The Heliosphere and Its Boundaries

A. Balogh[a] and V. Izmodenov[b]
[a]*The Blackett Laboratory, Imperial College, London, UK*
[b]*Faculty of Mechanics and Mathematics,*
Lomonosov Moscow State University, Moscow, Russia

Introduction

The concept of a large volume of space, surrounding the Sun, in which the Sun controls the properties of the medium was first suggested by L.E. Davis in 1955[1]. This suggestion predates the theoretical prediction of the existence of the solar wind[2] and its observation by the first interplanetary missions[3]. The phenomenon that led to the idea of a "heliosphere", as this volume of space was named, is the modulation of the intensity of cosmic rays in anti-phase with the well-known 11-year solar activity cycle[4]. The intensity of cosmic rays is highest near solar minimum and is lowest around solar maximum. Cosmic rays are high-energy particles that are accelerated by shock waves driven through interstellar space by supernova explosions and fill the Galaxy quite uniformly. For the Sun to influence their intensity near Earth, it is necessary that large-scale magnetic fields originating in the Sun fill interplanetary space and that these fields (and the plasma flow - the solar wind - that carries them) vary with the solar cycle.

The implication of the intensity modulation of cosmic rays is that there is a cavity in interstellar space filled by solar material. There are several ways in which this cavity can impede the access of cosmic rays to the vicinity of the Earth. Simply listed, these are variations in (a) the properties of the medium in the cavity, (b) the size of the cavity, and (c) the properties of the boundaries of the cavity. In reality, the likelihood is that all three affect the access of cosmic rays. In this brief review of the heliosphere, we discuss these three aspects of the heliospheric cavity.

A simple diagram of its overall shape is sketched in Figure 1[5]. The heliospheric cavity is formed by the interaction of the solar wind with the Local Interstellar Medium (LISM)[6]. It is usual to make a distinction between the inner heliosphere, from the Sun to about the orbit of Saturn, and the outer heliosphere, reaching out to its boundaries. The structure of the interaction region, often called the heliospheric interface, is shown in Figure 1. The first boundary is the termination

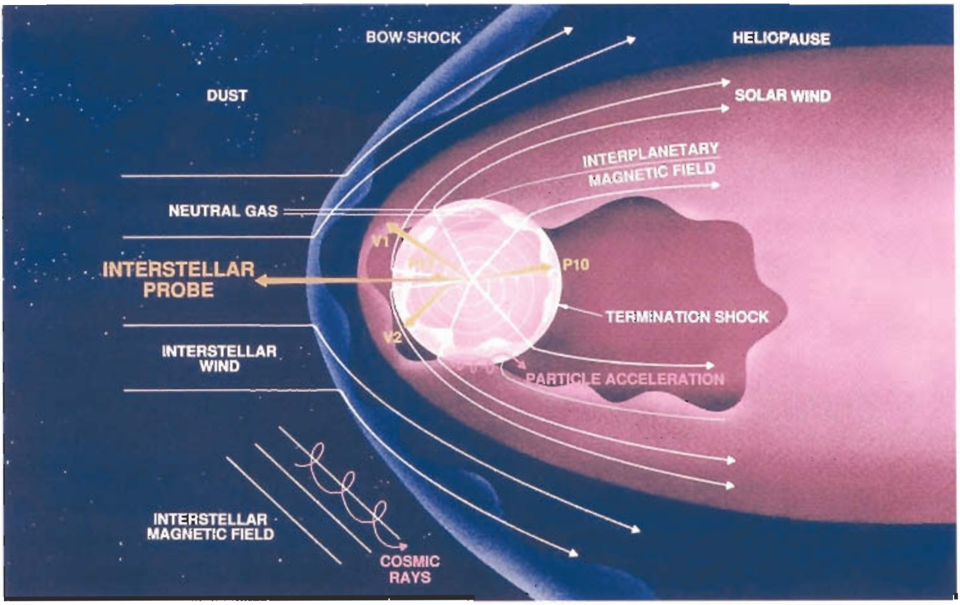

Figure 1. The heliosphere and the heliospheric interface. The heliopause is a contact discontinuity which separates the solar wind from interstellar plasma component. The termination and bow shocks are formed to decelerate the supersonic solar and interstellar winds, respectively, before these ionised gases reach the heliopause. The interaction region is often called *the heliospheric interface*. Interstellar atoms can easily penetrate through the heliospheric interface into the heliosphere, because their mean free path is comparable with the size of the heliospheric interface. (Courtesy of NASA/JPL)

shock, where the solar wind turns from a supersonic to a subsonic flow. Further out, the boundary between interstellar and solar matter is the heliopause. Although not directly observed so far, there are very firm theoretical grounds for the existence of these two boundaries. It is less clear whether there is a shock wave that would form opposite the direction of the interstellar wind; its existence depends on the velocity, density and temperature of the Local Interstellar Medium with respect to the heliosphere. It is possible now to envisage a space mission to the heliospheric boundaries and beyond, into the Local Interstellar Cloud and the Voyager spacecraft are already clearly approaching the inner boundary, so that direct observations in the future will place current theoretical ideas on the boundaries on a firm grounding.

The Heliosphere at Solar Minimum

The inner boundary of the heliosphere is the solar corona. The solar wind, a supersonic flow of tenuous plasma originating in the solar corona, fills and defines the volume of the heliosphere[3,7]. There are two kinds of solar wind. In space-based observations, they are distinguished by their speed, density, temperature and, most importantly, elemental and charge composition. In terms of their origin in the solar corona, the two kinds of winds are distinguished by the magnetic structure of the underlying coronal regions. Fast solar wind (usually with speeds > 600 km s^{-1}, low density, lower temperature) originates in coronal holes, regions that are cooler in the corona and have open magnetic field lines. The coronal regions with which low-speed solar wind (speed < 500 km s^{-1}, higher density and higher temperature) is associated are more complex in terms of their magnetic structure, generally in the form of closed loops. Coronal matter may escape at the edges of closed loops, or through opening of loops as a result of complex acceleration processes that are likely to involve magnetic reconnection, a process whereby closed magnetic field lines open and release the plasma usually confined in the loops. The inner heliosphere, between the Sun and the Earth where the structure of the heliospheric medium is still closely related to the coronal regions in which the solar wind originates, was explored in the 1970s by the Helios mission[7]. The compositional differences between fast and slow solar wind include different elemental abundance ratios (the so-called "First Ionisation Potential", or FIP effect) and different degrees of ionisation of the plasma ions, corresponding to the different coronal temperatures in which the solar wind originates[8]. The solar wind also carries into the heliosphere magnetic field lines that are rooted in the Sun[9]. Magnetic field lines in the heliosphere form the large-scale structures; these structures and fluctuations in the strength and direction of the magnetic field affect the propagation of cosmic rays.

The contrasting appearances of the solar corona at solar minimum and maximum activity are illustrated in Figure 2. At the minimum in the solar activity cycle,

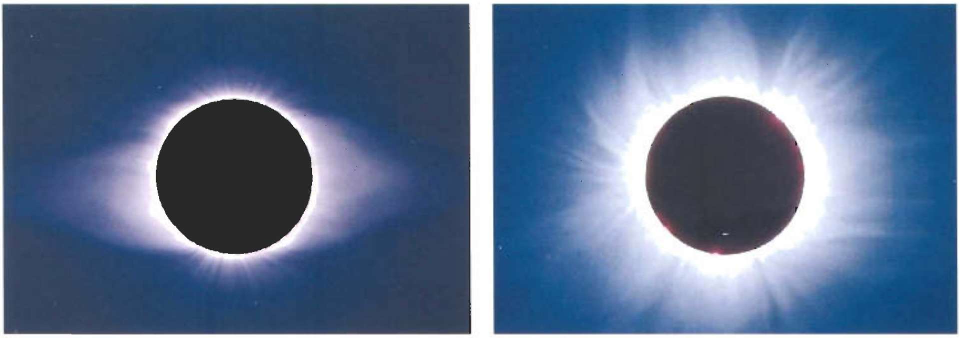

Figure 2. Solar-eclipse photographs of the solar corona. On the left, the corona at solar minimum on 24 October 1995, which shows the magnetically closed equatorial streamers and the large polar coronal holes with the characteristic polar plumes. On the right, the corona on 11 August 1999, close to solar maximum, which shows the complex, mostly closed coronal magnetic regions at all heliolatitudes. These photographs illustrate the different boundary conditions for the solar wind and the heliosphere at solar minimum and maximum. (Photos used by permission of Fred Espenak.)

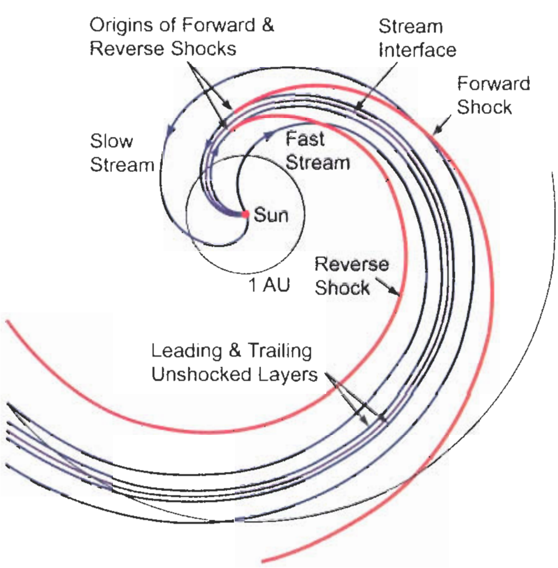

Figure 3. The equatorial projection of Corotating Interaction Regions, illustrating their main features (after Crooker *et al.*, in Ref. 11). The shock fronts are in red, magnetic field lines in blue, and the contact surface between two plasmas in magenta.

large and stable coronal holes form in the polar regions of the Sun. Closed magnetic field loops can be found generally near the Sun's equatorial region. Fast solar wind from the polar coronal holes fills most of the heliosphere at that time, except near the Sun's equator, where both fast and slow streams are emitted. The structure of the heliosphere in the years surrounding solar minimum is relatively simple: at heliolatitudes away from the equatorial region the heliosphere is filled with uniform high speed solar wind, while near the equator both slow and fast speed streams are present. Most of what is known of the heliosphere in three dimensions comes from the observations made by the Ulysses spacecraft[10].

The two different kinds of solar wind are emitted from the corona radially in distinct streams. As the Sun rotates, the streams of different speeds interact, faster solar wind catching up with the preceding slow solar wind stream to form Corotating Interaction Regions, or CIRs[10, 11]. Figure 3 shows the key features of CIRs projected into the solar equatorial plane. CIRs are formed within a latitude band about ±30° around the Sun's equatorial plane, at times approaching solar minimum, when the sources of the solar wind remain stable over many solar rotation periods. The shock waves that are formed leading and trailing the CIRs are important features in the acceleration of energetic particles, but also impede the access of cosmic rays near the solar equatorial plane at solar minimum.

Beyond a distance of about 10 AU, clear signatures of CIRs are no longer found, as successive interaction regions coalesce and propagate as large-scale pressure pulses towards the outer regions of the heliosphere. In the outer heliosphere, a new phenomenon, the effect of "pickup" ions becomes an important feature. Pickup ions are of interstellar origin, mostly hydrogen atoms that have penetrated the heliosphere and then become ionised, either through a process of charge exchange with solar-wind protons or by photo-ionisation. Once ionised, the ions are "picked-up" by the heliospheric magnetic field and carried in the outward-flowing solar wind. Their effect is to slow down the solar wind, but also to increase the overall pressure; in addition, the pickup process also contributes to heating of the solar wind. Pickup ions increasingly influence the properties of the solar wind when it propagates towards the boundaries of the heliosphere[12].

The largest connected structure in the heliosphere is the Heliospheric Current Sheet (HCS)[13]. This is a large surface that extends from the corona and represents the boundary between the two polarities of magnetic field lines ("away" and "towards" the Sun) as they are carried out from the corona into the heliosphere by the solar wind. Near solar minimum, the open coronal magnetic fields correspond to a near-axial magnetic dipole, so that the HCS is a slightly warped surface close to the solar equatorial plane; the warps in the surface are due to non-axial or higher order terms in the solar magnetic field. Around solar maximum, open coronal magnetic fields present a much more complex picture and the HCS becomes highly inclined, significantly warped and changeable on the timescale of solar rotations. For a space-based observer, the HCS separates the dominant magnetic polarities, leading to the so-called "magnetic sector structure".

The Heliosphere at Solar Maximum: Coronal Mass Ejections

In the years around solar maximum activity, the solar corona and the solar wind have a considerably more complex structure than at solar minimum. Most of the corona consists of closed magnetic loops; coronal holes are small and transient. As a result, the solar wind is generally slow and highly variable. The structure and dynamics of the heliosphere reflect the coronal complexity, generally forming frequent but small-scale and transient interaction regions[14].

The large-scale disruption of solar coronal magnetic fields at times of solar storms results in Coronal Mass Ejections (CMEs) that inject both large amounts of coronal plasma (~10^{12} kg) and complex magnetic-field structures into the ambient solar wind. CMEs are infrequent near solar minimum, but their frequency and size increase considerably (up to about one a day) for several years around solar maximum. As a result, CMEs significantly influence the structure of the heliosphere. Their passage through the ambient solar wind results in modifications to its density, temperature and composition. Their magnetic structures are frequently in the form of loops, or magnetic clouds. As they propagate away from the Sun, CMEs can interact with one another, eventually forming Merged Interaction Regions (MIRs) that act as barriers to cosmic rays[4,15]. A sketch of a

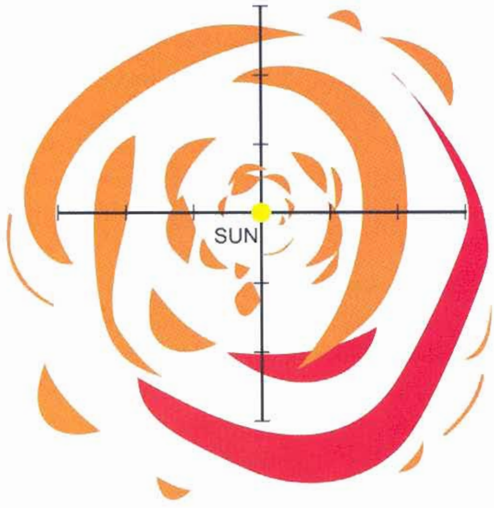

Figure 4. In the years surrounding solar maximum activity, Coronal Mass Ejections are emitted frequently, and as they propagate away from the Sun, plasma and magnetic structures associated with them fill a significant volume of the heliosphere.

cut through the heliosphere in Figure 4, schematically showing CMEs near solar maximum as they propagate away from the Sun, indicates the way in which frequent CMEs emitted at all heliolatitudes can populate the heliosphere.

In the distant heliosphere, the two Voyager spacecraft, launched in 1977, are currently approaching the expected first boundary, the termination shock. Both spacecraft observe magnetic fields and solar-wind structures that are generally the complex end-products of the long dynamic evolution from the Sun; the travel time of the solar wind to 70 AU is more than 300 days at a speed of ~400 km s^{-1}. Following the last solar maximum in 2000, Voyager 2 observed large fluctuations in the dynamic pressure of the solar wind, as shown in Figure 5. Successive large CMEs in the years after solar maximum have formed dynamically

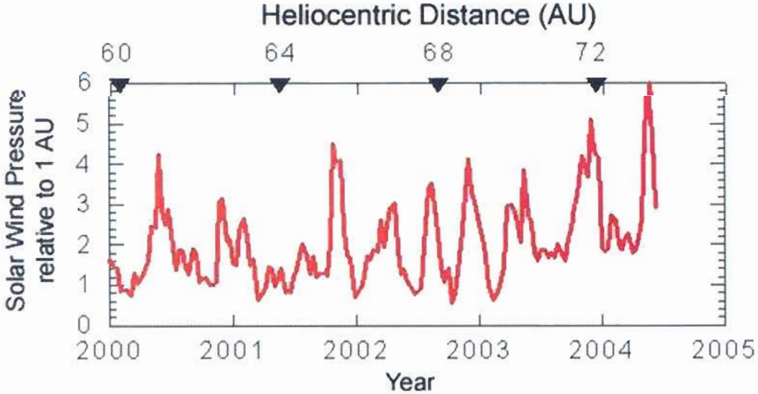

Figure 5. The variable pressure of the solar wind in the outer heliosphere, as measured by the Voyager 2 spacecraft in the years following the last solar maximum. (The pressure has been normalised to 1 AU.) The pressure pulses arise from the coalescing remnants of Coronal Mass Ejections and affect the position of the solar-wind termination shock. (Data courtesy of J.D. Richardson, MIT.)

combined pressure pulses that are a factor of up to 8 above the average, background solar-wind pressure. Such pressure waves continue to propagate towards the heliospheric boundaries and influence the location and dynamics of the termination shock.

The Boundaries of the Heliosphere

The heliospheric boundary lies beyond the Solar System at distances ~120 - 150 AU. This is the most distant and most unknown region in the heliosphere. The structure of the heliospheric boundary is determined by the interaction of the solar wind with the interstellar neighbourhood of the Sun – the Local Interstellar Cloud (LIC) consists of the termination shock, the heliopause and the bow shock (Fig. 1). At present there is no doubt that the LIC is a partly ionized cloud with size of several parsecs. This cloud is a part of a small group of partly ionized clouds, which is embedded in the hot (~10^6 K) and rarefied (~0.001 cm^{-3}) gas, the Local Bubble (Fig. 6). At present still there is no complete consensus between scientists on the origin of the Local Bubble.

The LIC temperature (~6700 K) and velocity (~26 km/s) can be inferred independently from direct measurements of interstellar helium atoms by the Ulysses/GAS instrument[17] and from analysis of absorption features in the stellar spectra[18]. The first method is based on the fact that the atoms of interstellar heli-

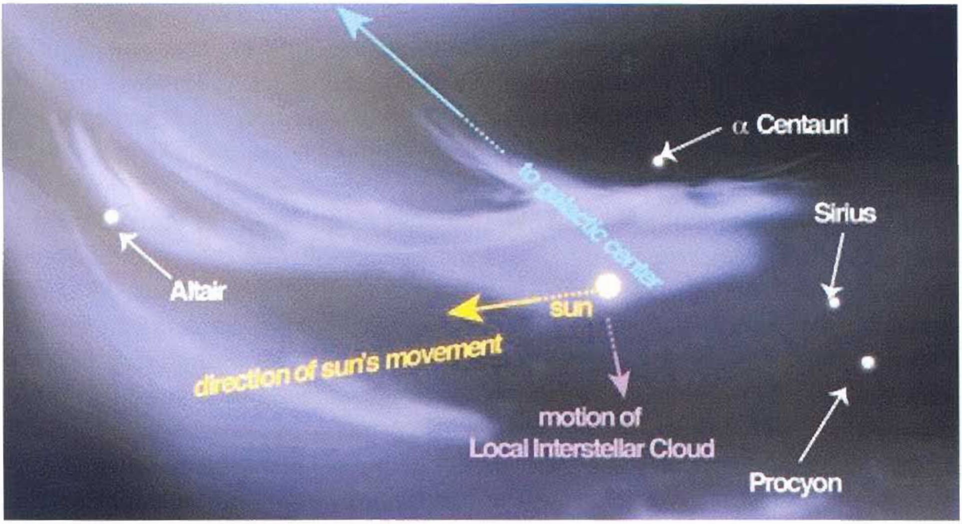

Figure 6. The structure of the Interstellar Medium in the vicinity of the Sun – within 10 light-years. The Sun is embedded into the partly ionized Local Interstellar Cloud (LIC). Directions of the Sun's motion, the LIC motion and towards the galactic centre are shown. (After P. Frisch, Ref. 16.)

um penetrate the LIC/SW interaction region undisturbed due to their large mean free path. Therefore, being measured inside the heliosphere at one or several astronomical units, the speed and temperature of these atoms correspond to the speed and temperature of pristine interstellar medium. Despite the second method – interstellar absorption study – providing mean values along lines of sight toward nearby stars in the LIC, a comparison of local interstellar temperatures and velocities derived from stellar absorption with those derived from direct measurements of interstellar helium shows quite good agreement. Other parameters of the LIC – such as its ionization, composition, magnetic field direction and intensity – are not observed directly and, therefore, are less known.

At the present time, the heliospheric interface structure and local interstellar parameters can only be explored with remote indirect measurements. To reconstruct the structure of the interface and the physical processes inside the interface using remote observations at one to several astronomical units, a theoretical model should be employed. Theoretical studies of the heliospheric interface have been performed for more than four decades, following the pioneering papers by E. Parker and V. Baranov[19]. However, a complete theoretical model of the heliospheric interface has not yet been constructed. The difficulty in doing this is connected with the multi-component nature of both the LIC and the solar wind. The LIC consists of at least five components: plasma (electrons, protons, and singly ionized helium), hydrogen atoms, interstellar magnetic field, galactic cosmic rays, and interstellar dust. The heliospheric plasma consists of original solar particles (protons, electrons, alpha particles, etc.), pickup ions and the anomalous cosmic-ray component (ACR). Pickup ions modify the heliospheric plasma flow starting from ~ 20-30 AU. ACRs may also modify the plasma flow upstream of the termination shock and in the heliosheath. Spectra of ACRs measured by Voyagers can serve as remote diagnostics of the termination shock strength and location[20].

To develop a theoretical model of the heliospheric interface, it is necessary to choose a specific approach for each of the interstellar and solar-wind components. Interstellar and solar-wind protons and electrons can be described as fluids. At the same time, the mean free path of interstellar H atoms is comparable with the size of the heliospheric interface. This requires kinetic description for the interstellar H atom flow in the interaction region. For the pickup-ion and cosmic-ray components, the kinetic approach is also required. Big progress in the development of multi-component models of the heliospheric interface and application of the model to the interpretations of different remote diagnostics of the interface was done in the frame of several ISSI and two INTAS-ISSI teams. Recent reviews on the modelling of the heliospheric interface can be found in papers by Izmodenov and others[21].

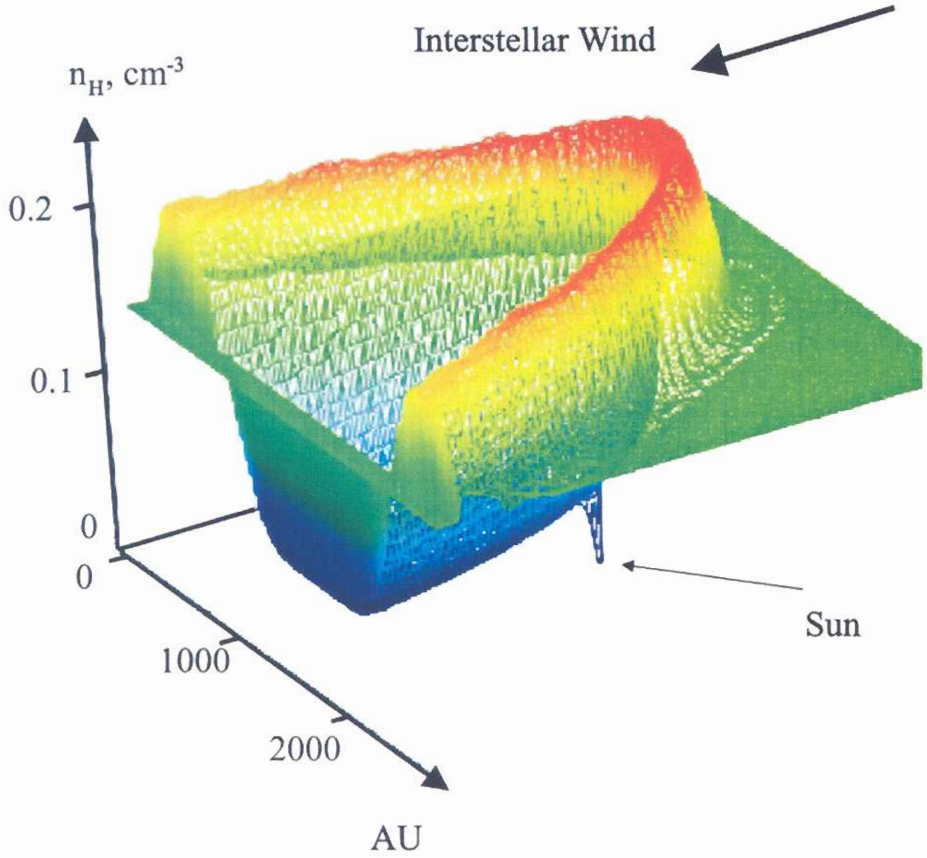

Figure 7. Hydrogen wall – an increase in the number density of interstellar H atoms around the heliopause of the Sun. The hydrogen wall is created due to charge exchange of primary H atoms with interstellar protons decelerated in the vicinity of the heliopause. (After Gruntman *et al.*, Ref. 32.)

One of the important findings obtained firstly theoretically by Baranov & Malama[22] is the existence of the interstellar hydrogen wall – an increase in the density of interstellar H atoms around the heliopause (Fig. 7). The existence of the hydrogen wall was confirmed later by observations. It was shown by Linsky & Wood[23] that the absorption spectra towards the Alpha Centauri cannot be explained without assuming the existence of the absorption produced by the "H wall" or heliospheric absorption. Later, the heliospheric absorption was detected towards several other stars. Absorption spectra toward both Alpha Centauri and Sirius detect existence of "H walls" around the stars[24]. The existence of the "H wall" around a star requires presence of the stellar wind and a partially ionized interstellar gas moving with respect to the star. Therefore, an analysis of absorption becomes a new tool to detect the stellar wind. For the stars similar to the Sun such winds had not been detected before.

Other major sources of information on the heliospheric interface structure and position of the termination shock are the following: (1) direct measurements of interstellar pickup ions, which are interstellar atoms ionized in the heliosphere by charge exchange and photo-ionization and measured by Ulysses and ACE spacecraft; (2) anomalous cosmic rays - those pickup ions that are accelerated to high energies and measured by Voyagers, Pioneers, Ulysses, ACE, SAMPEX and Wind; (3) backscattered (by interstellar atoms of hydrogen) solar Lyman-α radiation measured at 1 AU by SOHO/SWAN, Hubble Space Telescope (HST), and in the outer heliosphere by Voyager and Pioneer spacecraft; (4) direct measurements of the solar wind at large heliocentric distances by Voyager 2 spacecraft; (5) first detections of heliospheric energetic atoms (ENAs) by SOHO and IMAGE, which proved that the detailed imaging of the heliospheric interface in ENAs will be possible in the near future.

The observations together with modelling could provide constraints on both less-known interstellar parameters – such as interstellar proton and H atom number density and upper limits for interstellar magnetic field – and physical process in the heliospheric interface. One of the many important results of such a study is determination of the local interstellar abundances of the heavier elements and their isotopes. These abundances are of fundamental interest for cosmological models. The measurements of the interstellar abundance now becomes possible due to measurements of pickup ions of H^+, $^4He^+$, $^4He^{2+}$, $^3He^+$, N^+, O^+, Ne^+ by SWICS (Solar Wind and Interstellar Composition Spectrometer) instrument onboard Ulysses[25]. To obtain the local interstellar cosmic abundances of the elements, which are strongly coupled to plasma (as, for example, hydrogen and oxygen), a model of penetration of these elements through the interface is employed[26].

Another discovery connected with the heliospheric boundaries is the presence of radio emission in the 2-3 kHz range, first detected in 1983 and later in 1992-93 by the Plasma Wave Subsystem (PWS) of the two Voyagers[27]. This emission is associated with the heliopause and considered as an echo from the impacts of the CMEs on the heliospheric boundaries. The time delay between the solar ejection of the CMEs and the radio-emission detection allows estimation of the distance to the heliopause at 150-160 AU[27], which corresponds to the distances to the heliopause obtained in modern models of the interface[28].

Based on measurements of the low-energy particle fluxes, spectra, and composition by the Voyager 1 Low-Energy Charged Particle instrument and of an indirect determination of the solar-wind speed using particle anisotropy measurements, Krimigis *et al.*[29] reported the probable crossing of the Termination Shock by Voyager 1 at 85 AU in the summer of 2002 and the return to the TS upstream region about 6 months later. McDonald *et al.*[30] suggested another interpretation of the Voyager data, arguing that the spacecraft remained in the supersonic wind, but in the precursor region. In any case, recent Voyager 1 data suggest that the TS was close to 85 AU in the Voyager 1 direction. To compare this evidence with model predictions, the distance to the TS in the middle of 2002 into the Voyager direction for different interstellar proton and H atom number densities was computed in Reference 28. For ($n_{H,LIC}$, $n_{p,LIC}$) comparable to the ionization range of interstellar helium[31] of 0.3 - 0.4, the TS location is 104 ± 4 AU, which is ~20 AU farther from the Sun than Voyager 1. One possible solution to get the TS at ~85 AU in the model is to increase the interstellar atom and proton number densities. Model calculations show that for $n_{p,LIC} = 0.11\text{-}0.12\ \mathrm{cm}^{-3}$ and $n_{H,LIC} \sim 0.22\ \mathrm{cm}^{-3}$, the TS was at 85-86 AU in 2002 and the number density of H atoms at the TS, $n_{H,TS}$ is $\sim 0.1\ \mathrm{cm}^{-3}$, which is in agreement with the value deduced from Ulysses/SWICS observations of pickup protons[25], but contradicts 30-40 % of interstellar helium ionization.

Finally, growing interest in heliospheric interface studies is connected with expectations that Voyager 1 recently crossed the termination shock or will cross the shock in the near future. Many predictions of the time of the termination shock being crossed by Voyager appeared in the literature. However, it seems that much more work should be done to explain and reconcile all available indirect observations of the heliospheric interface based on the unique model of the heliospheric interface. This work should be done especially because NASA plans to explore the interaction region remotely using ENA imaging[32] (HIGO, or the future NASA Interstellar Boundary Explorer (IBEX) mission scheduled for launch in 2008) and to send a spacecraft (the Interstellar Probe) to a heliocentric distance of at least 200 AU with a flight-time of only 10 or 15 years. Intensive theoretical study will help to optimize goals, instrumentation and, finally, the scientific profit from the "interstellar" missions.