; the past is relevant only as it guides our future work.

Discussion

Gold: A new situation arises in the study of flares and their correlation with high-energy particles. We see these particles coming from flares but we do not know how. The mechanism is not understood. One conjecture, and to my mind a very useful one, is that the flare represents the sudden instability of a volume current in its own magnetic field (pinch effect), as suggested earlier by Alfvén.

If this is right it may be of interest to refer to what may be the most interesting laboratory analogy. Dr Kurchatov reported in April about the Russian experiments on high intensity discharges, and one result was the production of particles with energies in excess of the energies available from the applied field. This may be a laboratory analogy to the cosmic ray production in flares.

Biermann: The limitations of the simple model suggested that Bagge and myself in 1949 were only in part considered in that paper. A discussion based on less limitations was carried further by Schlüter in 1952. In all my own later discussions of the subject of the electromagnetic acceleration to cosmic ray energies (e.g. in *Kosmische Strahlung*, ed. by Heisenberg (Springer, 1953), and in *Amer. Rev. Nuclear Sci.* vol. 2) care was taken to emphasize the limitations as well as the positive aspects of the theory.

Bostick: To return to the subject of solar prominences and a laboratory analogue of these, a pair of sources at the poles of a magnet produce a picture as given by Fig. 1. The streamers occur only when current flows from A to B or from B to A, but not otherwise.

Another experiment consists of firing plasmoids at a screen along a magnetic field with no current flowing between the screen and the plasma source. This produces on the screen a cross instead of a 'spot', that is, no streamers (see Fig. 2). Finally, a source placed near the end of a coil as shown in Fig. 3 produces several spots on a screen which is placed perpendicularly to the axis of the coil when a current of several hundred amperes flows from the source to the screen. The number of spots observed at the screen indicates that the current from the source is shredded into streamers directed along the magnetic-field

https://doi.org/10.1017/S0074180900237686 Published online by Cambridge University Press

lines. Along each streamer there presumably results a helical magnetic field, and this may possibly act as a plasma guide along which the projected plasma may travel. There exists a hypothesis to account for the shredding of current along a magnetic field into streamers, but it is too complicated to be developed here in a brief discussion.

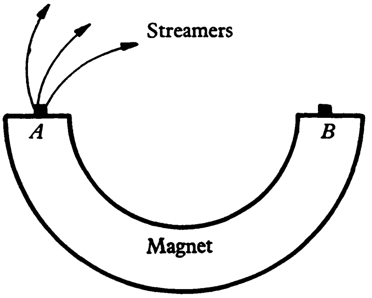

Fig. 1

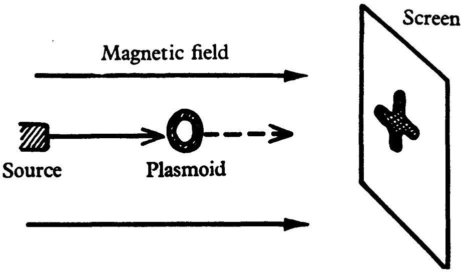

Fig. 2

Fig. 1. A pair of plasmoid sources (A, B) placed at the poles of a magnet. The streamers occur only when current flows from A to B or from B to A, but not otherwise.

Fig. 2. Firing of plasmoids at a screen, along a magnetic field. A 'cross' is produced at the screen.

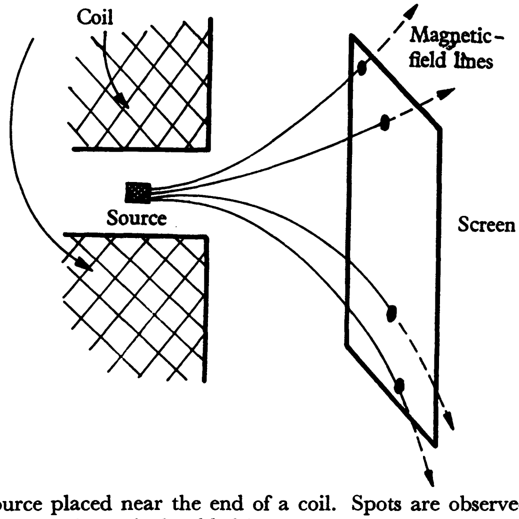

Fig. 3. Plasmoid source placed near the end of a coil. Spots are observed at a screen which indicates that the beam is shredded into streamers along the magnetic field.

Cowling: I should like to emphasize the difficulties involved in applying the results of experiments to cosmic problems. In the laboratory discharges are limited by tubes; there are electrodes to introduce potential differences; and gravity, which is certainly important in prominences, has very little influence. A careful examination of the effects of these and of all the scale effects involved is necessary before cosmic phenomena can be interpreted in the light of laboratory results.

https://doi.org/10.1017/S0074180900237686 Published online by Cambridge University Press