48

ANDRÉ BALOGH* AND LENNARD A. FISK**

The heliosphere

The heliosphere is the vast region of space surrounding the Sun and the Solar System that is filled with the solar wind. Its inner boundary is the outer atmosphere of the Sun, the solar corona. Its outer boundary is defined, to a first approximation, as a surface across which there is a balance of pressure between the solar wind and the local interstellar medium (LISM). This outer boundary is to be found at probably about two and a half to three times the orbital distance of the furthest planet, Pluto. The main regions of the heliosphere, described in this chapter, are illustrated schematically in Figure 1, although there are many uncertainties concerning this simple picture. The objective of this chapter is to provide an overview of the heliosphere as understood, at the beginning of the twenty-first century, complete with our very comprehensive understanding of the solar wind and its large-scale dynamic structures that shape the heliosphere. Together with the development of solar wind studies, the study of the propagation of cosmic rays in the heliosphere has also been of great importance in shaping our understanding. These topics are reviewed at some length. The space missions that particularly contributed to the exploration of the heliosphere are also summarized.

The many topics related to the heliosphere have been the subject of several books, as well as numerous review articles. Of particular interest and relevance are the books edited by Schwenn and Marsch (1990, 1991) on the results of the Helios mission in the inner heliosphere; four books related to the three-dimensional heliosphere and the Ulysses mission (Marsden 1986, 1995, 2001; Balogh *et al*. 2001); a volume on cosmic winds and the heliosphere (Jokipii *et al*. 1997); three volumes published under the imprint of the International Space Science Institute, Bern, on the interstellar medium and the heliosphere (von Steiger *et al*. 1996), on cosmic rays in the heliosphere (Fisk *et al*. 1998), and on corotating interaction regions (Balogh *et al*. 1999). A volume that gathered his many contributions to heliospheric physics was published by Burlaga (1995). The outer heliosphere has been the subject of a volume edited by Grzedzielski and Page (1990). In addition, there have been, over the past 15 years, numerous review papers on aspects of the heliosphere; these, together with many original references are listed in this chapter.

THE EXISTENCE OF THE HELIOSPHERE AND ITS MAIN PROPERTIES

The existence of the heliosphere, as a volume of space controlled by the Sun, was first proposed by Davis (1955), who based this concept on the modulation of cosmic rays in anti-phase with the solar activity cycle. Cosmic rays are very high-energy particles which are accelerated (by vast shockwaves generated by supernova explosions) in the Galaxy, at considerable distances from the Solar System. The intensity of cosmic rays observed at the Earth is nearly constant, but is a few percent higher at sunspot minimum than at sunspot maximum. This 11-year modulation cycle can only be explained by assuming that the sunspot cycle has an effect on the space surrounding the Sun which, in turn, affects the propagation of cosmic rays into the inner Solar System, so that it is easier for cosmic rays to reach the Earth near solar minimum than around solar maximum. This explanation therefore implies that the medium surrounding the Sun undergoes changes between solar maximum and minimum. Given that these changes affect the propagation of high-energy cosmic rays (which are electrically charged particles), the changes in the medium around the Sun need to involve magnetic fields and, in effect, changes in the magnetic field itself, in response to the solar

* Imperial College, London, United Kingdom
** University of Michigan, Ann Arbor, MI, USA

André Balogh and Lennard A. Fisk, The Century of Space Science, 1141–1161

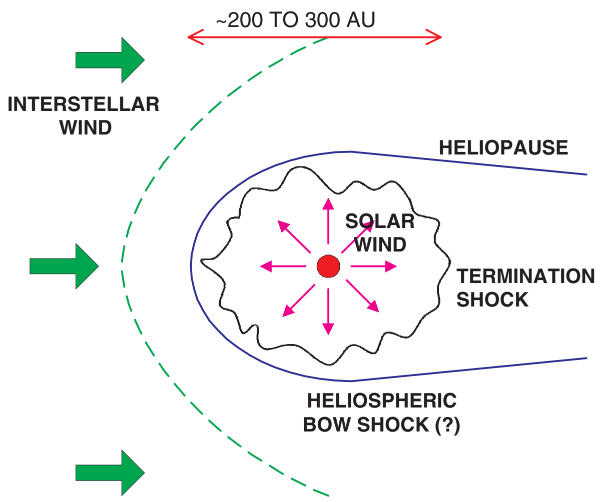

Figure 1 A schematic view of the heliosphere and its boundaries.

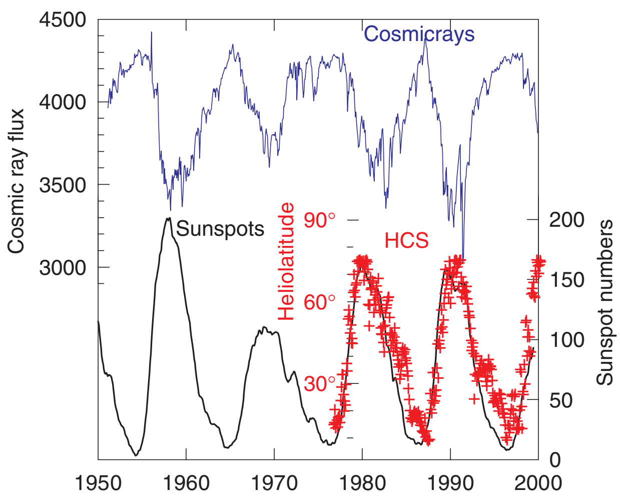

Figure 2 A composite plot showing the monthly number of sunspots (black trace), the daily average intensity of cosmic rays measured on the ground (blue trace) and the maximum heliolatitude of the coronal magnetic neutral line, representing the coronal imprint of the HCS (red symbols). There is clearly a very close correspondence between sunspots and the extension of the coronal magnetic neutral line; the anticorrelation between cosmic-ray intensity and sunspot numbers is indicative of the modulation of galactic cosmic rays by solar activity in the heliosphere.

activity cycle. Given also the large energies of cosmic-ray particles that are affected, the volume involved in the modulation also has to be large on the scale of interplanetary distances.

It is this relationship between cosmic-ray fluxes and sunspots that is illustrated in Figure 2, in which 50 years of observations, covering close to five solar cycles, are shown. The historical background to the early cosmic-ray modulation studies has been reviewed recently by Simpson (1998) and McDonald (2000). It is remarkable that the recognition of the relationship came so early, in the mid-1950s, even before it had been observed for a complete sunspot cycle. There is also a third quantity plotted in Figure 2; this quantity, labelled HCS in the figure, is the calculated maximum heliolatitude of the heliospheric current sheet, in effect the physical extension of the Sun's magnetic equator into the heliosphere. The nature and importance of the HCS in structuring the heliosphere are described below. However, its is clear from the close match of this quantity to the sunspot numbers (once the two are superimposed, as in the figure) that the variation in the heliolatitude of the Sun's magnetic equator through the solar cycle is an important indicator of solar activity and, given the need for a physical description of the solar modulation of cosmic rays, the extension of the magnetic equator into the heliosphere is at least a part of the physical mechanism of the modulation process.

The necessary advance in understanding the magnetized medium surrounding the Sun assumed by Davis (1955) came about at the same time, from the recognition that the Sun continually emits a 'corpuscular' radiation. The existence of such a radiation, to explain the geomagnetic effects that could be related to activity on the Sun, had been assumed much earlier by Chapman (see Chapman and Bartels 1940, and references therein). The existence of the solar wind, as a constant outflow of plasma from the solar corona was first suggested by Biermann (see Biermann, 1957, and references therein) from the study of the ion tails of comets. The current understanding of the reason for the existence of the solar wind as a permanent outflow of supersonic plasma originated with Parker's theoretical model (Parker 1958, 1963). Within four years of Parker's first theoretical predictions, the existence of the solar wind was confirmed by the early interplanetary space missions (for a summary of the early results, see e.g. Lüst 1967). This confirmation (that included the charting of the interplanetary magnetic fields necessary for affecting the propagation of cosmic rays) finally placed on a firm foundation the concept of the heliosphere (e.g. Axford 1972) and opened the way for its exploration in the space age, from the early 1960s.

The phenomenology of the heliosphere covers numerous space plasma phenomena on all scales, from the kinetic plasma scale where the dynamics of particle distributions dominate the physical processes, to the large magnetohydrodynamic (MHD) scales where the large-scale dynamics of the solar wind and the magnetic field define the global properties of the heliosphere. Space plasma processes on many intermediate scales provide a constantly changing, dynamic environment in the different regions of the heliosphere. The relative accessibility of the heliosphere to direct observations by deep-space probes has made it the largest astrophysical plasma regime that can be studied in detail. Heliospheric physics is now a discipline in its own right, with obvious and important links to solar physics, space plasma physics and, by extrapolation, with many topics in astrophysics in general.

The size of the heliosphere is not yet known precisely. The outer boundary is currently estimated to be somewhere between 80 and 120 AU from the Sun. Both the size of the heliosphere and the nature of the outer boundary depend on the properties of the solar wind at large distances from the Sun, on the properties of the LISM and on the physical processes involved in their interaction. Even though these properties are not known in sufficient detail to draw definitive conclusions on either the size of the heliosphere or the nature of the boundary, there are several estimates based on indirect observations which, surprisingly, agree quite well with each other. Such indirect evidence has also made it possible to introduce useful models that can be tested as further observations become available.

A great deal is known about the heliosphere from the Sun out to the distance of the furthest direct, *in situ* observations, currently to 74 AU by the Voyager 2 spacecraft. The *in situ* observations, made since the early 1960s by deep-space missions as well as by the continuous monitoring of the solar wind in the vicinity of the Earth, have provided a vast, even if in some respects selective database for determining the structure and dynamics of the inner and middle heliosphere. At the same time, progress in understanding the physics of the solar corona, the inner boundary of the heliosphere, has led to a good understanding of the important relationships between solar and heliospheric phenomena. Space missions have also played the major role in increased understanding of the complex dynamics of solar and coronal phenomena relevant to heliospheric physics.

The role of space missions in exploring the heliosphere has led to phenomenological descriptions that emphasize the dependence of heliospheric phenomena on distance from the Sun. Although any division of the heliosphere into different regions is somewhat arbitrary, the inner, the middle and the outer heliospheric regions have been distinguished, based mostly on the dominant dynamic features observed.

The inner heliosphere is the region from the solar corona to about the orbit of the Earth. In this region the coronal sources of the different solar wind streams recognizably dominate the dynamics of the medium. The middle heliosphere, from about the Earth's orbit to Saturn's orbit at 10 AU, is the region where the dynamic evolution of solar wind structures forms large-scale structures that begin to mask the solar origin of the different solar wind streams. In the outer heliosphere, the structures formed in the middle heliosphere continue to evolve, but dissipative processes and the intrusion of material from the LISM make the link between the observed structures and their solar origin recognizable only on the largest temporal and spatial scales.

There is, however, another equally important way to divide the heliosphere. For as much as five or six years around solar minimum in each solar cycle, the polar regions of the heliosphere have significantly different properties from the equatorial region. Relatively uniform, high-speed solar wind fills both polar regions, extending down to between 20 or 30 degrees within the equator. In the equatorial region, fast and slow solar wind streams intermingle and interact, making this region more structured and dynamic: it is this region which can be conveniently subdivided into the three regions, inner, middle and outer heliospheric regions described above. Due to the absence of direct observations in the polar regions at different heliocentric distances, its structure and evolution with heliocentric distance can only be extrapolated from the unique set of observations made at high heliolatitudes by the Ulysses spacecraft at distances of about 2 AU.

For several years in each 11-year solar activity cycle, around maximum activity, the heliosphere becomes considerably more complex than around solar minimum. This is due in part to the fragmentation of solar wind streams and in part to the considerable increase in the occurrence rate and intensity of solar transients, in particular coronal mass ejections (CMEs). Instead of the relatively simple distribution of the solar wind sources in the corona at solar minimum when they are divided into large-scale coronal holes emitting fast solar wind and the equatorial belt of slow solar wind, non-uniform but mostly slow solar wind is emitted from the whole corona at solar maximum. Short-lived, small-scale coronal holes still emit fast solar wind, but at lower speeds than wind from the large polar coronal holes near solar minimum. Superimposed on the slow solar wind from the corona, CMEs occur not only more frequently than at solar minimum, but are distributed more or less evenly at all latitudes. Until Ulysses explores the polar regions of the heliosphere during the current solar maximum activity period in 2000–02, it is possible only to extrapolate the near-ecliptic conditions to those regions.

The terms used in this chapter, referring to the heliospheric medium and the heliospheric magnetic field, intend to underline the three-dimensional nature of the heliosphere and its phenomena. Historically, the terms 'interplanetary medium' and 'interplanetary magnetic field' have been extensively used to describe phenomena in the space between the orbits of the planets. This use of the terms was justified as most of the space missions that made observations in the solar wind remained in orbits restricted to close to the ecliptic plane, the Earth's orbital plane. The main impetus for change came from the Ulysses mission, described below, with its objective to explore the three-dimensional properties of the heliosphere in its unique, polar orbit around the Sun. However, other missions, Pioneer 11 first, when in transit between Jupiter and Saturn, and, later on, the two Voyager spacecraft left the ecliptic plane, following their final planetary encounters, reaching now far beyond the furthest planets in the Solar System. It has become therefore more appropriate to use the "heliospheric" terminology when describing phenomena in the three-dimensional volume around the Sun.

THE EXPLORATION OF THE HELIOSPHERE

Space probes have played a key role in the exploration of the heliosphere. Since the early 1960s, numerous missions have explored different regions in the vicinity of the Earth; in the distant, outer heliosphere; between the Sun and the Earth; and above the poles of the Sun. In the following, a few key missions that contributed in a significant way to our knowledge of the properties and structures of the heliospheric medium are described, highlighting their contributions to our knowledge of the heliosphere.

The Soviet space missions to the Moon (Lunik 1 and 2) carried plasma detectors that first observed the solar wind. However, the first extensive set of data confirming the existence of the solar wind and its main properties, such as its velocity, density, temperature, as well as its variability came from the Mariner 2 mission to Venus in 1962, launched on 27 August 1962 (Neugebauer and Snyder 1962). Some of the properties of the embedded magnetic field were also observed on this mission (Coleman *et al*. 1962), in particular the general agreement of the orientation of the magnetic field with Parker's predicted spiral geometry (Davis *et al*. 1966).

In the 1960s, early Earth-orbiting spacecraft, with apogees reaching beyond the magnetosphere, provided further evidence concerning many of the basic properties of the solar wind and the magnetic field embedded in it. All the most important phenomena that shape and characterize the heliospheric medium were first noted by Earth-orbiting spacecraft. In particular, good agreement was found in general between the Archimedean spiral structure of the magnetic field lines proposed by Parker (1958, 1963) and the observed orientation of the magnetic field (Davis *et al*. 1966; Ness and Wilcox 1964) on the Interplanetary Monitoring Platform (IMP 1) mission. On the same mission, the sector structure of the magnetic field, showing a recurring pattern of alternating polarities of the field as a function of solar longitude, was also identified (Wilcox and Ness 1964). A later spacecraft in the same series, the remarkable IMP 8, launched in 1974, has continued to provide observations in the solar wind for nearly 30 years. These observations now constitute a historic data set that is used for studying long-term trends in solar wind and cosmic ray parameters.

The first space mission that provided a constant monitoring of the solar wind, the heliospheric magnetic field, as well as energetic charged particles and cosmic rays in the vicinity of the Earth, yet removed from the influence of terrestrial effects, was the International Sun-Earth Explorer (ISEE 3) mission. This was the first spacecraft to be launched to orbit one of the Sun–Earth system's gravitational 'neutral' points, named after Lagrange. The Lagrange point L1 is situated at about 1% of the distance (or about 1,500,000 km) between the Sun and the Earth, ahead of the Earth. An orbit around this point in space allows the spacecraft to monitor the solar wind, and even to provide a warning of 30 to 60 minutes if a solar storm (in effect a CME) is about to hit the Earth's magnetosphere. The ISEE 3 spacecraft, launched on 12 August 1978, remained in this orbit for four years, through the solar maximum in 1979–80, before becoming the prototype of space wanderers by being retargeted first to the Earth's distant magnetospheric tail and then, following some spectacular manoeuvres, to the first spacecraft encounter with a comet in 1985.

The L1 orbit, first used by ISEE 3, has been used very successfully, more recently (since the mid-1990s) by the Solar and Heliospheric Observatory (SOHO) and the Advanced Charge Composition (ACE) spacecraft. SOHO was launched to provide a comprehensive and often spectacular monitoring of the Sun and its corona, and ACE is following up, two solar cycles after ISEE 3, and with a more up-to-date instrumentation, the monitoring of the solar wind and its magnetic field.

The Pioneer series of missions, from the early 1960s, have explored the region of space between the planets Mars and Venus, as well as the distant heliosphere past Jupiter and Saturn. The first spacecraft to the outer planets were the Pioneer 10 and 11 missions. Pioneer 10 was launched on 2 March 1972; it reached Jupiter on 3 December 1973 to make the first close-up observations of the giant planet. After its encounter with Jupiter, the spacecraft followed a trajectory to the outer reaches of the heliosphere. Pioneer 11 was launched on 5 April 1973 and it reached Jupiter on 2 December 1974. At Jupiter, the spacecraft was targeted in such a way that the flyby would allow it to reach Saturn. This involved a trajectory that, for the first time, reached a heliolatitude of 16° above the ecliptic plane. The two missions were terminated in March 1997 and November 1995, respectively. At the time when the missions ended, Pioneer 10 was at a heliocentric distance of 67 AU, while Pioneer 11 was at 43 AU.

These two spacecraft were the first to make detailed observations of the heliospheric medium out to Saturn and beyond. The first phase of the Pioneer mission took place around the minimum activity period in solar cycle 21, in the mid-1970s. They were the first to observe the recurring sequence of corotating interaction regions (CIRs) and their

Springer
http://www.springer.com/978-0-7923-7196-0

The Century of Space Science
Bleeker, J.A.; Geiss, J.; Huber, M. (Eds.)
2001, XLIX, 1846 p., Hardcover
ISBN: 978-0-7923-7196-0