A history of early work on the heliospheric magnetic field

E. N. Parker
Enrico Fermi Institute, and Departments of Physics and Astronomy
University of Chicago, Chicago, Illinois

Abstract. The idea of a magnetic field in space around Earth began 400 years ago with Gilbert’s recognition that the magnetic field of Earth extends outward into space to form what we now call the magnetosphere. The concept of the solar wind and the heliosphere had its first glimmerings with the recognition, about 270 years ago, that geomagnetic activity is correlated with solar activity. It was suggested about a 100 years ago that the connection is through solar corpuscular radiation with velocities of the order of $10^3$ km/s. The observed acceleration of comet tails indicated the universal nature of solar corpuscular radiation, giving rise to the first speculation on the existence of the heliosphere. The hydrodynamic expansion of the million degree solar corona was then shown to provide the solar wind (solar corpuscular radiation), stretching out the magnetic fields of the Sun to fill the heliosphere with a spiral magnetic field. The advent of the space age soon verified the solar wind and magnetic field with direct measurements.

1. Early Concepts

The scientific pursuit of magnetic fields in interplanetary space (the heliosphere) began about 400 years ago with research on the form of the geomagnetic field. The crucial step was made by Gilbert [1600] who first recognized that Earth possesses a dipole magnetic field extending far out into space and declining in intensity as $1/r^3$ with distance $r$ from the center of Earth. Geomagnetic fluctuations were discovered and associated with auroral activity through the coordinated work of Graham [1724] and Celsius [1741]. De Mairan [1754] suggested that Earth orbits through the extended corona of the Sun, and it is the entry of solar coronal particles into the geomagnetic field that causes the fluctuations and the aurora. This prescient concept was based on the mistaken idea that the zodiacal light represents the outward extension of the corona observed close to the Sun during an eclipse of the Sun. The zodiacal light extends far beyond the orbit of Earth. As the reader will soon see, the association of the zodiacal light with interplanetary free electrons (a concept totally unknown to de Mairan) persisted until about 1962 when direct measurements in space showed otherwise. Brief histories of the development of geomagnetic activity [Chapman, 1967] and space science [Parker, 2000] are available in the literature.

It must be appreciated that interplanetary space was regarded as a hard vacuum by many physicists, the zodiacal light notwithstanding. Space was considered to be so empty that electrostatic potential differences of $10^6$ volts or more could be postulated without fear of ridicule. To cite an extreme case, Kelvin asserted that solar activity could not be the cause of geomagnetic fluctuations $\Delta B$ in spite of the overwhelming evidence of a connection with a 2-day delay. Kelvin considered $\Delta B$ at Earth as the superposition of a $\Delta B$ in a solar dipole magnetic field. Extrapolating back to the Sun with $1/r^2$ implied a $\Delta B$ at the Sun some $10^2$ times larger than the modest $\Delta B = 3 \times 10^3$ gauss observed at Earth, that is, $3 \times 10^6$ gauss at the Sun. Over the $10^{33}$ cm$^3$ volume of the Sun, this represents a magnetic energy of $4 \times 10^{29}$ ergs, equivalent to the solar output ($4 \times 10^{33}$ ergs/s) over a period of about 4 months, whereas the magnetic fluctuation $\Delta B$ occurs in an hour. Hence Kelvin concluded that a solar cause is impossible. He ignored the suggestion that the geomagnetic variations might be the result of a beam of solar corpuscular radiation.

The concept of isolated collimated (cold) beams of solar corpuscular emission of protons and electrons ($10^3$ km/s) developed through the last decade of the nineteenth century and the first half of the twentieth century [cf Chapman and Ferraro, 1933, 1940] as the cause of geomagnetic storms and enhanced aurora. The ideas began to take form with the discovery of the electron and the superficial resemblance of the auroral rays to cathode ray streamers in the laboratory Crooke’s tube. The velocity ($10^3$ km/s) of the corpuscular radiation was estimated from the delay of a couple of days between a flare event at the Sun and the associated geomagnetic activity. The ultimate step in the concept of solar corpuscular radiation was Biermann’s [1951, 1957] realization that the solar corpuscular radiation is responsible for the observed strong antisol solar acceleration of the gaseous comet tails. Since all comet tails show this acceleration, Biermann pointed out that the Sun emits solar corpuscular radiation in all directions at all times. That is to say, the corpuscular radiation is not confined to beams created by special events on the Sun, but rather is a universal and continuing property of the Sun. Assuming that the principal interaction of the corpuscular radiation with the comet tail gas is through charge exchange, Biermann estimated that the corpuscular radiation has a density of the order of 500-1000 protons and electrons/cm$^3$ at the orbit of Earth. This agreed with the 500 electrons/cm$^3$ estimated from the brightness of the zodiacal light on the assumption that the light is scattered sunlight from free electrons. In fact, the zodiacal light is principally sunlight scattered by dust grains. What is more, the solar corpuscular radiation interacts with the comet ions and electrons principally through the magnetic field carried with corpuscular radiation. However these quantitative points were not understood until later. The essential step was the recognition that the corpuscular radiation from the Sun is an everyday phenomenon and does not depend upon special flare events at the Sun.

Now there are two obvious places to look for the source of the interplanetary magnetic field. One is the interstellar or galactic...
magnetic field, which may extend through the solar system, and the other is the Sun itself. Hiltner [1949] and Hall [1949] observed the systematic plane polarization of starlight that has been reddened by passage through interstellar dust, indicating that nonspherical dust grains are preferentially aligned by some large-scale field. Spitzer and Tukey [1951] and Davis and Greenstein [1951] made the point that the only plausible field is a galactic magnetic field oriented more or less along the galactic arm. They proposed ferromagnetic grains or paramagnetic grains, respectively. It was not possible to deduce the field strength from the degree of polarization without precise knowledge of the composition and shape of the individual grains, but numbers of the order of \(10^2\) gauss were often mentioned. Fermi [1949, 1954] appealed to the galactic magnetic field to confine the galactic cosmic rays, and, in fact, suggested that cosmic ray particles are accelerated by bouncing off Alfvén waves and other moving disturbances in the galactic magnet field.

However, it soon became clear that the interstellar magnetic field does not penetrate into interplanetary space, because of the powerful outward sweep of the solar corpuscular radiation. With the numbers available at the time (500 protons and electrons/cm\(^2\) and 1000 km/s) Davis [1955] estimated that the interstellar magnetic field (\(10^{-2}\) gauss) is pushed away to a distance of several hundred AU. Hence, it would appear that the Sun must be the source of any interstellar magnetic fields. The magnetic fields of the Sun were first detected and measured by Hale [1908] through their Zeeman effect in sunspots, where the field strengths are commonly 2000-3000 gauss. Hale [1913] also thought that he detected polar fields of the general order of 40 gauss, indicating that the Sun has a general dipole magnetic field. In fact, when Babcock and Babcock [1955] began systematic mapping of the solar photospheric magnetic fields, the polar fields of the Sun proved to be only 10 gauss or less, reversing every 11 years near the peak of the sunspot cycle. Re-examination of Hale's old plates indicated a noise level of the order of 40 gauss.

2. Indirect Evidence

Now 50 years ago the only direct probes of interplanetary conditions were the solar corpuscular radiation impacting the geomagnetic field and the observed variations \(\Delta I\) [Forbush, 1937, 1954] of the galactic cosmic ray intensity. Unfortunately, the corpuscular radiation gave no clue as to the interstellar magnetic field, whereas the cosmic ray variations were clearly the result of some form of interplanetary electromagnetic effect on an otherwise steady external field of galactic cosmic rays. Forbush made his historic ground-based measurements with ion chambers, which respond primarily to the muons produced by collisions of the incoming protons at the top of the atmosphere. Thus the ion chambers track the intensity of incoming cosmic ray protons with energies of the order of 10-15 GeV and higher. Of particular interest was the occasional abrupt intensity drop of from a few percent over a period of hours, gradually recovering over the next several days, usually simultaneously with a geomagnetic storm, generally known as a Forbush decrease in honor of its discoverer. The similarity and close tracking of the two phenomena suggested to Forbush that the geomagnetic storm may somehow increase the geomagnetic cutoff energy at each latitude, thereby reducing the intensity of the incoming cosmic rays. But whatever the mechanism for the manipulation of the cosmic rays, the Forbush decrease clearly indicated the heavy hand of solar corpuscular radiation. Simpson [1951, 1954; Simpson, Fonger, and Treiman, 1953] developed the neutron monitor to measure the cosmic ray intensity at energies in the range 1 – 10 GeV, where the variations \(\Delta I\) are much larger. Simpson established five cosmic ray neutron monitors from Huancayo, Peru, on the geomagnetic equator, where the geomagnetic field admits particles only above about 10 GeV, to Chicago, where the geomagnetic cutoff is about 2 GeV. In this way he used the geomagnetic dipole field as a magnetic spectrometer. By intercalibrating and intercomparing \(\Delta I\) at these stations, he was able to show the energy dependence of the Forbush decrease and of the 11-year variation of the mean cosmic ray intensity with the magnetic sunspot cycle. He found that the energy dependence could be explained by neither an electrostatic potential in interplanetary space nor an increase in the geomagnetic cutoff energy. He noted, then, that the only remaining possibility is manipulation of interplanetary magnetic field by solar corpuscular radiation in such a way as to partially exclude the lower energy particles.

A number of magnetic effects were proposed. Alfvén [1956, 1958] pointed out that the abrupt expansion of Hale's 40-gauss dipole field of the Sun would adiabatically decelerate and dilute the cosmic rays, producing a reduction in measured intensity not unlike the Forbush decrease. A particularly naive idea was proposed by Parker [1956], who suggested that interstellar plasma clouds containing magnetic fields might be captured by the gravitational field of Earth, thereby forming a temporary magnetic barrier around the geomagnetic dipole and inhibiting the arrival of the lower energy cosmic rays. Fortunately, it was soon realized that the feeble gravitational field of Earth is grossly inadequate for the task and the idea was abandoned. Morrison [1959] suggested that a plasma cloud containing a closed tangled internal magnetic field is emitted by the Sun. Expanding out through interplanetary space, the cosmic rays within the cloud are adiabatically decelerated and diluted, so that a Forbush decrease is observed as the cloud engulfs Earth. Gold [1959] suggested that a plasma cloud is emitted from a bipolar active region on the Sun, stretching the bipolar field out through interplanetary space to provide an expanding magnetic tongue, whose interior would contain a decelerated and diminished cosmic ray intensity. The Forbush decrease occurs when the tongue engulfs Earth.

The great cosmic ray flare of February 23, 1956 provided an unanticipitated and effective opportunity for probing interplanetary magnetic conditions. Meyer, et al. [1956] used the time variations recorded by the five neutron monitor stations, plus a sixth on shipboard (on the way to Antarctica) in the harbor at Wellington, New Zealand to determine the energy spectrum. The prompt arrival of solar cosmic rays at precisely determined times at different terrestrial longitudes and latitudes showed that the first arrivals came straight from the Sun. The intensity peaked, and the directionality disappeared, after about 15 min, subsequently declining as \(t^{-3/2}\) for a few hours before going over into an exponential decline \(\exp(-t/T)\). It was clear that any interplanetary magnetic field between Earth and Sun was more or less radial. The simplest model for the declining phase was a uniform diffusive magnetic barrier beginning at about 1.5 AU and extending out to perhaps 5 AU.

Collectively, the cosmic ray variations suggested a variety of interplanetary magnetic field scenarios at different times. The indirect inferences from cosmic rays, comet tails, and geomagnetic variations were insufficient to pin down the precise form and behavior of the interplanetary magnetic field. Presumably, the magnetic field is dominated by the solar corpuscular radiation, but the origin of the corpuscular radiation...
was unclear. There were vague thoughts about corpuscular emission from sunspots, $M$ regions, flares, etc., all confounded by Biermann’s revelation that the Sun emits solar corpuscular radiation at all times and at all heliocentric latitudes.

3. Solar Wind

The next advance was made by Chapman [1959], who calculated the extremely high thermal conductivity and small thermal emission of the million degree coronal gas. He showed, then, that the corona, strongly bound by gravity near the Sun, extends far beyond the orbit of Earth, with the temperature $T$ declining outward only as $1/r^2$. The ghost of de Mairan must have been pleased with the result.

Parker [1958] noted two fundamental difficulties with the idea that Chapman’s corona could remain static. The first difficulty arose from integration of the equation for static equilibrium,

$$dp/dr + GM/Mr^2 = 0,$$

where $p = 2NkT$ represents the pressure of ionized hydrogen at a temperature $T$ and number density $N$, $M$ is the mass of the Sun, and $r$ is the mass of a hydrogen atom. The result is nonvanishing pressure $p$ at $r = \infty$ for the simple reason that far from the Sun the slow decline of the temperature leads to the thermal energy per ion exceeding the gravitational binding energy of the Sun.

The second difficulty was that a tenuous static corona would not permit the passage of Biermann’s universal solar corpuscular radiation. Even if there were no transverse magnetic fields in space, the two-stream plasma instability would quickly arrest the flow of one plasma through the other. It was evident that, somehow, the strongly bound and seemingly static corona near the Sun must become the corpuscular radiation at large distance. If the corona cannot be static, suppose, then, that there is a steady radial velocity $v(r)$, so that, instead of the static equilibrium equation, we have the momentum equation

$$NMv(r)/dr + dp/dr + GM/Mr^2 = 0,$$

with the condition that $Nvr^2 = constant$ for conservation of particles. With the million degree temperature extending far out into space, the only solution starting at small $r$ with small $v$ and strong gravitational binding and extending smoothly and continuously to $p = 0$ at $r = \infty$ is the supersonic expansion solution providing the solar wind [cf. Parker, 1958, 1963, 1965, 1969; Hundhausen, 1972; Roberts and Soward, 1972].

With the general hydrodynamic expansion of the solar corona as the origin of the solar corpuscular radiation it was immediately clear that the interplanetary magnetic field is made up of the solar magnetic fields wherever they are not strong enough at the Sun to prevent the free expansion of the corona [Parker, 1958]. That is to say, the active corona is forcibly confined by the strong $(-10^9$ gauss) bipolar magnetic fields of the active regions, whereas the corona is able to push its way out through the relatively weak $(-10$ gauss) fields of the broad active regions. So the interplanetary magnetic field arises primarily from the forced extension of the quiet region fields. Coronal expansion and the resulting solar wind create the heliosphere and fill it with magnetic field drawn out from the Sun and extended to the outer boundary at some 100 – 200 AU.

The basic character of the interplanetary or heliospheric magnetic field is easily illustrated [Parker, 1958] for the idealized case of a precisely radial solar wind with uniform velocity $v$ beyond some radius $r = a$ near the Sun. A smokeplot placed at some point $(\theta_a, \phi_a)$ on the sphere $r = a$ rotating with the angular velocity $\Omega$ of the Sun at the polar angle $\theta_a$ produces a line of smoke in the wind with the spiral conical form $\theta = \theta_a, r - a = (v/\Omega)(\phi - \phi_a)$. The magnetic field line of force through the smokeplot lies everywhere in the line of smoke, so that the field line has the same Archimedian spiral form as the line of smoke. The magnetic field intensity is determined by the field at $r = a$ by

$$B_r(r, \theta, \phi) = B_r(a, \theta, \phi - (r-a)/v) (a/r)^2,$$

$$B_\theta(r, \theta, \phi) = 0,$$

$$B_\phi(r, \theta, \phi) = B_r(r, \theta, \phi - (r-a)/v) a^2 \Omega \sin \theta/v.$$

In the real heliosphere the solar wind velocity is not uniform around the Sun, nor is it always constant in time at any given heliographic longitude of the rotating Sun. So the magnetic structure in interplanetary space is complicated by diverging flow (at high latitude), converging flow (at low latitude), fast spiral streams overtaking slow spiral streams, and blast waves (nowadays recognized as coronal mass ejections). These possibilities were sketched at some length [cf. Parker, 1963, Chaps. 9, 10, 11 and references therein; Sarathbai, 1963]. One of the most interesting aspects has turned out to be the overtaking of slow regions of wind by fast regions, providing forward and backward propagating shocks along the spiral zones of collision. Burlaga [1997] has followed these effects over the years as spacecraft have ventured farther out through the solar system, finding that the phenomenon provides a sawtooth structure of the solar wind and magnetic field at the orbit of Jupiter and beyond. The large fluctuations in the otherwise spiral magnetic field were investigated by Belcher and Davis [1971] and the dynamics of the deflection of the solar wind was treated by Carovillano and Siscoe [1969].

4. Space Observations

Observational studies of the solar wind and its magnetic field got underway with Explorer 10, on an orbit that carried it out to about 15 $R_E$ ($R_E$ is the radius of Earth) in the general direction of the Sun. The presumed solar wind density of 500 H ions/cm$^3$ at velocities of 500–1000 km/s suggested that the sunward boundary of the geomagnetic field was pushed in to about 5 $R_E$, so Explorer 10 would be well out in the wind for most of the time. It was baffling, therefore, to find only an intermittent wind of 4 – 8 H atoms/cm$^3$ and 200 – 400 km/s alternating with relatively steady magnetic fields of $10^{-4}$ gauss [Heppner et al., 1963] and no wind for irregular periods of the order of an hour. At first sight it suggested a solar wind loosely filled with magnetic filaments from the Sun. However, it was soon pointed out [Bonetti et al., 1963] that if the solar wind density is only 4–8 H atoms/cm$^3$, then the sunward magnetopause lies at 10–15 $R_E$ rather than 5 $R_E$. Thus Explorer 10 was moving along the bouncing sunward magnetopause, which alternately retreated, exposing the spacecraft to shocked solar wind, and advanced, placing the spacecraft within the shelter of the geomagnetic field. Bonetti et al. [1963] sketched the configuration, with the upstream bow shock and the comet shaped magnetosphere extending in the antisolar direction.

The next spacecraft to explore the interplanetary magnetic field was the Mariner 2 mission to Venus. It provided the first long-term quantitative record of the solar wind velocity and
density [Snyder and Neugebauer, 1964; Neugebauer and Snyder, 1966, 1967]. Unfortunately, the Mariner spacecraft was not magnetically clean, there being active electric current loops as well as magnetic structural material, so that the magnetometer was able to record rapid magnetic fluctuations $\Delta B \sim 0.6 \times 10^4$ gauss in the interplanetary magnetic field carried past the spacecraft in the 500 km/s wind [Davis et al., 1964], but there was no way to determine the mean field.

Thus it remained for the Interplanetary Monitoring Platform 1 to be carefully designed, under continued pressure and persuasion from N. F. Ness, to allow the necessary precision magnetic measurements of both the mean and the fluctuation components of the field. Thus it was that Ness et al. [1964] succeeded in making the first measurements of the mean interplanetary magnetic field at the orbit of Earth. The measured field fluctuated rapidly, with $\Delta B \sim B$. From their detailed measurements, however, they found a mean field $B \sim 0.6 \times 10^4$ gauss inclined to the radial direction by the expected spiral angle ($\psi = \tan^{-1}(r/\Delta v)$) of $30^\circ$ to $45^\circ$ depending upon the prevailing wind velocity at the time. Thus the interplanetary magnetic field is nearly radial inside the orbit of Earth and more nearly azimuthal beyond the orbit of Mars. One could begin to see the basis for the behavior of the solar cosmic rays of February 23, 1956.

The long-term monitoring of the interplanetary magnetic field got underway at that time, subsequently accumulating the variety of effects arising from the irregularities of the solar wind. The immediate interest, however, was in tracing the field back to its origins at the Sun. For it must be appreciated that the observations of the solar wind at the orbit of Earth and the observations of the corona and magnetic fields at the Sun do not show precisely what parts of the corona are able to expand to provide the solar wind observed at the orbit of Earth.

Note then that a simple $1/r^2$ radial extrapolation of a radial magnetic field back to the surface of the Sun provides a factor of about $5 \times 10^6$, so that the mean $0.6 \times 10^4$ gauss at the orbit of Earth becomes about 3 gauss at the Sun. In fact, the quiet region (and nowadays the coronal hole) fields are typically 10 gauss, but over only a fraction of the solar surface, so their fields spread out over the entire 4r solid angle filled by the solar wind at large distance from the Sun, providing an equivalent mean field at the Sun of only a few gauss. Thus the mean field at the orbit of Earth checks out right, as nearly as one can tell.

A curious feature of the interplanetary field was the tendency to point away from the Sun for about a week and then toward the Sun for about a week, then away again, etc., so that the field in the plane of the ecliptic, and presumably also in the equatorial plane of the Sun, showed four distinct sectors rotating with the 27 day period of the Sun (as viewed from the orbiting Earth and spacecraft). In some years there were only two sectors, each rotating by in about 2 weeks. A detailed study by Ness and Wilcox [1964; Wilcox and Ness, 1965, 1968] and Schatten et al. [1969] showed that the interplanetary magnetic field is stretched out from the large unipolar magnetic regions on the Sun. These regions do not fully respect the heliographic equator, so that an outward magnetic flux centered north of the equator may spill southward across the equator at one longitude, and an inward flux may spill northward across the equator somewhere else. The result for the field extended out into space near the equator is the alternating inward and outward sector structure (see review by Dessler [1967]). Rosenberg and Coleman [1969] showed that the sign of the interplanetary magnetic field traced back to the Sun along the spiral field lines correlates nicely with the Babcock magnetograms of the solar surface.

Subsequent studies showed that the sign of the interplanetary magnetic field correlates best with the fields of the Sun at latitudes of about $+30^\circ$ and $-30^\circ$, showing how the corona is confined by the strong magnetic fields of the active regions at low latitude so that the wind at Earth comes mainly from the free coronal expansion at middle latitudes [Wilcox, 1968; Wilcox and Howard, 1968; Wilcox and Colburn, 1969].

This brief history traces the investigation and thinking on conditions in space, especially the magnetic fields in space, from the early days of indirect probing and inferring to the later direct studies that became possible with the dawn of the space age. The hydrodynamic expansion of the solar corona proves to be the controlling factor, stretching out the weak regions of solar magnetic field all the way out through the heliosphere, through the presumed termination shock, and into the subsonic downstream wake in the local interstellar wind. The following paper by N. F. Ness picks up where the present historical narrative leaves off, carrying the space exploration of the interplanetary magnetic field into the fascinating details omitted in the gross picture outline here.

Acknowledgments. Janet G. Luhmann thanks both of the referees for their assistance in evaluating this paper.

References

Alfvén, H., The Sun’s general magnetic field, Tellus, 8, 1, 1956.


