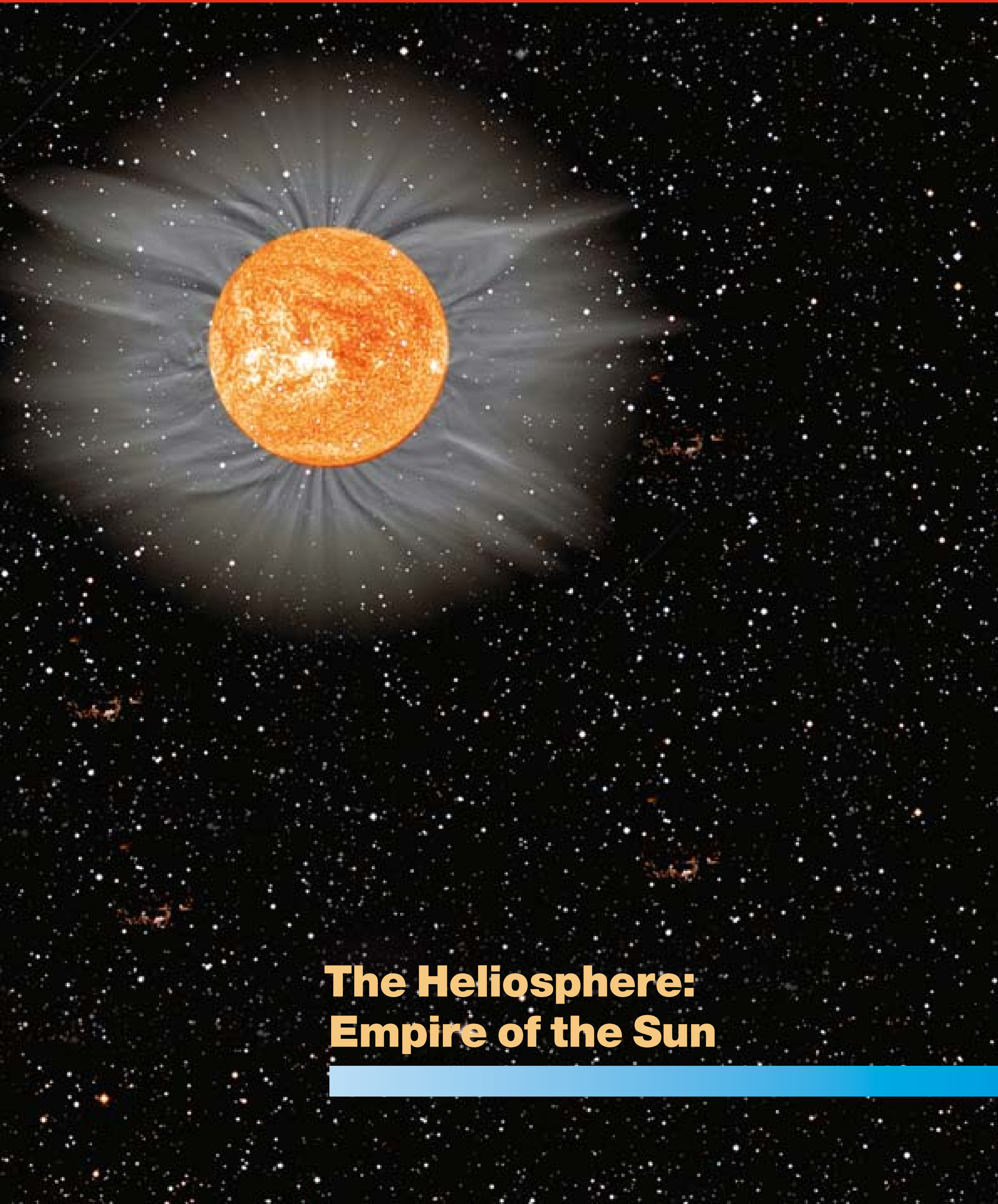
INTERNATIONAL
SPACE
SCIENCE
INSTITUTE

SPATIUM

Published by the Association Pro ISSI

No 17, October 2006



The Heliosphere: Empire of the Sun

Editorial

In a landscape rich in lakes and forests in what was then eastern Pomerania, nowadays a part of Poland, there was a large farm with the wonderful name "Kornburg". It was into that environment that a boy was borne on 4 September 1926. The child grew up on his parent's estate and at that time no one could foresee that this youngster eventually would become one of the leading space scientists of the twentieth century.

He visited the schools in the nearby village and at the age of ten he was sent to the Gymnasium in Stolp (now Slupsk), the town near the Baltic coast. Mathematics and natural sciences were the subjects fascinating him most; so it was a logical choice to study physics at the University of Göttingen, one of the intellectual centres of Germany just recovering from World War II. There, he received the diploma and the Ph.D. in physics, but soon Carmen, a pretty girl just about to finishing her Gymnasium, attracted his interest just as well and the two married a few years later. Many stations marked their common further way to the Universities of Chicago, Miami and Toulouse and to the NASA centres at New York, Maryland, Houston and Pasadena; but during all these years, Bern was their favourite city.

It is with gratitude that Pro ISSI devotes the present issue of Spatium to Johannes Geiss, the boy from Pomerania, at the occasion of his eightieth birthday and to Carmen, the charming wife who accompanied Johannes through all the busy years.

The heliosphere is the subject of the present issue of Spatium, the empire of the Sun, or, to be more specific, that part of the universe where the Sun dominates over the surrounding interstellar medium. We are very thankful to Professor André Balogh, Director at ISSI, for his kind permission to publish herewith his fascinating talk held for the Pro-ISSI audience on 28 March 2006.

It is the Sun, from which Johannes Geiss learned about the composition of the pre-solar cloud that created our central star and the planets 4.6 billion years ago. His solar wind experiment, mounted five times on the Moon by the Apollo astronauts, allowed him and his co-investigators at the University of Bern to get deep insights not only into the solar system but into the secrets of the entire universe.

We are deeply indebted to you, Johannes, for all what you have done for science, for Bern and Switzerland and for the cultural progress of our society towards understanding this wonderful world!

Hansjörg Schlaepfer
Neerach, October 2006

Impressum

SPATIUM
Published by the
Association Pro ISSI
twice a year

INTERNATIONAL
SPACE
SCIENCE
INSTITUTE

Association Pro ISSI
Hallerstrasse 6, CH-3012 Bern
Phone +41 (0)31 631 48 96
see
www.issibern.ch/pro-issi.html
for the whole Spatium series

President
Prof. Heinrich Leutwyler,
University of Bern
Publisher
Dr. Hansjörg Schlaepfer,
legenda schläpfer wort & bild,
CH-8173 Neerach
Layout
Marcel Künzi, marketing · kommunikation, CH-8483 Kollbrunn
Printing
Schellenberg Druck AG
CH-8330 Pfäffikon

Parts of the Sun's outer atmosphere are jettisoned into space in the form of the solar wind. This artist's view shows an image of the Sun by the SOHO extreme ultraviolet imaging telescope together with the corona as seen from the Earth during a total eclipse. (Credit: Hansjörg Schlaepfer)

The Heliosphere: Empire of the Sun[1]

Introduction

The Sun dominates, through its expanding atmosphere in the form of the solar wind, a large volume of space that extends to a distance over twice the orbital distance of Pluto or more than 80 Astronomical Units (AU)[2]. This is the heliosphere, the empire of the Sun. At its outer boundary, the solar wind is stopped by the surrounding interstellar medium. The structure of the heliosphere follows the variations of the solar activity cycle; it is formed by the ongoing interaction between solar wind streams of different speeds and the magnetic field of the Sun that is carried in the solar wind. At times of high solar activity, solar outbursts lead to very disturbed conditions that affect even the immediate neighbourhood of the Earth. Space missions have explored the properties in the different regions of the heliosphere, from the orbit of Mercury to its outer reaches. The study of the heliosphere is immediately relevant to a better understanding of our own space neighbourhood, but it is also a case study of a much larger class of stellar environments.

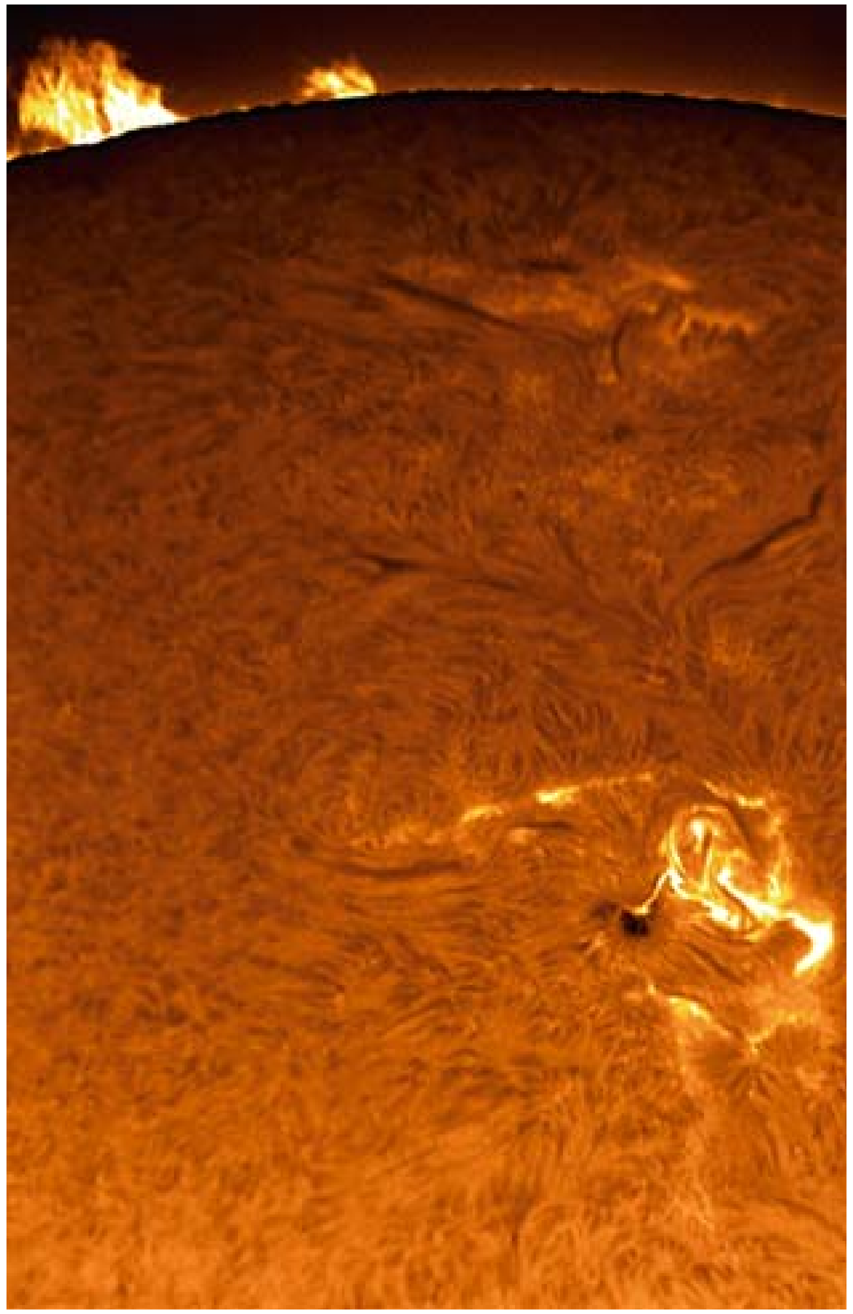


Figure 1: This image shows part of the Sun's surface in the Hα wavelength that is at the emission line of hydrogen. The sunspot has a diameter exceeding the size of the Earth. Relatively cool regions appear dark while hot regions appear bright. On the far left, solar prominences hover far above the Sun's surface. (Credit: Greg Piepol, sungazer.net)

---

[1] This text is based on a lecture by Prof. André Balogh for the Pro ISSI Association on 28 March 2006.

[2] The Astronomical Unit is defined as (approximately) the mean distance of the Earth from the Sun. The currently accepted value is 149,597,870 kilometres.

The Solar Wind and the Existence of the Heliosphere

The large volume of space around the Sun that we call the heliosphere is filled with the expanding atmosphere of the Sun. The upper solar atmosphere, the solar corona, visible from the Earth only at the time of solar eclipses **(see Figure 2)**, continuously expands into space at high speeds, varying between about 300 km/s and 1,000 km/s. This is the solar wind. The corona is a rarefied, very hot gas, with temperatures well in excess of one million degrees. At these temperatures, many of the electrons around the atomic nuclei are stripped away; leaving positively charged ions and negatively charged electrons in the gas. This is plasma.

Why the corona is so hot, when the surface of the Sun, the photosphere, is only about 6,000 degrees, has remained a mystery since the high coronal temperatures were identified in the 1940s. There is enough energy emerging from the Sun in the form of convective motions (mass transport from the interior to the surface) to supply the energy needed to heat the corona, but the way that this convective mechanical energy is transmitted to the gas in the corona is not really understood. An important factor in this is the magnetic flux that is also transported with the material to the solar surface. In the lower corona (the layers closest to the solar surface) magnetic fields emerging from the solar surface form very complex magnetic loops that can be observed by space-based solar telescopes or even from the ground at the times of total solar eclipses. It is likely that some form of waves and some form of magnetic dissipation are the main contributors to the heating of the corona. In any case, the corona is not only heated to these very high temperatures, but also, a part of it is expelled into space at high, supersonic speeds in the form of the solar wind.

The solar wind streams away from the Sun in all directions. Its composition is identical to the Sun's corona, that is approximately 95% protons (hydrogen cores) by particle densities, 4% alpha particles (helium cores) and 1% of carbon, nitrogen, oxygen, neon, magnesium, silicon and iron are the most abundant. These components are present as a plasma. The first detailed composition measurements were performed by Johannes Geiss's Solar Wind Composition (SWC) experiment on the Moon in the frame of the first five US Apollo missions 1969–1972. The solar wind was collected using a specially prepared metal-foil and then brought back for analysis **(see Figure 3)**. Today, the exact composition of the solar wind is routinely measured by instruments on Ulysses and ACE, two spacecraft carrying a Solar Wind Ion Composition Spectrometer.

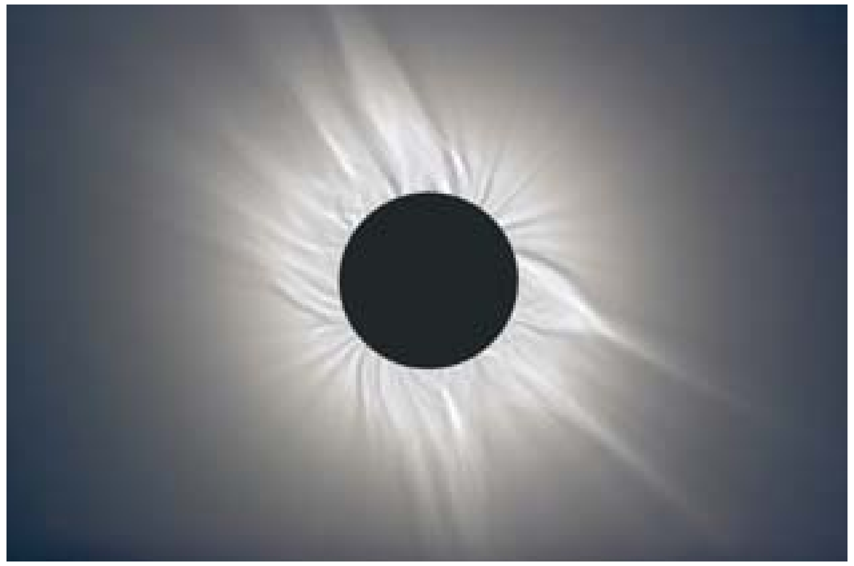


Figure 2: During a total solar eclipse, the Sun's corona is a marvellous sight. The subtle shades and features span a brightness range of over 10,000 to 1, making them difficult to capture in a single picture. But this composite of 33 digital images ranging in exposure time from $^1/_{8000}$ to $^1/_5$ second comes very close to revealing the solar corona in all its details. (Credit: Koen van Gorp)

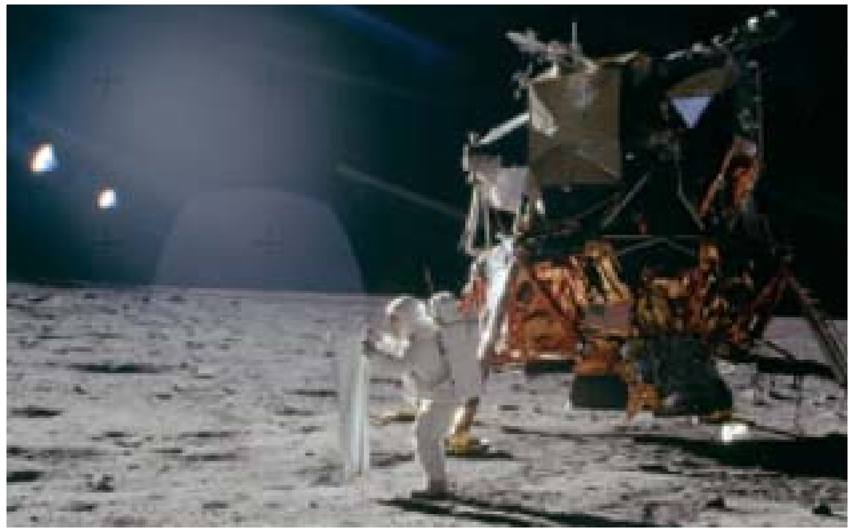


Figure 3: The US astronaut Edwin E. Aldrin unfurling the Solar Wind Composition (SWC) experiment of Johannes Geiss in the Mare Tranquilitatis on the Moon on 20 July 1969. This experiment was the first attempt to quantitatively analyse the composition of the solar wind. (NASA Photo S11-40-5872)

After more than four decades of observations, the properties of the solar wind have been well documented. The solar wind has an average density of 7 particles/cm^3 at the orbit of the Earth (but really highly variable from about 1 to 100 particles/cm^3), with a speed that varies from less than 300 km/s up to about 1000 km/s. The solar wind causes the Sun every second to loose up to 10^9 kg of matter. The density of the solar wind decreases as the inverse square of the distance to the Sun (it is a spherically expanding gas), but the speed of the solar wind varies very little all the way out to the outer boundary of the heliosphere. Both, the solar wind and the heliosphere have been postulated before space missions, since the early 1960s, have brought conclusive evidence about their existence. The two are really related, because if there is a continuously flowing solar wind, it is more than likely that it will be powerful enough to create the space for a "heliosphere" around the Sun, from which the interstellar medium (an even more rarefied gas than the solar corona and the solar wind) is excluded.

Manifestations of the Solar Wind

Perturbations of the Earth's Magnetic Field

Four kinds of observational evidence had been interpreted in terms of a solar wind and the heliosphere. First of all, periodic perturbations in the Earth's magnetic field were noted that matched the periods of the solar rotation (about 27 days) and the solar sunspot cycle (about 11 years), **see Figure 4.** The shorter, 27-day periodicities were associated with the passage, as the Sun rotates, with specific regions on the Sun that appeared to cause a higher level of auroral activity and other manifestations of geomagnetic disturbances. These regions were called M-regions that were thought to be somehow more magnetically active than other parts of the Sun. Their occurrence also changed with the 11-year solar cycle, and such M-regions appeared to be more active (caused

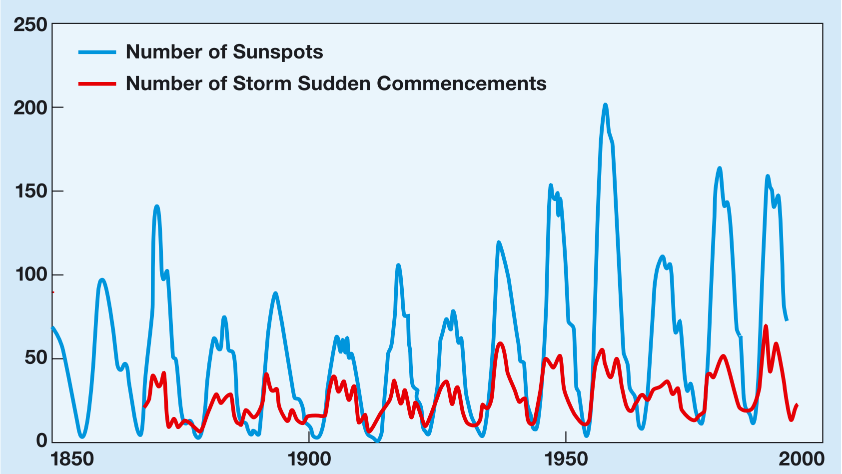


Figure 4: The periodicities observed in geomagnetic phenomena follow solar periodicities: the 11-year solar sunspot cycle and the 27-day solar rotation period. (Credit: Ellis, Maunder, Forbush, Bartels, Chapman)

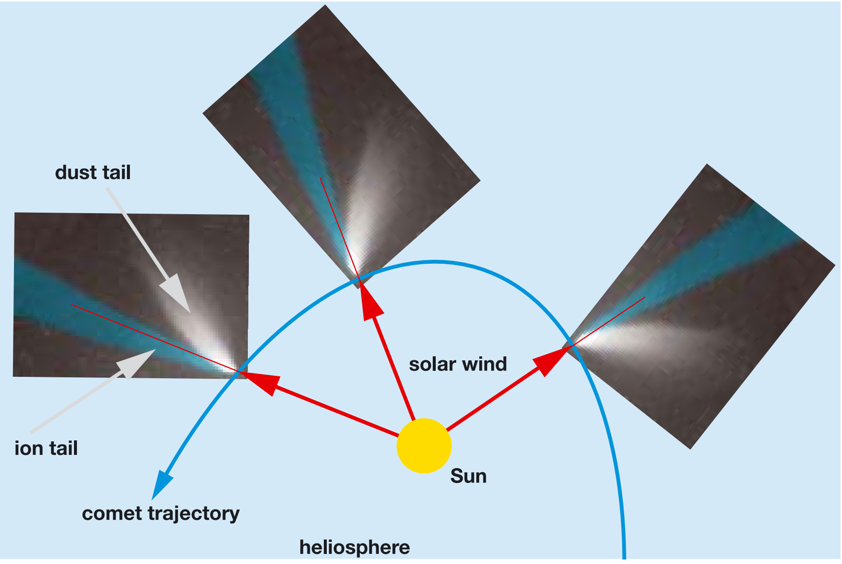


Figure 5: L.Biermann proposed that ion tails arise from atomic particles in the coma that are ionised by solar ultraviolet radiation which are then "entrained" by a "continuously flowing" corpuscular radiation from the Sun (now known to be the solar wind).

more terrestrial disturbances) not at the time of maximum in solar activity, but rather away from it. It is now known that such periodic disturbances are really caused by Corotating Interaction Regions which, as described in more detail below, are large scale structures in the solar wind caused by the collision of fast and slow solar wind streams. As the solar regions with which these structures are associated remain stable over several solar rotations, the interaction regions in the solar wind that cause geomagnetic disturbances also recur at the same time in each solar rotation. This is then observed as a 27-day periodicity in the terrestrial effects that was known well before the space age.

In addition, generally following large solar flares, large geomagnetic storms were observed. These geomagnetic storms are seen in the records as large, fast depressions, followed by a recovery over many hours or even a few days in the magnetic field measured in geomagnetic monitoring stations. This can be explained (and this explanation has proved valid even in the latest state of our knowledge) by the compression of the geomagnetic field by some large scale wave-front that travels from the Sun after large solar flares. These are now known to be Coronal Mass Ejections, described in more detail below. Both the matching periodicities at the rate of the solar rotation and the nature of the geomagnetic disturbances pointed to some agent that brought solar phenomena and disturbances to the vicinity of the Earth. Of course, this is the solar wind and the various structures that it carries in response to events on the Sun.

The Secrets of Cometary Tails

The second line of reasoning that indicated a continuous emission of particles from the Sun was developed in the 1950s. Ludwig Biermann worked out that the bluish tails of comets (now called the plasma tail) that always point radially away from the Sun can only be caused by particles constantly steaming also away from the Sun **(see Figure 5)**. Even though the orientation of comet tails had been known for centuries, for long it was thought that the pressure of radiation (the visible light) from the Sun was responsible. Biermann's contribution was to demonstrate that radiation pressure was not enough and that particles travelling from the Sun at hundreds of km/s were necessary to create the plasma tails of comets. The other tail, the dust tail, curves away from the comet, still in a generally anti-sunward direction, is in fact generated by solar pressure, photons from the Sun striking the micron-sized dust particles emanating from the comet. Both tails are very visible in the photograph of comet Hale-Bopp in **Figure 6.**

Parker's Supersonic Flow Theory

In 1958, a highly controversial idea was put forward by Eugene Parker, then a young researcher in the University of Chicago. He had been influenced by the ideas of Biermann, but he set out to calculate the consequences of the million-degree solar corona on what happens to this very hot part of the Sun's atmosphere. His theoretical solution to the problem included many simplifications, but still provided a sophisticated mathematical model for a solar wind that would escape from the solar atmosphere at supersonic speeds. We need to explain what supersonic speed means in the context of the solar wind. In a plasma, unlike in an ordinary gas like air, three kinds of waves can propagate: the so-called Alfvén wave that is just the wave which propagates along a magnetic field line as along a stretched string, and two kinds of sound waves (longitudinal compressional waves), the slow- and the fast-mode waves. Parker demonstrated that the solar wind, as it escapes the Sun's corona, travels at speeds in excess of the speed of the fast-mode sound or acoustic wave. This makes the solar wind a supersonic flow of plasma in interplanetary space.

Figure 6: This image shows comet Hale-Bopp with its two tails: the dust tail arises from mass loss in the form of small particles shed by the comet, while the ion tail, a fainter feature, arises from ionized atoms and molecules. (Credit: J. C. Casado)

Parker's idea was controversial at the time, as the scientific establishment favoured a different solution for the solar atmosphere, the so-called solar breeze that just evaporated at the outer edges of the corona. Nevertheless, Parker's theory found very soon, by 1962, uncontroversial proof from the first space missions in interplanetary space, in particular from NASA's Mariner 2 probe to Venus, that the solar wind was continuously observed at speeds of a few hundred km/s, close to what was predicted by Parker's theory. The density of the solar wind close to the Earth's orbit was found to be much less, about 7 particles per cubic centimetre, compared to 30 to 50 than in Parker's original theory. This discrepancy was mainly due to the simplifying assumption made by Parker that the temperature of the corona is constant. Since then, much has been learned about the solar wind. We know that it is a more complex phenomenon than treated by Parker, however, many of his conclusions remain valid, and in any case his thoughts have shaped the way we think about the solar wind and the heliosphere.

The Boundaries of the Heliosphere

If the solar wind flows from the Sun all the time in all directions, even if at speeds and densities that vary, its pressure can hold away the interstellar medium from the space around the Sun. Information about the properties of the interstellar medium are difficult to obtain, as no space mission has up to now reached it to provide direct information. We rely on indirect inferences about such parameters as the density, temperature, composition and magnetic field in the interstellar medium. Very sophisticated remote sensing techniques

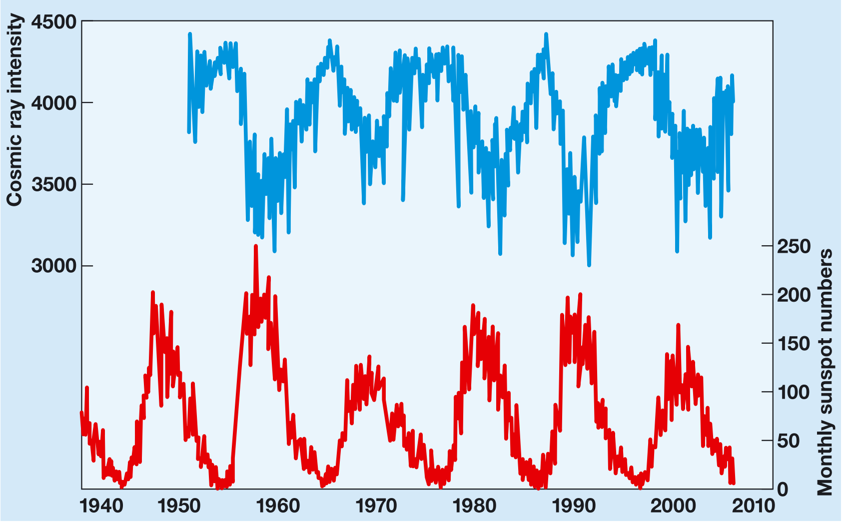


Figure 7: The 11-year variation in the intensity of cosmic radiation at Earth, in anti-phase with the sunspot cycle that implies a large volume of space around the Sun to which the access of galactic cosmic rays is somehow remotely controlled by the Sun.

have shown that even in the distant neighbourhood of the Sun the medium is not uniform but rather lumpy on very large scales. This means that calculating the size of the heliosphere from a balance of pressures between the solar wind and the interstellar medium is not very easy. Further below, the size of the heliospheric cavity will be estimated based on the best currently available estimates of the parameters of the interstellar medium.

Modulating the Cosmic Rays

The fourth and last early inference that a large volume of space was under the control of the Sun was made by Leverett Davis in 1956. His suggestion was based on the observation that the intensity of high-energy cosmic rays at Earth (see Spatium Nr. 11) increased when the number of sunspots was low and decreased when the number of sunspots was high. Sunspot numbers follow an 11-year cycle and are related to how "active" the Sun is. Activity on the Sun is measured in many different ways, but the way they change is closely related to the oldest indicator, the number of sunspots. There are generally more energetic outbursts from the Sun when the sunspot numbers are high, even though the outbursts are not directly related to sunspots. Almost certainly, underlying all measures of solar activity is the magnetic field of the Sun and the way the solar dynamo in the interior of the Sun is changing periodically.

The way cosmic ray intensities and sunspot numbers change in opposite directions is shown in **Figure 7.** How does this lead to the concept of the heliosphere? Cosmic rays are generated in the far reaches of the galaxy, in cataclysmic events such as supernova explosions; cosmic rays exist everywhere in the galaxy. As they reach the volume of space around the Sun, their propagation is impeded by the outward flowing solar wind and the diverse magnetic structures that are carried in the solar wind. If the solar wind and its structures change according the level of activity on the Sun (as they indeed do in response to the 11-year activity cycle), then cosmic rays will be more or less impeded in reaching the Earth. This means that when the Sun is in its more active state, with the largest number of sunspots, fewer cosmic rays can reach the Earth than when solar activity is low. Another way of looking at this is that at high solar activity a larger amount of energy needs to be expended by cosmic rays to reach the Earth, but as there are fewer cosmic rays of such higher energies, the number detected at Earth at such times is lower. Conversely, during low activity levels, lower energy cosmic rays can reach the Earth and as they are more numerous, their intensity increases.

This modulation of the intensity of cosmic rays in synchronism with the solar cycle can be explained by the existence of a large volume of space around the Sun in which the solar wind and its magnetic structures can impede the penetration of the cosmic rays that arrive from all directions in the galaxy. This is the volume of space around the Sun that Leverett Davis was the first to call the heliosphere.

The Size of the Heliosphere

The best indication of the size of the heliosphere was obtained when NASA's Voyager 1 spacecraft, launched in 1977, crossed the Termination Shock, one of its key outer boundaries, in December 2004, **see Figure 8.** The distance of Voyager from the Sun was then 94 Astronomic Units or about 14,100,000,000 km. This was a very important event in the history of heliospheric research, because after four or more decades of theoretical speculation, the measurement of the distance to this outer boundary provided a firm foundation to the calculations and modelling of not just the size, but also of our understanding of the heliosphere. Voyager 1 in fact made a set of important observations that will be described below; some of these are providing new questions because they don't fit the previously accepted theoretical mould.

Early estimates simply assumed that the flow of the solar wind, as it becomes rarefied the further we go from the Sun, will simply drop to a sufficiently low level of dynamic pressure which will balance the low pressure that the interstellar medium exercises to contain the solar wind. This is not only an oversimplified picture, but it is not conform to the important physical principles that describe the collision of a supersonic flow with another medium.

The Properties of the Interstellar Medium

In any case, the parameters of the interstellar medium in the vicinity of our Sun are not very well known. There have been many indirect estimates of the relevant parameters, but interstellar space is not at all uniform and has very large variations in its density, temperature, as well as the relative speed between different regions, **see the table below.** Primarily by examining starlight from different stars located in directions all around the Sun, it has been deduced that our distant neighbourhood in space, beyond the heliosphere, is rather emptier than interstellar space in general. This large and very rarefied volume, but with very high temperatures, is about 100 parsec in dimension (where 1 parsec, a unit normally used to measure astronomical distances, is 206,264.8062 Astronomic Units, or 3.1×10^{13} km, or yet 3.3 light years). But in fact, inside this bubble there are small irregular regions that are considerably cooler and denser than the average of the bubble. Even then, the Sun's immediate neighbourhood is described as a "warm, partially ionized diffuse interstellar cloud", this is the Local Interstellar Cloud or LIC whose properties define, together with those of the solar wind, the size of the heliosphere and the nature of its boundaries. The key parameters of the LIC are the density of neutral hydrogen atoms (0.24/cm^3), the density of electrons (0.09/cm^3), the ratio of ionized hydrogen (or number of electronless protons) to hydrogen

Interstellar Medium (ISM) Phases

Component	*Fractional volume (%)*	*Temperature (K)*	*Density (atoms/cm^3)*	*State*
Cold Neutral Medium (CNM)	1– 5%	50 – 100	1–10^3	neutral hydrogen atoms
Warm Neutral Medium (WNM)	10–20%	1000–6000	$4 \cdot 10^{-4}$	neutral hydrogen atoms
Warm Ionized Medium (WIM)	20–50%	10^3–10^4	10^{-2}	ionized hydrogen
Hot Ionized Medium (HIM)	30–70%	10^6–10^7	10^{-4}–10^{-2}	highly ionized

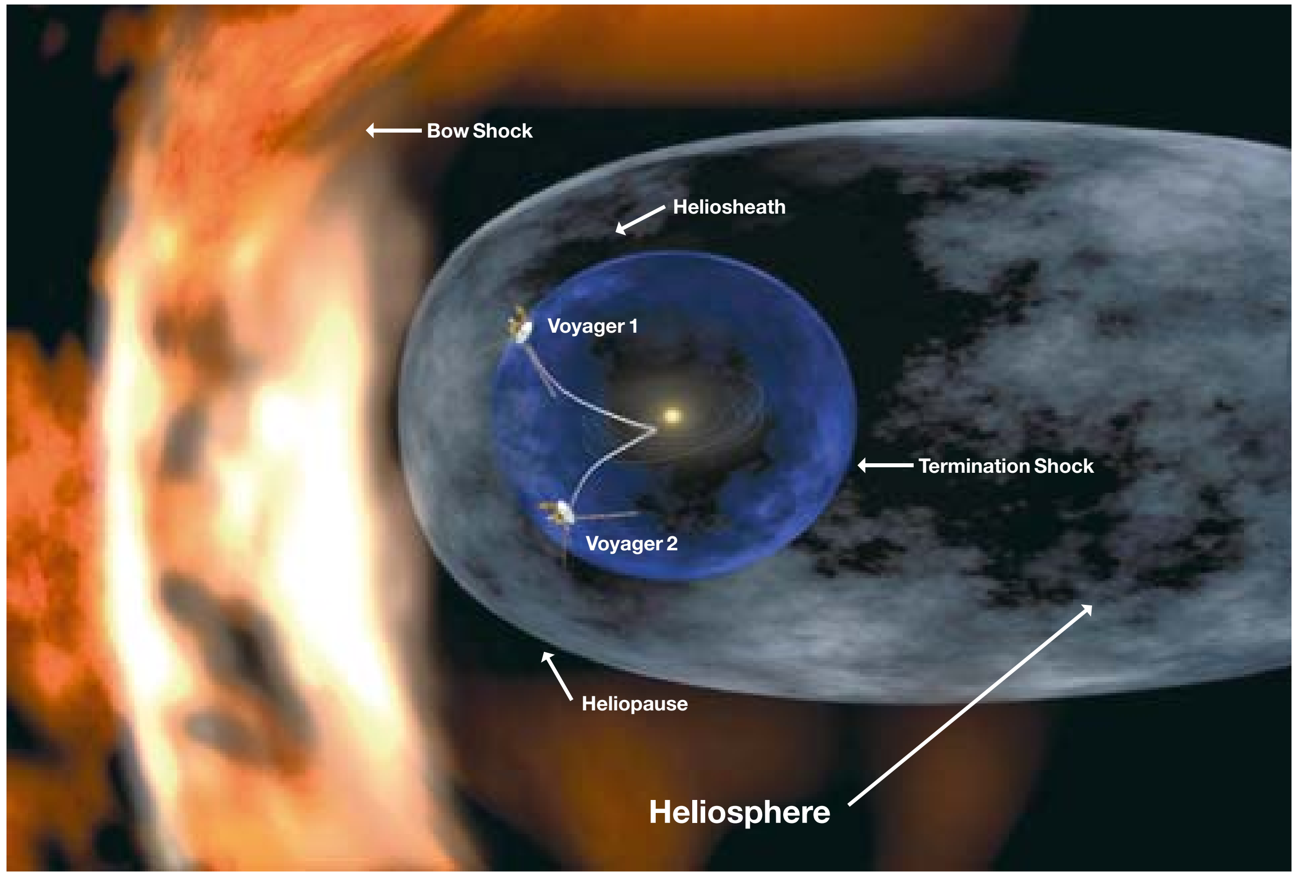


Figure 8: The boundaries of the heliosphere consist of the Termination Shock (where the solar wind slows down to subsonic speeds) and the heliopause (where the solar wind reaches the interstellar medium). In front of the heliopause there is the bow shock, where the interstellar medium is compressed by the heliosphere.

atoms (about 23%), the density of helium atoms (0.014/cm^3), the ratio of ionized to neutral helium (about 45%), the temperature (about 6,400 K, very similar to the temperature of the Sun's surface, the photosphere, but there are many orders of magnitude difference in their respective densities). Just for comparison, it is estimated that the temperature of the large local bubble is about a million degrees, but it consists mostly of very low-density ionized hydrogen, about 0.005 particles/cm^3.

Defining the Boundaries

For estimating the size of the heliosphere, using these parameters, we also need the physical principles of the interaction between the solar wind and the LIC. It has already been said that the solar wind is highly supersonic. As this flow runs into a near stationary medium, it needs to slow to subsonic speeds. In all kinds of flows in nature, the transition from supersonic to subsonic flow can only be a shock wave. There is a very large variety of shock waves, depending on the flows and the media in which the slowdown happens. Frequently quoted examples are supersonic aircraft, explosions, even the cracking of whips! In the astrophysical context, there are shock waves around planetary magnetospheres (where the solar wind slows down to flow around the obstacle) and extremely large shock waves generated by supernova explosions. In the case of the solar wind and the heliosphere, the shock wave is called the Termination Shock (this is the boundary that was crossed by Voyager 1 in late 2004) where the solar wind becomes subsonic, **see Figure 9.**

The Boundaries of the Heliosphere

Beyond the Termination Shock, the medium still consists of the now slowed down solar wind, until we reach the heliopause which is the boundary between the solar wind and the LIC. In space plasma physics, it is shown that plasmas of different origin cannot really mix (except in very special circumstances), so the LIC and solar wind plasmas have a distinct boundary that separates them. But there are complications. The LIC (unlike the solar wind) is not a fully ionized plasma, it contains neutral hydrogen and helium which are not subject to the laws of plasma physics. Many of these neutral atoms from the LIC penetrate into the heliosphere, from its distant boundaries all the way to the Earth's orbit where they have been observed, providing important information on the LIC. However, many neutral atoms are ionized when they encounter the solar wind (mostly by a process called charge exchange with the solar wind protons). Once ionized, these particles are "picked up" by the solar wind and can, if they are numerous, influence the local properties of the solar wind. Furthermore, such pick-up ions can be energized at the Termination Shock and, at high energies, become part of the cosmic ray population; as they are generally quite recognizable as a different population from the more generally observed galactic cosmic rays, they are called the "anomalous"

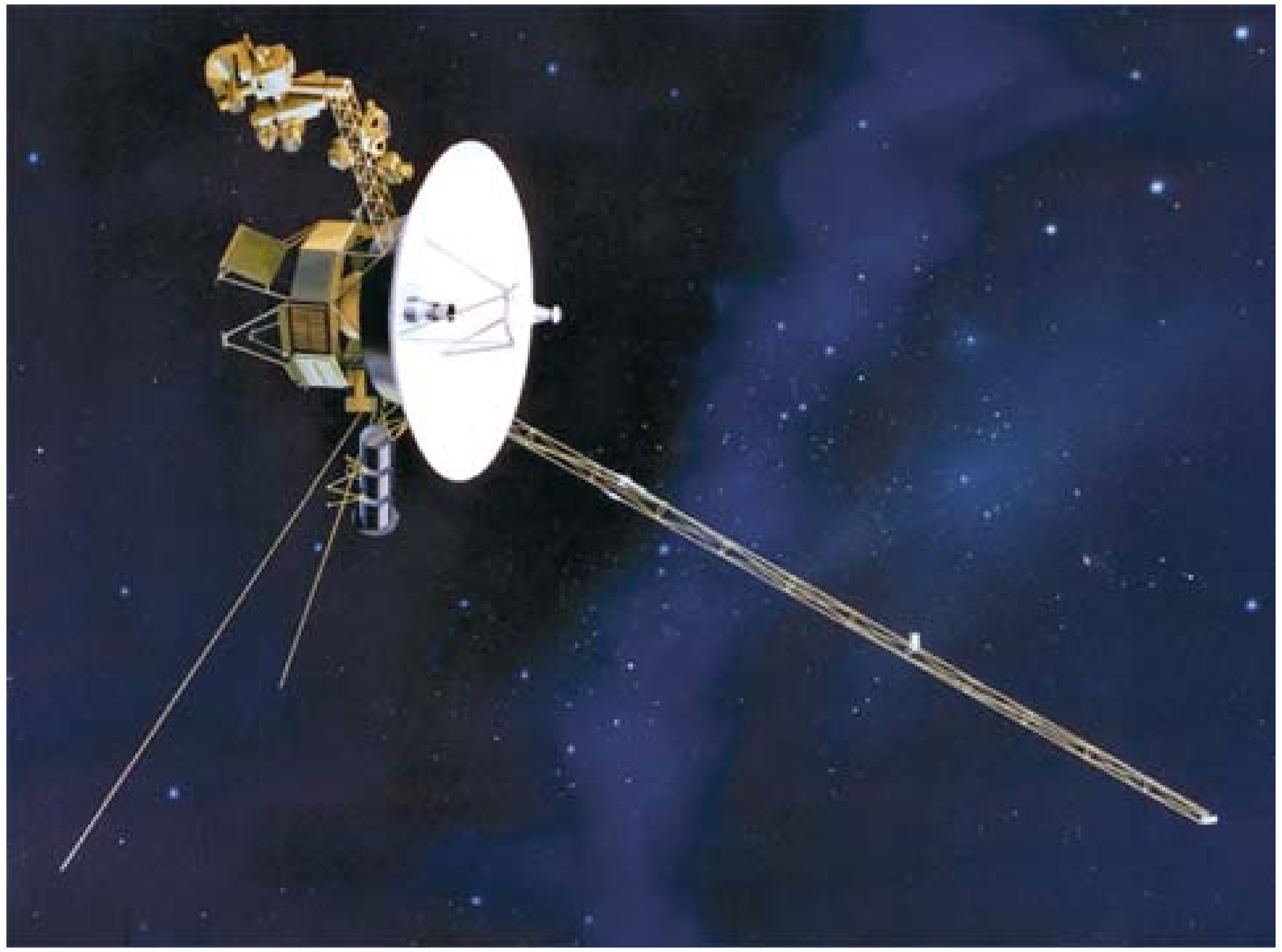


Figure 9: NASA's Voyager 1 spacecraft crossed the historical 100 AU milestone (equalling 15 billion kilometres) on 17 August 2006. (Credit: NASA / JPL)

cosmic rays. Just where these anomalous cosmic rays are energized has become one of the puzzles of the Voyager 1 observations after it crossed the Termination Shock, as this process does not appear to happen where expected from theory.

There have been many estimates in the past of the distance to the Termination Shock, all with many very tentative parameters. The very early estimate of Leverett Davis who could not know about the solar wind at the time (in 1955), but just made based on the modulation of cosmic rays, was 100–200 AU. Since then models with ever increased sophistication, but still with a limited certainty of the parameters, have suggested a distance of about 80 to 120 AU. The Voyager 1 observation at 94 AU falls well within the estimated range, thus justifying a posteriori many of the assumptions and parameters that have been used in the models.

Of course, the distance to the Termination Shock cannot be a constant, as the properties of the solar wind vary significantly, in particular with the 11-year solar activity cycle. As a result, the Termination Shock, as all other shock waves in space that we know, is in constant motion, moving in and out at speeds probably about 100 km/s. When Voyager 1 observed it, it was moving inward with about that velocity. The amplitude of the motion of the Termination Shock probably varies, depending on timescales: it makes large excursions (maybe as much as 10 AU or more) in response to solar cycle variations, but smaller ones in response to the always changing conditions in both the solar wind and, presumably, in the LIC as well.

There are question marks concerning the existence of an outer shock wave, outside the heliopause and surrounding the whole heliosphere. This depends partly on the relative velocities of the Sun and the LIC, measured to be about 25 km/s, but partly also on the other parameters such as the density and temperature of the two colliding media. At this stage, scientific opinion is divided on this, while awaiting better observations of the relevant parameters.

By studying the relative motion of the Sun and the heliosphere in the local interstellar medium, it is estimated that occasionally, many times during the lifetime of the solar system, conditions in the LIC change significantly as large, cool and dense molecular clouds are encountered (with presumed densities of up to 10 particles/cm^3, or maybe 30 to 40 times greater than the present LIC). When that happens, the heliosphere is significantly compressed to perhaps less than a half or a third of its current size. How such conditions affect the Earth and its immediate environment, has not been really investigated, but clearly there will be noticeable consequences. A smaller heliosphere would lead to a weakened capacity to shield cosmic rays and therefore to a higher level of irradiation of the Earth by cosmic rays, which in turn would have important consequences on the Earth's climate and the evolution of life on the Earth. However, such changes are unlikely to happen on timescales measured by human generations; rather it is assumed that the heliosphere has been immersed in this cloud for the last perhaps 10,000 to 100,000 years.

It is interesting to compare the size of the heliosphere to the size of the solar system occupied by the planets. The Earth is well in the inner heliosphere, but really so are Jupiter (at about 5.5 AU) and Saturn (at about 10 AU). Even the gas giants Uranus and Neptune are well within the inner half of the heliosphere, while Pluto (lately defined a dwarf planet) is, when furthest away from the Sun, just about half-way out towards the Termination Shock. It can be safely said that the heliosphere, the empire of the Sun, is, on our human, solar-system-bound scales truly enormous. But then, on astrophysical scales, whether in the local bubble or in our galaxy, it is dwarfed by the vastness of the universe, **see Figure 10.**

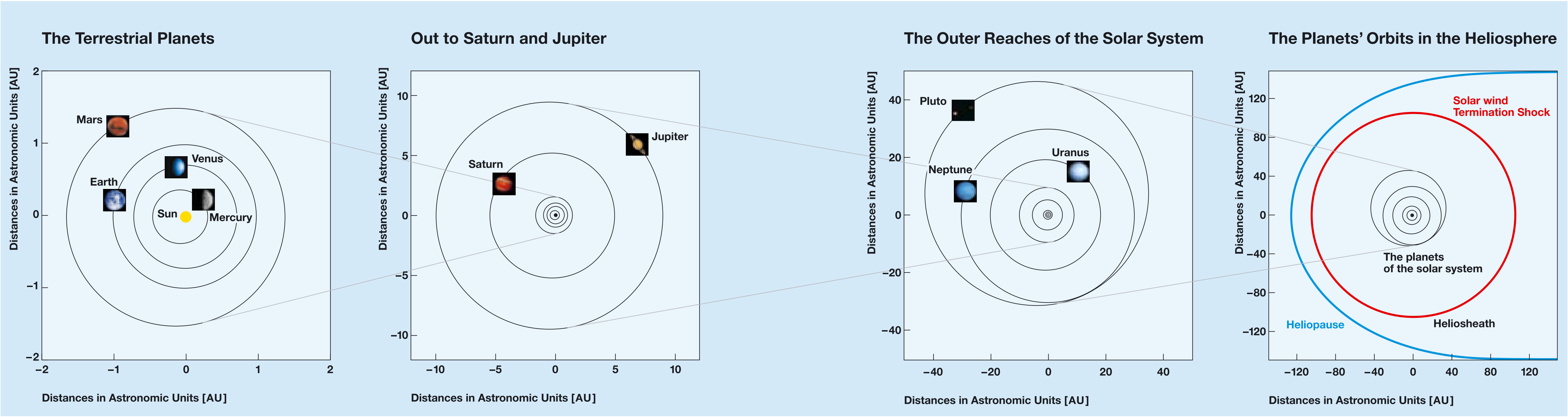


Figure 10: This schematical image shows on successively increasing scales the inner solar system with the terrestrial planets, the gas giants Jupiter and Saturn, the outer solar system with Uranus, Neptune and Pluto, and finally the solar system in the heliosphere.

The Dynamic Structures of the Heliosphere

Having visited the outer reaches of the heliosphere and visited its boundaries, we now turn to the inner half of the heliosphere; perhaps even just the innermost tenth of it, out to about the orbit of Saturn. This is the part of the heliosphere that we know best and one that is most relevant to Earth-dwellers. The structure of the heliosphere in this inner region is wholly dependent on the Sun and the solar wind. There is a whole range of dynamic phenomena, mostly dependent on the 11-year solar activity cycle that characterizes the inner heliosphere.

Probing the Heliosphere

Thanks to a large number of interplanetary spacecraft, since Mariner 2 in 1962, the solar wind is a well known and well understood medium. Questions of course remain, but the key processes and phenomena, their dependences as a function of position in the heliosphere and of time have been well observed and described. Among the key space missions dedicated to heliospheric studies, special mention must be made of the two USA-German Helios probes in the mid 1970s, the Pioneer 10 and 11 probes, and the two Voyagers, 1 and 2. Helios remains to date the first spacecraft ever to have observed the solar wind in the innermost heliosphere, well inside the Earth's orbit. The Pioneers and Voyagers have explored the middle and outer heliosphere. Several spacecraft have observed the solar wind at the orbit of the Earth, from the early IMP series through ISEE-3 and now Wind and ACE. The Ulysses mission launched in 1990 and scheduled to operate at least until mid-2008 has been unique in its orbit over the poles of the Sun to provide observations of the three-dimensional heliosphere. These missions and the many scientific teams that have worked with them now over two generations have brought the knowledge of the heliosphere to its current advanced level.

The Solar Wind through the Solar Activity Cycle

The principal changes in the solar wind, as a consequence of changing solar activity levels, take place in three dimensions, mainly out of the ecliptic plane which contains the Earth's orbit and is within 7.25° of the solar equatorial plane. Ground- and space-based solar observations have shown that there are many important changes in the Sun and its corona that accompany the sunspot cycle. One of the main discoveries concerning the solar wind in the 1970s was that there are two kinds of solar wind, fast streams (at speeds in excess of about 650 km/s) and slow solar wind (with speeds less than about 550 km/s), **see Figure 11.** Most of the parameters that describe the solar wind, such as its density, temperature and composition, are significantly different in the two kinds of solar wind, not just its speed. Also in the 1970s, solar observers identified dark, relatively cool regions in the solar corona as the origin of the fast solar wind streams, these regions are the coronal holes. One of the characteristics of coronal holes is that the solar magnetic field that they contain is open to interplanetary space, so that the magnetic field lines dragged into the heliosphere by the solar wind originate mostly from coronal holes.

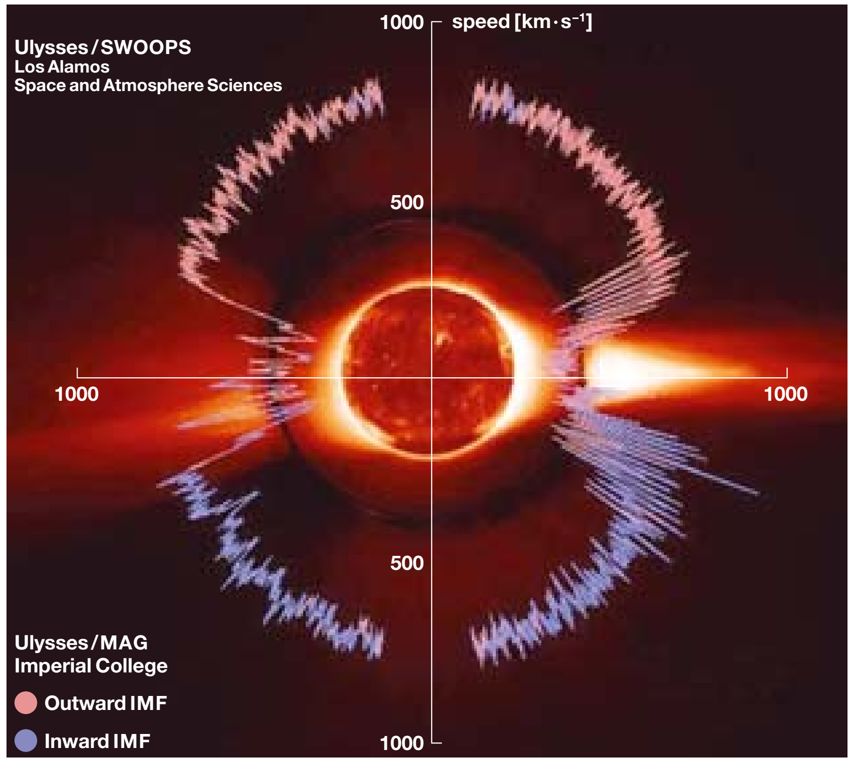


Figure 11: The solar wind profile as a function of solar latitude probed by the Ulysses spacecraft around solar minimum activity. The low speed portion of the solar wind is found mainly in the Sun's equatorial plain, whereas the high speed portion covers the higher solar latitudes. (Credit: Ulysses SWOOPS team, PI: D.J. McComas)

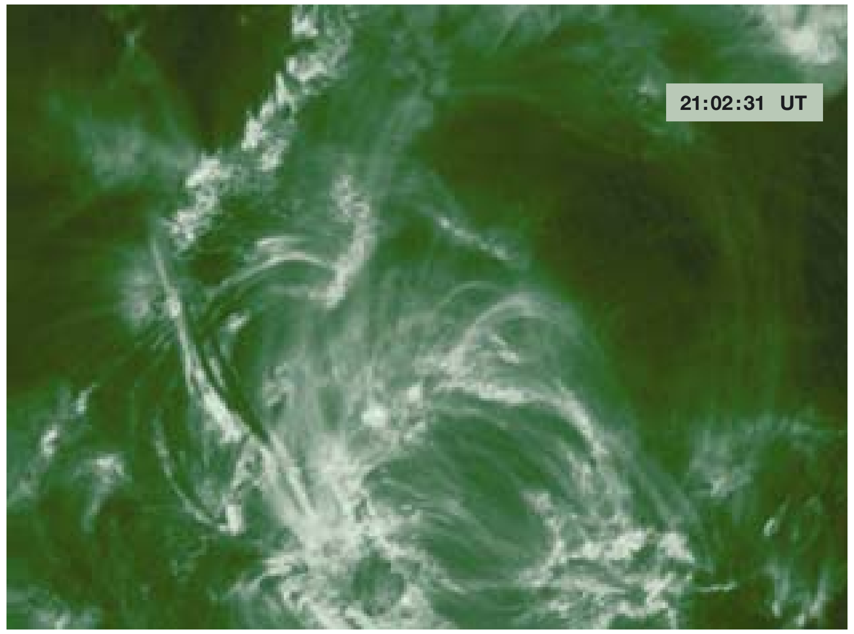


Figure 12: NASA's Transition Region and Coronal Explorer (TRACE) spacecraft discovered new features of the Sun's surface, termed "solar moss". The solar moss consists of hot gas at about two million degrees centigrade which emits extreme ultraviolet light. It occurs in large patches, about 6,000–12,000 miles in extent, and appears between 1,000–1,500 miles above the Sun's visible surface, sometimes reaching more than 3,000 miles high.

The origin of the slow solar wind is less well understood, it is far more variable in its parameters than the fast solar wind, and it is likely to be generated near or at the edges of hot regions in the solar corona where the magnetic field is tightly held in complex loop systems, **see Figure 12.**

The magnetic field of the Sun that threads the solar corona is carried into space by the solar wind from the open-field regions, the coronal holes. The Sun, unlike the Earth, for instance, does not have simple magnetic North and South Poles. Around the minimum activity phase, there are large coronal holes covering the heliographic poles of the Sun, and the magnetic fields that emanate from these are of opposite polarities in the North and the South. This is the closest the Sun ever gets to showing a magnetic dipole such as the magnetized planets. As solar activity increases, these polar coronal holes shrink and fragment, so that there is apparently much less open magnetic flux and the polarities are much more mixed in the corona.

As the magnetic field is carried out in the solar wind, the magnetic polarities are separated by a so-called neutral line, separating the inward and outward pointing magnetic fields. In interplanetary space, the surface that separates the polarities is called the Heliospheric Current Sheet (HCS). Near solar minimum, this vast surface is close to the equatorial plane of the Sun, while near solar maximum, it becomes very complex and highly inclined with respect to the solar equator. Pictures of the solar corona, taken at the time of total eclipses **(see Figure 2)** or rather more routinely from a space-based observatory such as SOHO **(Figure 13),**

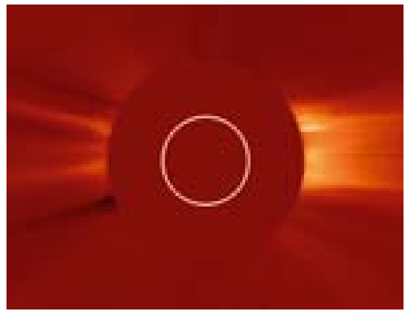


Figure 13: During solar minimum, the solar corona consists of bright streamers along the Sun's equator. Image taken by SOHO's EIT instrument. (Credit ESA)

show long, bright streamers that are confined near the solar equator near solar minimum, but point at high heliographic latitudes when the sunspot number is high **(Figure 14).**

The two kinds of solar wind are well delineated around solar minimum: all the fast solar wind comes from the large polar coronal holes and the slow wind from above the equatorial regions where the magnetic field remains in the form of loop systems. The HCS is always embedded in slow solar wind, just as the streamers are seen to originate above the hot loop system in the equatorial corona. A further effect that shapes the heliospheric medium is what is called the over-expansion of the fast wind: even though coronal holes above the poles have an angular extent of only about 30 degrees away from the poles of the Sun, the fast wind fills the inner heliosphere to much lower heliolatitudes, it is as though at least at the edges of coronal holes, the solar wind flows bend in a direction away from the solar poles. Around solar minimum, the polar coronal holes remain approximately the same for many periods of solar rotation.

At solar maximum, coronal holes are small and can be found everywhere on the Sun, not just near the poles. They are also generally short lived, often appearing and disappearing within a single solar rotation that is within a few days. As a result, the solar wind is rarely fast, but generally mixed, some not-so-fast wind and slow wind streams mingle at all heliolatitudes. The contrast between the solar wind at solar minimum and solar maximum is well illustrated by the observations of Ulysses.

Closer to Solar Minimum Activity: Corotating Interaction Regions

The formation of polar coronal holes, following solar maximum activity, is not a simple process, but involves the migration of magnetic regions of opposite polarities towards the heliographic poles. While these large coronal holes form, parts of them reach to low latitudes, towards the equator. This means that, at equatorial latitudes, both fast and slow solar wind streams are generated in successive solar longitude ranges. As the Sun rotates, fast and slow solar wind streams are emitted alternately in a given radial direction. But then fast wind catches up with the slow stream ahead of it. In the interaction between fast and slow wind, the solar wind plasma is compressed and heated. As the compressed plasma travels out away from the Sun, it even forms travelling shock waves at its leading and trailing edges. If, as happens close to solar minimum, the flow pattern of fast and slow streams remains the same over successive solar rotation periods (each solar rotation is ~ 26 days long), the outward travelling interaction regions appear to rotate with the Sun, hence their name "Corotating Interaction Regions" or CIRs. Such CIRs persist for a year or more, usually in the period between solar maximum and minimum activity and constitute the major structuring process in the inner heliosphere. In addition, the shock waves that form at their leading and trailing edges constitute an important source of energetic particles in the heliosphere, not competing with cosmic rays, but still providing an important example of how energetic particles can be produced in the universe.

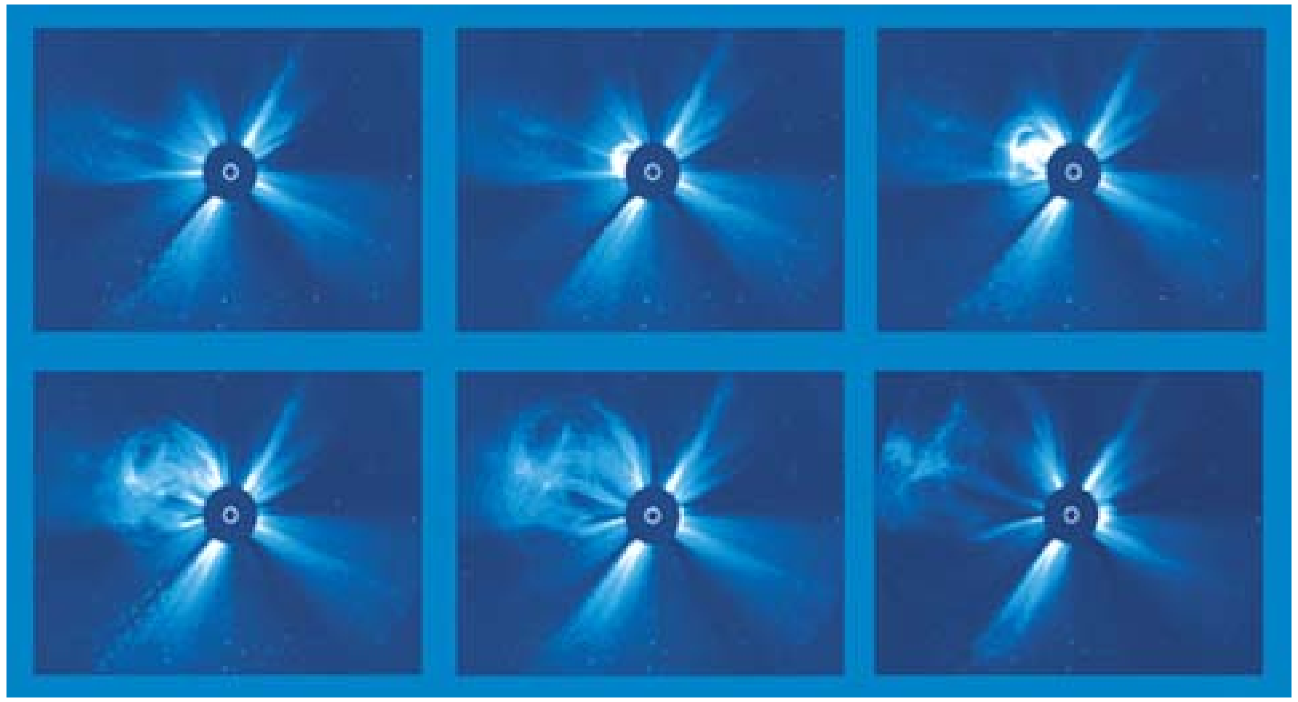


Figure 14: During solar maximum, the coronal streamers tend to point to higher elevations. In addition, large amounts of material are often ejected into space in the form of Coronal Mass Ejections. (Credit: ESA)

Around Solar Maximum Activity: Coronal Mass Ejections

Solar maximum activity has been historically measured by the number of sunspots. This remains a useful criterion, although we now know that sunspots themselves are only a symptom of considerable changes in the Sun's upper layer, the convection zone. Solar activity in fact increases in response to the complexity of magnetic fields emerging to the surface of the Sun; the increased complexity restructures the magnetic fields in the solar corona. In the course of this restructuring process, very large and occasionally explosive amounts of magnetic energy are transformed into sudden heating of the corona that expel very large amounts of coronal material into space, embedded in the solar wind. These events are Coronal Mass Ejections (CMEs); their number increases from close to zero at solar minimum to several per day close to solar maximum.

In many cases, CMEs appear to be closed magnetic structures, unlike the ordinary solar wind that has open magnetic fields embedded in it. CMEs, when directed towards the Earth, often carry enough hot plasma and strong magnetic fields to cause major disturbances in the terrestrial magnetic field; these events are the magnetic storms. The strongest of these storms, usually before or after solar activity maximum, depress the Earth's magnetic field and cause an increase in the intensity of radiation near the Earth in such a way that can damage spacecraft and even affect large sections of terrestrial power supply grids. The very large-scale CME "bubbles" occupy large volumes in the heliosphere around solar maximum activity (when the most powerful CMEs occur) and modify significantly the structure of the heliosphere. As the CMEs propagate out towards the outer heliosphere, they often amalgamate and the largest among them form what has been called a Global Merged Interaction Region (GMIR) which acts as a barrier between the inner and outer heliospheric regions and significantly impedes the access of cosmic rays to the vicinity of the Earth.

One of the effects of large-scale, probably merged CMEs is the effect they have on the outer boundaries of the heliosphere. Some intriguing radio noise observations by the Voyager spacecraft in 1983 and 1992/93 have been interpreted as the signs of radio emission from the heliopause when the large-scale CMEs reach it and disturb it. This provides yet another tentative measure of the distance of the heliopause: estimating the travel time of the CMEs that cause the radio emissions gives a distance of about 150 AU to the heliopause, a figure that is broadly in line with theoretical expectations.

Large-Scale Structures and the Modulation of Cosmic Rays

The cause of cosmic ray modulation is the increase in complexity of heliospheric structures from solar minimum to solar maximum activity. However, the precise processes that control the access of cosmic rays into the inner heliosphere are not fully understood. Almost certainly, several factors jointly play a role in this process. One such process is the formation of GMIRs; it has been noted that cosmic ray intensity decreases significantly, almost in a step-like way, after the passage of a GMIR. But the modulation process as measured at the Earth is smoother and almost certainly involves higher levels of turbulence (or simply disorder) in the heliospheric magnetic field around maximum activity. The access of cosmic rays is controlled by a combination of GMIRs, turbulence and the heliospheric boundary; the modulation of cosmic rays is caused by changes in these factors associated with the activity level through the solar cycle.

Models of the Heliosphere and its Boundaries

There are many models of the heliosphere that attempt to take into account the known characteristics of the solar wind as it propagates to large distances from the Sun and the characteristics of the Local Interstellar Cloud that are deduced from remote and indirect observations. The models tend to agree on the basic parameters, such as the approximate distance of the outer boundaries, the Termination Shock and the heliopause. However, the problem of modelling in detail is quite difficult, as there are several components of the solar wind and the interstellar medium to be brought into the model. Current modelling work is greatly helped by the increase in computer power that enables multiple parametric studies to be carried out. The nature of the boundaries and their effects remain uncertain, as was discovered by Voyager 1 recently, as described below.

One aspect seems generally accepted: this the Hydrogen Wall, a region in front of the heliopause in which there is a significant increase in the density of interstellar neutral hydrogen atoms. First predicted by models, the existence of such a wall has found strong support in the observation of an increase in the absorption of radiation, selectively in the hydrogen spectrum, from nearby stars, such as Alpha Centauri and Sirius. This observation could also be used to try and detect stellar winds similar to that of our Sun around other stars.

The Termination Shock and Beyond: Voyager 1 Results

Most models of the heliosphere estimated the distance to the first boundary, the Termination Shock, to be around 100 AU or perhaps somewhat less, dependent on the activity cycle of the Sun. The two Voyager spacecraft have been approaching such distances in the past few years, so the expectation has been strong that at least Voyager 1 (the furthest away) would meet the Termination Shock sooner rather than later. This significant event occurred in December 2004, at a distance of 94 AU from the Sun, when the instruments on board Voyager 1 unanimously indicated the crossing of a shock front, clearly the Termination Shock wave where the solar wind becomes subsonic. At first sight, the observations seemed to match the predictions, but as Voyager 1 moved further out into the heliosheath, the expected increase in the so-called Anomalous Cosmic Ray (ACR) population was not observed. The ACRs are interstellar atoms that have been ionized and raised to high energies, somewhere at or near the heliospheric boundaries; their existence has long been considered as a strong indication of the strength of the Termination Shock.

Previously, ACRs had been expected to be energized at the Termination Shock, in fact close to the part of the shock front which directly faces the flow of interstellar matter. A recent suggestion to explain these observations is that the geometry of the Termination Shock is not suitable for the energization process at the expected location, but that ACRs originate around the flanks of the heliosphere, where the geometry of the Termination Shock enables the energization to be carried out.

This result has shown the importance of observations to verify the theoretical models. Voyager's great achievement is that it has lasted the long voyage to the edge of the heliosphere: the 27 years it took from Earth to travel to this first boundary illustrates the scale of the heliosphere when measured against what our space missions can achieve.

The Future of Heliospheric Research

Interplanetary spacecraft have explored the solar wind and the heliosphere since the early 1960s. Of these many were in Earth orbit and some just ahead of the Earth in the direction of the Sun near a Lagrange point in space, a point where the Earth's and the Sun's gravity balance. These spacecraft have measured in great detail the properties of the solar wind and monitored the changes in these properties on all time scales. As a result, a very large data base is now available on the solar wind.

But the properties of the solar wind change as a function of distance from the Sun and as a function of solar latitude in ways that are usually not easily predictable. Space missions that have travelled towards the Sun or away from the Sun have contributed to our understanding of how the solar wind and the heliosphere changes as a function of distance away from the Sun. The still unique Ulysses mission, a joint undertaking between ESA and NASA, has been in its polar orbit around the Sun since early 1992, opening a completely new, and in many ways unexpected perspective to observe changes in the structure of the heliosphere with solar latitude.

Continued monitoring of the near-Earth solar wind as well as remote solar observations will extend the data base for heliospheric research. A few space missions are dedicated to such research. NASA's Stereo mission, to be launched in autumn 2006, consists of two identical spacecraft that will accompany the Earth around the Sun. They will follow the same orbit as the Earth, but one ahead, the other behind along the orbit. These two spacecraft will make observations of CMEs from two observation points, to build up a stereoscopic image of these events as they propagate away from the Sun, towards the Earth. This will allow three-dimensional models of the shapes and structures of CMEs to be verified and new models to be developed. Another mission that is under construction by NASA is the Interstellar Boundary Explorer (or IBEX) that will use a novel technique to study the heliosphere's Termination Shock and the heliosheath. In fact, IBEX will remain in Earth orbit, but will monitor energetic neutral atoms that are generated at the heliospheric boundaries and can travel unimpeded through the solar wind. By building up images of the intensities of such neutral atoms, the dynamics of the heliospheric boundary can be better understood. A more ambitious mission, ESA's Solar Orbiter, is also being planned that will combine a relatively close approach to the Sun (maybe a third of the distance between the Sun and the Earth) and moving out of the Sun's equatorial plane, following the example of Ulysses, but not aiming to reach polar latitudes. The objective of the Solar Orbiter, currently planned for a launch in 2015, is to combine remote observations of the Sun and the solar corona with in situ observations of the solar wind to link in more detail than before the corona with the heliospheric medium.

Possibly the most ambitious space mission in the planning is NASA's Interstellar Probe. No launch date has been set and there are some key technical issues to be resolved. The Interstellar Probe is attempting to shorten the time needed to reach the boundaries of the heliosphere and go beyond them, into interstellar space. In order to achieve this aim, a large solar sail, using the radiation pressure of the Sun's photons will be used to accelerate the spacecraft away from the Sun. The speed expected to be achieved is about 70 km/s, or four times faster than was achieved by the Voyager craft. This speed would allow the Interstellar Probe to reach a distance of 200 AU in about 15 years. Voyager reached the first boundary, the Termination Shock at nearly 100 AU, almost 30 years after launch, clearly the Interstellar Probe would bring a better, quicker answer to the questions concerning the boundaries of the heliosphere and just what lies beyond.

The Author

André Balogh is one of the many Hungarians who had to leave their country in the wake of the Hungarian revolution against the Soviet occupation in 1956. He first completed his secondary education in France and then went to Paris to study telecom engineering at the École Nationale Supérieure des Télécommunications. In 1964 he moved to England where he found an employment at the Imperial College in London as a Research Fellow of the then European Space Research Organization ESRO. There he first specialized in research of the Earth's magnetosphere and the nearby interplanetary medium. Capitalizing on his knowledge as a telecom engineer he soon became involved in the development of scientific instruments for space research, which led to his appointment as Principal Investigator for the magnetometers onboard Ulysses (launched 1990 and still working perfectly) and on the four Cluster spacecraft (launched in 2000). In parallel he became a professor for Space Physics in the Physics Department of Imperial College, an appointment which he held until his retirement in 2005 when he was appointed Emeritus Professor of Space Physics. André Balogh knew ISSI from his former visits in Bern participating in Workshops, International Teams and as a visitor. When he was offered the task of a part-time Director of ISSI he gladly accepted as this new appointment allows him to actively shape the space research programmes in his favourite field but also beyond.

Space has attracted the interest of the very young André already. Later, after his studies of engineering sciences, he decided to realize his youth's dreams and to become a space scientist. In the vastness of the universe, the heliosphere is of primary interest to André Balogh as this is the area which is accessible to space missions and can be explored by in situ observations. The complexity of physical processes involved in the creation of the heliosphere is another element challenging space scientists like A. Balogh. And last but not least, it is the sheer dimension of the heliosphere which makes this part of our environment so amazing: yet it is our extended home, still imaginable on the human scale.