ISSI Scientific Report SR-003

The Solar System and Beyond

Ten Years of ISSI

Johannes Geiss & Bengt Hultqvist (Eds.)

INTERNATIONAL SPACE SCIENCE INSTITUTE

SR-003
June 2005

The Solar System and Beyond
Ten Years of ISSI

Editors

Johannes Geiss and Bengt Hultqvist

Cover: Earth rising above the lunar horizon as seen by the Apollo 8 crew - Frank Borman, James Lovell and William Anders - when orbiting the Moon in December 1968 (Photo: NASA/Apollo 8 crew)

The International Space Science Institute is organized as a foundation under Swiss law. It is funded through recurrent contributions from the European Space Agency, the Swiss Confederation, the Swiss National Science Foundation, and the University of Bern.

Published for: The International Space Science Institute
Hallerstrasse 6, CH-3012 Bern, Switzerland

by: ESA Publications Division
ESTEC, PO Box 299, 2200 AG Noordwijk, The Netherlands

Publication Manager: Bruce Battrick
Layout: Jules Perel

ISBN: 1608-280X
Price: 30 Euros

Contents

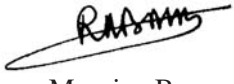
Foreword

When it was created ten years ago, through the initiative of Professor Johannes Geiss, the International Space Science Institute (ISSI), like every new venture, faced the risk of a short-lived existence. Instead, the initial impetus provided by both Johannes Geiss and Bengt Hultqvist set the Institute on a winning course. In the hands of these two talented and famous European pioneers of space science, an ambitious programme was established for ISSI, which gradually linked the Institute's original series of Scientific Workshops with a growing number of International Teams of scientists, thereby involving the science community more proactively.

Scientists from all around the globe have been drawn to ISSI to take advantage of the unique opportunities and facilities that it offers for conducting first-class science by making use of space as well as ground and laboratory data. The result has been an exceptional output in terms of scientific achievements and discoveries that are a testimony to the uniqueness of the Institute. The recognition of its role and its usefulness continues to grow day by day, to the point where ISSI is now capable of attracting the World's best scientists.

After the first decade and with a new team at the helm, we thought it would be an appropriate moment to review the Institute's achievements so far through the summaries and reflections of the main players who have so actively participated in its early life. We also thought that such an initiative could be beneficial not only for the wide scientific community involved in ISSI's activities, but also for those perhaps less-specialised people who already know something about the Institute and its work but would like to learn more. That is how the idea of this book was born.

The fact that ISSI's ten-year anniversary happens to coincide with the celebration of the centenary of Einstein's Theory of Relativity offers a unique chance to showcase the Institute's role. There was nobody better placed to take responsibility for this book's production than the two pioneering founders of ISSI. I would like to express here my great admiration for their dedication and their work, and my strong belief that their invaluable work will continue to serve the cause of international science for many years to come.



Roger-Maurice Bonnet
Executive Director of ISSI

Part A

Introduction

The "space age" began nearly half a century ago. One of the highlights of this epoch was the first view by humans of the Earth rising above the lunar horizon, as shown on the cover of this volume. These early pictures of the Earth seen from a distance have created a new public perception of the World around us. The deserted lunar landscape and the dark space looming behind the Earth contrast starkly with the blue planet – it appears as a beautiful and lonely island in an otherwise hostile cosmic environment.

The change in the public's perception of the World has been matched by advances in our scientific understanding of the Universe, of the origin and evolution of the Solar System with the Sun, the planets, moons and comets, and last but not least of the Earth, including its environment and near-Earth space. All of this has been brought about during the first fifty years of the space age. For the last decade, the International Space Science Institute (ISSI) has participated actively in this development. The present volume provides an overview of the results of the many ISSI Workshops that have taken place, and it is hoped that the reader will thereby gain an insight into the role that the Institute has been playing in fostering these scientific advances.

The International Space Science Institute was established just ten years ago, and its first Workshop was held in October 1995. The various contributions were published as Volume 1 of the *Space Science Series of ISSI* books in the fall of 1996. By the end of 2005, a total of 21 volumes in this series, and four in the *ISSI Scientific Reports* series, will have been published. They contain the complete results of the Study Projects and Workshops that have formed a major part of ISSI's scientific programme over the past ten years.

International "ISSI Teams" were introduced in 1997 as a new type of support for the space-science community. In contrast to the Workshop projects, these Teams define their own goals and publish their results in the appropriate scientific journals. More than 1200 scientists from 40 countries have participated in ISSI Workshops and Team meetings so far, many of them visiting the Institute quite regularly. Their commitment has made ISSI a success, and their dedication to doing science has contributed greatly to the Institute's spirit and intellectual atmosphere.

ISSI's most important merits lie in its internationality, interdisciplinarity and interactivity, providing both support and a forum for international participation, bringing together groups of scientists with broad and diverse expertise, and

providing the logistics for efficient cooperation between scientists from various disciplines, so that the synergistic benefits can be optimized. ISSI's primary focus has been on the Solar System, where the need for interdisciplinary studies has been strongest. However, it reaches out far beyond the bounds of the Solar System, building bridges between Solar System Science, Earth Science, Astronomy and even Fundamental Physics.

This book summarizes the results of ISSI Workshop projects in a number of scientific areas, to which they have made important contributions, but we do not claim any completeness. There are two parts to the volume: Part A contains essays relating to the evolution and scientific goals of ISSI, while the articles in Part B relate more specifically to the *Space Science Series of ISSI* and *ISSI Scientific Reports* volumes (see below and covers reproduced on page 251). The book has been specifically written with non-specialists in mind, and hopefully will thereby provide not only scientists from other fields but also the interested layman with a basic understanding of ISSI's achievements. But this is easier said than done. It is quite a challenge to explain in simple terms – but without falling into the trap of oversimplification – any advance in basic science or new and evolving ideas, a dilemma expressed by Paul Valéry with the words *"Tout ce qui est simple est faux, mais tout ce qui ne l'est pas est inutilisable"*. If this book succeeds – even to some extent – in overcoming this dilemma, it will be due to the scientific competence and the writing skills of the authors, who have contributed so significantly over the years to the successes of ISSI's Workshops and publications.

The *ISSI Scientific Reports (ISR)* are published by the Publications Division of the European Space Agency, and the *Space Science Series of ISSI (SSSI)* as part of the *Space Science Reviews* by Kluwer Academic Publishers and more recently by Springer Verlag. In this book we have reproduced several figures from these volumes. We thank the publishing partners for ten years of fruitful cooperation, and we are most grateful to the authors of the articles, without whose efforts this book would have been impossible.

It has been a pleasure for the Editors, who led ISSI during the early years of its life, to organize and edit this ten-year anniversary volume at the request of the current Directors, and we are grateful for their support. We are indebted to George Gloeckler, Atsuhiro Nishida, Rudolf Treumann and other colleagues for their valuable help in reviewing and editing the articles, and we thank the Publication Manager, Bruce Battrick, for the congenial cooperation with ESA in publishing this volume.

Johannes Geiss & Bengt Hultqvist
Bern, May 2005

The Need for Interdisciplinarity

R. Lallement

Service d'Aéronomie du CNRS, Verrières-le-Buisson, France

An ever-increasing number of scientific fields are using the opportunities offered by space flight that began with sounding rockets in the middle of the 20th century and were quickly followed by the historic launch of Sputnik. Since then, measurements in space and observations from space have assumed an essential role in Solar System research, astronomy and cosmology, and in the earth sciences[1]. Indeed, today space research is indispensable for a multitude of scientific fields. But that is not enough. To cope with the growing complexity of scientific problems that are now being studied with the vastly improved sensitivity and resolving power of space-borne instruments, an interdisciplinary approach is required. Such an approach can also provide us with early warnings of the undesirable side effects that go along with every rapidly evolving technology. All of these considerations were the guiding motivation in creating ISSI. In the brochure issued in 1994 by the Association Pro-ISSI one reads:

"… Perhaps the most important aim for ISSI lies in its interdisciplinarity, providing the means to draw as necessary on the methods and arguments of the appropriate branches of physics, astronomy, chemistry and earth science… " [2]

"Interdisciplinarity" is often invoked, but not as often implemented. There are practical reasons for this and, for space research, they are probably mostly the result of the involvements of the space scientists in projects that continue from one to the next. ISSI's activities are, as a rule, not related to specific space projects, and this has allowed for true interdisciplinarity. I will try to show here that ISSI has done more than organize sequences of activities in different scientific fields, which would be multi-disciplinarity. ISSI has been a place where distinct scientific communities learn from each other. It has thus been a true interdisciplinarity institution. Good examples of interdisciplinarity "at work" at ISSI that I have witnessed were the first workshop held in 1995, and one of the recent working teams, which met in autumn 2004. These examples illustrate that throughout the first 10 years ISSI has initiated and maintained interdisciplinary curiosity.

Merging Communities at ISSI

One instance of interdisciplinarity, and a most obvious one, is linked to a "physical boundary" between two physical domains. The heliospheric interface, i.e. the transition region between the distant solar wind and the interstellar medium is a perfect example. To provide a complete description of the complex interplay between ionized and neutral species, and between pickup ions and accelerated particles in the outer heliosphere requires both interplanetary and interstellar data.

In the 1970's, there was a huge gap between the length scale of parsecs (pc's) characterizing local interstellar medium measurements, provided mostly by stellar observations, and the scale of astronomical units (AUs), the size of local inhomogeneities in the immediate environment of the Sun, required for heliospheric interface studies. During the last 50 to 100 million years, the Sun has been crossing an empty region of the galactic disk, the so-called Local Bubble, which is probably a remnant of supernova explosions. No dense clouds populate this Bubble, and only a few diffuse and partially ionized cloudlets have survived in the hot and extremely tenuous gas that fills most of the "bubble" space (Fig. 1a). Because the cloudlets in the Sun's vicinity are so tiny, the gas column densities intercepted along the path-lengths to nearby stars are very small. The detection of these local gas cloudlets (Figs. 1a and b) and their disentangling began only in the 1980's, with dedicated, long-duration ground-based observations and the first UV and EUV spectra recorded from space. In parallel, much progress was achieved in heliospheric physics thanks mainly to the outer-space missions of Voyager and Pioneer, and then Ulysses. Combining these new space-based data was the goal of the first ISSI Workshop "The Heliosphere in the Interstellar Medium"[7]. This first ISSI workshop illustrated so well ISSI's approach to interdisciplinarity. Two different communities came together, the "heliosphericists", observers and plasma-physics specialists studying the interplanetary medium, and astronomers dealing with the interstellar medium. It was also a perfect example of one of the original purposes of ISSI, encouraging and starting not only interdisciplinary but also multi-spacecraft/multi-experiment investigations. No particular space mission had motivated this meeting; it was initiated by the Directors of ISSI. Data and results from a "flotilla" of spacecraft (more than 15 in total!) were presented, compared and discussed. Interestingly enough, nowadays there is no longer any meeting that deals with the outer heliosphere without the participation of members of the interstellar community. The links between heliosphericists and interstellar-medium astronomers have been strengthened, and ISSI has played a major role in establishing this connection.

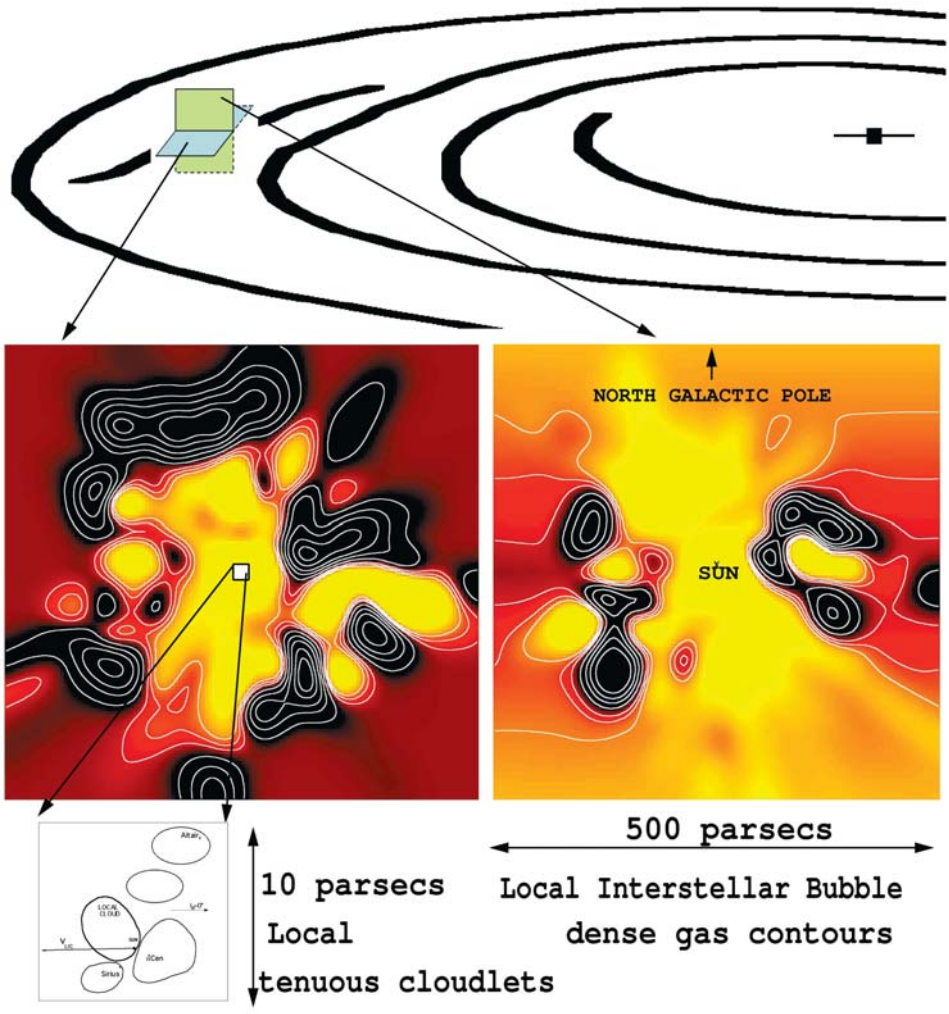


Figure 1a. The "Local Bubble" (contours from Ref. 3) contains small diffuse clouds, 5 to 10 parsecs wide[4,5]. One of these, the "Local Cloud", is presently traversed by the Sun.

During the following years, there have been a number of scientific results from this "interplanetary/interstellar connection", namely the interplay between the properties of the heliosphere and the physical state of the interstellar medium in our "galactic corner". The Hubble Space Telescope was used to measure the electron density in the local clouds. It was shown that the density is of the order of 0.05-0.1 cm^{-3} (Refs. 8-9), in agreement with some of the new dedicated ground-based data, but contrary to prevailing ideas. This number has since been used in models and in conjunction with heliospheric measurements.

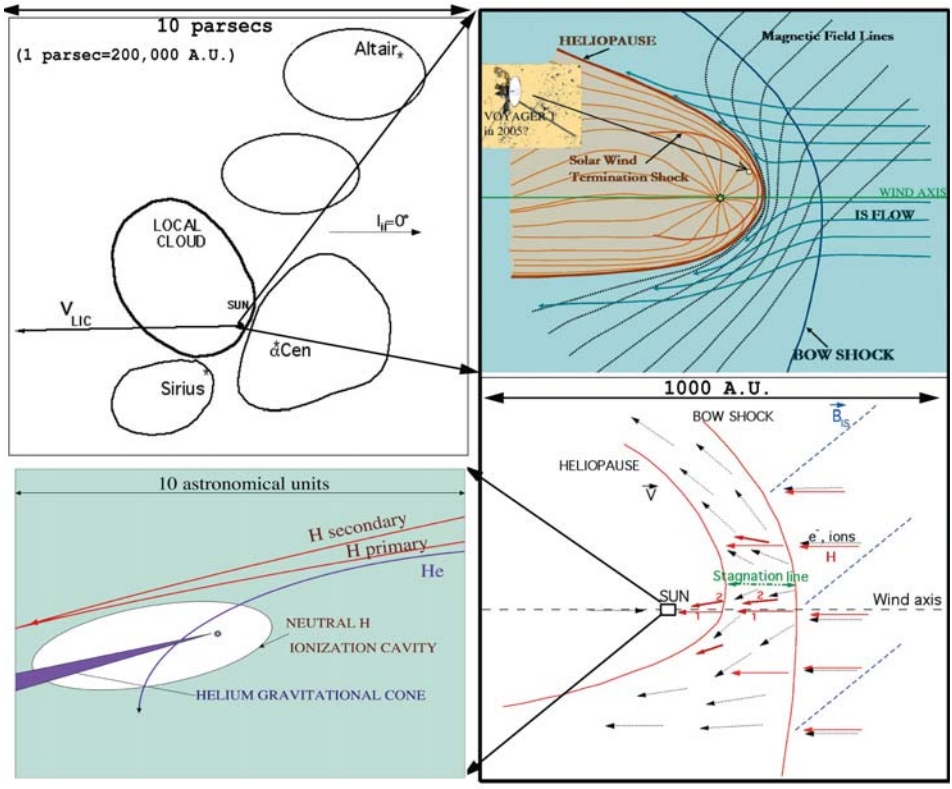


Figure 1b. Inside the Local Cloud the Sun is surrounded by the heliosphere, a tiny region filled with plasma of solar origin. The size, structure and shape of the heliosphere depend on the properties of the gas and magnetic field in the Local Cloud (scheme from Ref. 6). Neutral H atoms reaching the inner Solar System are made of primary particles and of secondary atoms which have kept the imprints of the heliospheric interface. Despite the broad scale-length range illustrated by the Figures 1a and 1b, there are links between the neutrals detected in the inner heliosphere on one hand, and the multi-phase interstellar medium on the other.

Another example is the ionization state of the interstellar helium in the local interstellar medium. Due to its high ionization potential, helium is generally neutral in cold and in diffuse, partially neutral interstellar clouds. Assuming helium neutrality within the circumsolar interstellar medium, it was difficult to explain self-consistently both the filtration or partial exclusion of neutral hydrogen entering the heliosphere and the measured velocities and temperatures of neutral helium and hydrogen inside the heliosphere. At the same time, the Extreme Ultraviolet Explorer (EUVE), a satellite dedicated to astronomy, was busy measuring the EUV spectra of white dwarfs, with the goal of studying atmospheric layers of these compact stars. However, a key ingredient (and thus a by-product of these studies) is the measured column density of hydrogen and helium

between the Sun and the target star. Using EUVE results for the column-densities of interstellar hydrogen and helium, it was found that, surprisingly, helium is significantly ionized in the local cloudlets surrounding the Sun[10].

Merging Multi-Spacecraft Data at ISSI

ISSI recently hosted a working team (the Interstellar Helium Team, led by E. Moebius from the University of New Hampshire) whose goal was to analyze all available data relevant to the flow of interstellar helium in the heliosphere in a consistent manner. This included not only observations of the main flow of the gas itself, but observations of interstellar helium outside the heliosphere, of helium derivatives inside the heliosphere, and finally of all external parameters affecting the flow, such as solar-wind and solar-radiation data.

This has been a well focused, clearly defined, very stimulating and productive working team, and in my opinion a good example of what can be conducted at ISSI. As a matter of fact, collecting and understanding extraneous data, that is data with which one is not familiar, is not something one does with enthusiasm, even if the new data appear to be directly related to one's main interest. Having in one room all the various specialists allows for the rapid removal of psychological barriers, and is a much more efficient process in reaching a scientific goal. I remember that, stimulated by the first ISSI kick-off meeting of this team, I decided to do what I would never have thought of doing otherwise, namely to reanalyze some of the "old generation" data on the helium resonance glow, gathered in the 1970's and 1980's. These older data were known to be completely contradictory to the results from the particle experiments and the EUVE spectrometers obtained in the 1990's and later. In the context of the team's activity, it appeared worthwhile to do this. In the light of the new results, and the "*in-situ*" help of the team, we finally found the cause of the contradiction. This is not "big science", but it resolved the discrepancies. Furthermore, new fits to the old data generated additional and valuable information.

One of the main products from the work of this ISSI team has been the derivation of a set of parameters for the interstellar helium flow, with unprecedented accuracy. Helium, because it enters the heliosphere without coupling to the plasma, and because it is gravitationally focused behind the Sun, to form the conspicuous helium cone (see Fig. 1b), is both a tracer of the physical conditions in the surrounding medium and our "interstellar wind flag". It can be used as a reference for the other elements, especially hydrogen. Recently, we used this set of parameters to show that charge-transfer with the ionized gas deflects the neutral hydrogen flow by a few degrees[6].

Stimulating Interdisciplinary Curiosity at ISSI

Long after the meetings or workshops are over, there is always something tangible that remains: the interdisciplinary way of looking at a problem! In practice, this means paying attention to scientific results from adjacent fields of study that one would not normally have noticed before. Because the danger of the interdisciplinary approach lies in the false interpretation or the over-interpretation of data due to a lack of experience in the "new" field, such meetings are needed. They help to avoid such misunderstandings, and permit research to proceed in a more comprehensive manner.

One concrete example of the advantages of the interdisciplinary approach to working on a problem that I personally have experienced through ISSI has to do with the Voyager spacecraft and its location with respect to the solar-wind termination shock. Since the first ISSI meeting, I have followed the scientific results obtained by both Voyagers, not too closely, but constantly, practicing what I call interdisciplinary curiosity. The recent measurements of energetic particle fluxes by Voyager 1 suggested that the spacecraft had crossed the solar-wind termination shock. However, the magnetometers onboard the same spacecraft indicated no magnetic field compression, contrary to expectations for a shock crossing. When results of a new model[11], potentially reconciling the observed anisotropies of energetic particles and the other data sets, were presented during a session of the last COSPAR meeting, I saw the connection with the difference of a few degrees between the flow directions of interstellar helium and hydrogen atoms that we were observing with the SWAN instrument on board SOHO[6]. If a distortion of the heliosphere and an offset of the heliophere apex from the wind direction were the clue in Voyager measurements, as suggested at the conference, it could also be the clue in our SWAN data. A galactic magnetic field of reasonable strength is expected to produce such a distortion, if the field is not aligned with the direction of the flow. According to the observed deviation of the neutral hydrogen flow, Voyager 1 is heading towards the most elongated part of the heliospheric "head", a favorable case for the "offset" interpretation.

Interdisciplinarity is a Win-Win Situation

In the previous paragraphs I have illustrated how the knowledge of the interstellar medium impacts on heliospheric physics. But, conversely, our "small" heliosphere is a unique astronomical tool. This reciprocity is one of the very positive aspects of interdisciplinarity. It works both ways!

The best example is the remarkable first measurement of the interstellar helium He^3/He^4 isotopic ratio[12]. This "cosmological" parameter has finally been measured for the first time with good accuracy within the heliosphere, using helium pickup ions collected by the SWICS instrument onboard the Ulysses spacecraft (pickup helium are created when interstellar neutral helium atoms, having penetrated the heliosphere (see Fig. 1b), get ionized and are convected outward with the solar wind). There is no better example of how heliospheric physics, in this case the study of the generation and propagation of interstellar secondary species, and the development of new-generation mass spectrometers mainly devoted to the solar wind, finally lead to a major result in astrophysics/cosmology.

Another example of a "feedback" result is the subject of our current investigations. From the measured deflection, deceleration, heating and filtration of interstellar H atoms, interface models will allow us to deduce the ambient galactic pressure (from neutrals and ions), and the galactic magnetic-field intensity and direction. In this sense the heliosphere is both an interstellar barometer and an interstellar magnetic compass. Together with the cosmic-ray pressure derived from Voyager measurements of cosmic-ray fluxes and gradients, and finally all the terms contributing to the total interstellar pressure will be determined "in situ". In turn, these results will shed light on the local interstellar medium. As a matter of fact, the interstellar gas pressure is only very indirectly inferred and the multi-phase structure of the interstellar medium in general is not very well known. In the case of our neighborhood, the Local Bubble (the 100 pc wide volume surrounding the Sun), there is a strong and unexplained discrepancy between the results for the interstellar gas pressure derived from the two different methods presently used. The soft X-ray method[13] makes use of the emission of the hot gas that fills the "bubble", and provides a gas pressure $nT \approx 10{,}000$-$15{,}000\ cm^{-3}K$. The second method uses abundance ratios of CI/CI* (neutral carbon and excited neutral carbon) in the diffuse clouds embedded in the Local Bubble[14], and derives a gas pressure $nT \approx 2000$-$5000\ cm^{-3}K$. This last result is in very good agreement with the gas pressure determined in the vicinity of the heliosphere[15]. A strong additional magnetic pressure in the clouds could account for the difference between these two values. In other words, the galactic magnetic field would be much higher in the local cloud than in the surrounding hot gas. An alternative explanation for the difference is a strong contribution to the soft X-ray diffuse emission by a "local" emission resulting from charge-transfer reactions between highly charged solar-wind ions and interstellar neutrals inside the heliosphere. Indeed, the actual level of this emission, which "contaminates" all X-ray emissions from extended sources, is still a matter of debate. "In situ" estimates of the local galactic pressure and magnetic-field heliosphere observations will help to place limits on the pressure of the hot gas and the degree of local soft X-ray emission.

Conclusion

I am convinced that interdisciplinary meetings, multi-experiment working teams and workshops hosted at ISSI have had educational and inspiring effects, in addition to being directly scientifically productive. This interdisciplinarity is certainly a quality that contributes to the uniqueness of the Institute.

I would like to conclude with a picture. Interdisciplinarity is not only a tool, it is also the source of surprising and amusing results. Figure 2 illustrates the serendipitous discovery of a "cometary shadow". The SWAN instrument on the SOHO spacecraft observes routinely the resonant glow of interstellar hydrogen (at 121.6nm, Ly-alpha line) over the whole sky, and in doing so it is also detecting the emission of bright comets. Solar UV dissociates water vapour released from the comet and newly formed H atoms backscatter the solar radiation exactly as do interstellar atoms. A movie based on comet Hale-Bopp glow images was being prepared in the laboratory, when our attention was drawn to a small feature looking like an instrumental "artifact". On further inspection, it appeared to be the shadow of the absorbing cometary cloud on the interstellar gas. This shadow results from blockage of solar radiation by cometary hydrogen atoms, thereby preventing the full illumination of the interstellar gas behind the comet, and thus reducing backscatter radiation from the hydrogen in this gas. Again, both "sides" benefit. On the "cometary side", this observation has provided us with a new way of estimating the evaporation rate of comet Hale-Bopp. On the "interstellar side", the discovery of the comet's shadow will be used in conjunction with other data to measure multiple scattering effects in the interstellar hydrogen ionization cavity.

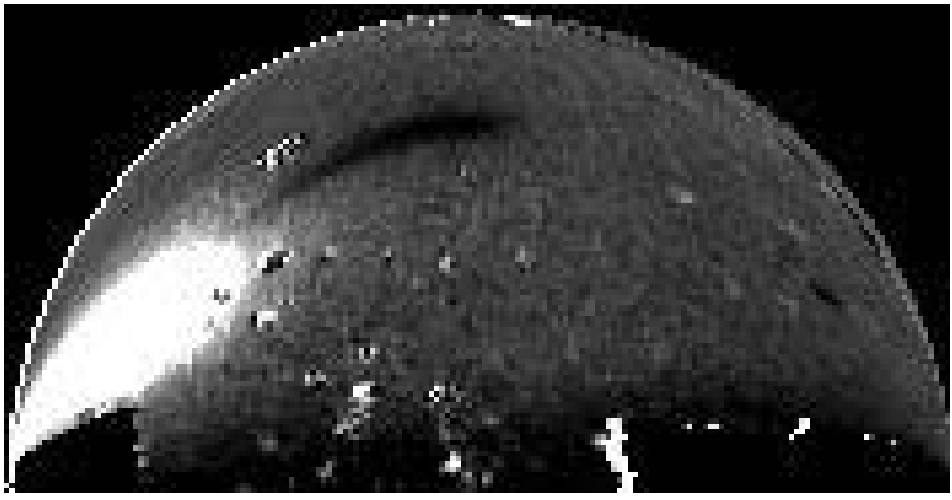


Figure 2. Comet Hale-Bopp: a shadow in the interstellar gas (from Ref. 16)

References

1. J.A. Van Allen, R. Lüst, J-E Blamont & B. Hultqvist in "The Century of Space Science" J.A.M. Bleeker, J. Geiss & M.C.E. Huber (Eds.), Kluwer Academic Publishers, Dordrecht, 2001.
2. J. Geiss *et al.,* "International Space Science Institute", Brochure published by the Pro-ISSI Association, Bern, Switzerland, 1994.
3. R. Lallement, B.Y. Welsh, J.L. Vergely & F. Crifo, *Astron. & Astrophys.*, **411**, 447 , 2003.
4. S. Redfield & J. Linsky, *Astrophys. J.*, **534**, 825, 2000.
5. R. Lallement, R. Ferlet, A.M. Lagrange, M. Lemoine & A. Vidal-Madjar, *Astron. Astrophys.***, 304**, 461, 1995.
6. R. Lallement, E. Quémerais, J.L. Bertaux, S. Ferron, D. Koutroumpa & R. Pellinen, *Science*, **307**, (5714), 1447, 2005.
7. R. von Steiger, R. Lallement & M.A. Lee (Eds.), The Heliosphere in the Local Interstellar Medium, SSSI Vol. 1, Kluwer Academic Publ., Dordrecht, 1996, and *Space Science Rev.,* **78**, Nos. 1-2, 1996.
8. B. Wood & J. Linsky, *Astrophys. J.*, **474**, 39, 1997.
9. R. Lallement, P. Bertin, R. Ferlet, A. Vidal-Madjar & J.L. Bertaux, *Astron. Astrophys.*, **286**, 898, 1994.
10. B. Wolff, D. Koester & R. Lallement, *Astron. Astrophys.*, **346**, 969, 1999.
11. R. Jokipii, J. Giacalone & J. Kota, *Astrophys. J.*, **611**, L141, 2004.
12. G. Gloeckler & J. Geiss, *Nature,* **381**, 210, 1996.
13. S. Snowden, M. Freyberg, K. Kuntz & W. Sanders, *Astrophys. Sup. Ser.*, **128**, 171, 2000.
14. E.B. Jenkins, *Astrophys. J.*, **580**, 938, 2002.
15. V. Izmodenov, J. Geiss, R. Lallement, G. Gloeckler, V.B. Baranov & Y.G. Malama, *J. Geophys. Res.*, **104**, A3, 4731, 1999.
16. R. Lallement, J.L. Bertaux, K. Szegö & S. Nemeth, *Earth, Moon & Planets*, **90**, 67-76, 2002.

The Exploration of the Heliosphere in Three Dimensions with Ulysses
– A Case Study in International Cooperation

L.A. Fisk

Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, USA

On 6 October 1990, the Ulysses spacecraft was launched to begin the exploration of the heliosphere in three dimensions, that vast region of space carved out by the influence of the Sun. A European Space Agency (ESA) spacecraft, launched and tracked by NASA, with instruments provided by both Europe and the United States, Ulysses is a classic example of international cooperation. Like all classic examples, however, Ulysses provides lessons on what to do and what not to do. (The Ulysses mission has had many names, from the Out-of-the-Ecliptic Mission to the International Solar Polar Mission to its current name, Ulysses, which was given to it in 1984. The mission has been the same, but the name has changed, and so for simplicity in this article, we will use the name Ulysses throughout.)

Figure 1. Ulysses, also known as the Out-of-the-Ecliptic or International Solar Polar Mission

Ulysses has been a long saga, dating from the early 1970s and continuing even today. It had its initial trials and, like all good stories, it ends in triumph. I will relate that story here, for the lessons it provides. The Ulysses story is also intertwined with the ISSI story. There are common participants who shaped ISSI in part by their Ulysses experience. And ISSI has been the venue where the triumphs of Ulysses have been exhibited.

We begin by discussing why the exploration of the heliosphere was considered to be important, and how the current Ulysses mission evolved. We then review some of the discoveries made that reserve for Ulysses a place in history. Our story will end with ISSI; the threads that developed in Ulysses evolve into ISSI, and at ISSI the discoveries of Ulysses are honed into understanding.

The Beginning

It is a simple fact of orbital dynamics that when you launch a spacecraft from Earth into the Solar System, the main velocity vector lies in the plane of the Earth's orbit, and thus it is confined to lie near the equatorial plane of the Sun. Prior to Ulysses, then, all spacecraft that were launched to explore the region of space influenced by the Sun were confined to a relatively narrow plane in an otherwise vast three-dimensional region. Indeed, prior to Ulysses, we referred to our region of exploration as the interplanetary medium, or interplanetary space, in recognition that it was the region between the planets, whose orbits all lie near this single plane. Ulysses, as we will discuss, has explored the full three-dimensional heliosphere, and created true heliospheric science.

The deficiencies of our exploration were well recognized early in the space programme. There was no expectation that the interplanetary medium that we were able to explore was representative of the broader heliosphere. Our sampling was certainly biased. The outer atmosphere of the Sun continually expands into space to form the solar wind. The fastest solar wind originates from so-called "coronal holes" on the Sun, where the magnetic field is open, giving easy escape to the solar wind and resulting in low densities in the solar corona that appear dark in X-rays (hence the name coronal holes). Coronal holes are very pronounced at the poles of the Sun during periods of low solar activity – solar-minimum conditions. There was an expectation, then unproven, that the solar wind would exhibit pronounced variations with solar latitude.

The expected configuration of the magnetic field in the solar wind results from the wind's outward expansion, which drags the field radially outward, and from the assumption that the field remains attached to the rotating Sun. A spiral pat-

tern for the field will result. At our location at Earth, near the solar equatorial plane, the spiral pattern is relatively tightly wound. But over the solar poles, the magnetic field should be essentially radial, which presents much different conditions for the inward access of cosmic rays from the Galaxy. Indeed, during certain periods of the solar cycle, we might expect that cosmic rays would have easy access over the solar poles; in particular, lower-energy galactic cosmic rays, which are otherwise excluded, should be present in the inner heliosphere.

Perhaps the most significant aspect of our pre-Ulysses ignorance of the three-dimensional heliosphere was our lack of knowledge of how the magnetic field of the Sun reverses polarity during the solar cycle. The Sun has a 22-year magnetic cycle; the average dipolar field of the Sun reverses polarity every 11 years, with the reversal occurring shortly after the period of maximum solar activity. The polarity change is manifested in the magnetic field in the solar wind, which also reverses polarity. The question is how does this physically occur. Is old magnetic flux shed and new magnetic flux emitted; is there a time when the magnetic field in the solar wind has mixed polarity, and on what scale? Interestingly enough, this was not a question that was really asked or speculated upon prior to the launch of Ulysses, nor was it an anticipated discovery. Perhaps it was the uncertainty of when Ulysses would perform its exploration, and for how long. If the time period around solar maximum was not included, the subject of the field reversal might not be addressed. As we shall see, however, understanding the field reversal of the Sun has been one of the seminal discoveries of Ulysses.

In the late 1960s and early 1970s, the clamour for a mission that would explore the three-dimensional heliosphere began to build. Fortunately, it began to build on both sides of the Atlantic, with two principal advocates, Harry Elliot of Imperial College in Europe, and John Simpson from the University of Chicago in America. There was thus an opportunity for a joint US/European mission. In 1974, a delegation of US scientists and NASA planners journeyed to ESTEC in Noordwjik, The Netherlands, to discuss the possibility of a joint mission. To fly over the poles of the Sun, you must first go to Jupiter and use its gravitational field to redirect the trajectory of the spacecraft into the polar direction, as is illustrated in Figure 2. The question, then, was how best to do this as a cooperative mission. In a productive session at that joint meeting in 1974, Ian Axford, then Director of the Max Planck Institute at Lindau, Germany, went to the blackboard and drew what was to become the mission design: two spacecraft, one provided by ESA and one by NASA, both launched together towards Jupiter. Their trajectories split at this point, with one spacecraft heading over the south pole of the Sun, and the other heading towards the north pole. It was to be an ideal science mission, with the conditions over each pole observed simultaneously, and temporal and spatial effects readily separable. It was also to be an ideal interna-

tional cooperation: clean interfaces, with separate spacecraft, so that each side could readily perform its tasks without being overly dependent upon the other (a situation that proved to be both a blessing and a curse).

ESA and NASA have separate and different approval processes for new missions. And so, with a mission design in hand, each side went back to begin their separate, arduous processes of approval. In the US, a new mission must be first sold to NASA and the Executive branch of Government for a so-called "new start", and then sold to Congress for funding approval. NASA responds to community pressure, and so a workshop was held in May of 1975 at the Goddard Space Flight Center (close to NASA Headquarters) to demonstrate community interest and excitement in the study of the three-dimensional heliosphere. Even with such community encouragement, there was bound to be stiff competition for new starts among the various science disciplines that NASA supports. Indeed, the general field of solar and heliospheric physics had previously suffered in this competition because it was not well represented at NASA Headquarters. Fortunately, shortly before the new-start discussions, NASA had reorganized to form a Solar-Terrestrial Division, on equal standing with the more powerful division that pursued planetary exploration, and these circumstances made it possible to obtain a new start for Ulysses. New starts must then be approved by Congress, and after much encouragement from the science community, the new start for Ulysses was obtained in 1978.

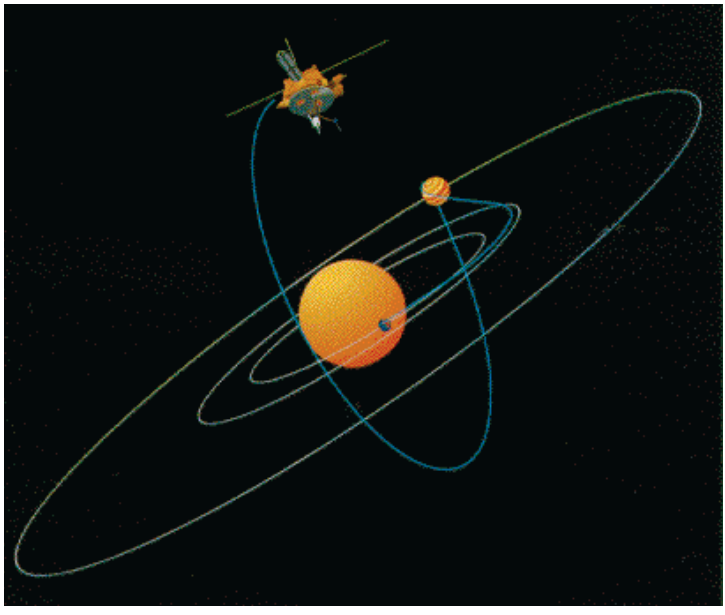


Figure 2. The trajectory of Ulysses, first to Jupiter to use its gravitational field to redirect the trajectory to fly over the poles of the Sun.

ESA has a different process from NASA, but with similarities. The issue is not persuading a Parliament, as with the US Congress, to approve Ulysses, but rather scientific representatives of the various Member States of ESA must be convinced. Herein lay the similarity. Someone had to be convinced that Ulysses was more worthy than competing projects, and the job fell to the science advocates of Ulysses. There were many acknowledged and unsung heroes in this effort, on both sides of the Atlantic, but since this is a story that weaves its way to ISSI, the role of Johannes Geiss should be acknowledged. As chair or member of various committees of the European Space Research Organisation (ESRO) and of ESA, Johannes Geiss was involved in many of the early discussions about an out-of-ecliptic mission. He became convinced of the importance of this exploration when the plans included flying to high heliospheric latitudes, and then as chair of the Solar System Working Group he defended the project energetically and with skill.

With the new starts well underway in Europe and the US, it was time to select the payloads. NASA, being ever more ambitious, was to fly a coronograph that would look downward from above the solar poles on the full corona of the Sun that flows outward and affects Earth. The payload on the ESA spacecraft was to be a full range of particle and magnetic-field experiments (see, for instance, the results in the volume edited by A. Balogh, R.G. Marsden & E.J. Smith[1]). This being a joint European/US mission the scientific community on both sides of the Atlantic distributed themselves so that almost every experiment was an international collaboration. The selection then resulted in instruments on the ESA spacecraft with a strong US role, and instruments on the US spacecraft with a strong European role. One interesting difference, however, was that the Europeans were prepared to take more risks than the Americans. The Solar Wind Ion Composition Spectrometer, known as SWICS, was a US/European collaboration between the University of Maryland, with George Gloeckler as Principal Investigator, and major hardware contributions from the Max Planck Institute at Lindau, Germany, the University of Braunschweig, Germany, and the University of Bern, Switzerland, with Johannes Geiss in the lead. SWICS depended on an innovative design that involved the highest voltages ever to be flown in space, up to 30,000 volts[2]. NASA was less comfortable with what they perceived to be the risk of SWICS, but the Europeans were more willing to accept it, perhaps because they delegate to the experimenters, and their national support, the responsibility for scientific instruments.

The unique orbit of Ulysses had to be matched by unique, advanced, and even entirely new scientific instruments. That was the opinion of Johannes Geiss and the Solar System Working Group, because – looking at the very limited resources of ESA's Science Programme – they did not believe a second mission

for high-latitude exploration would be likely. Thus, not only was SWICS accepted, but an experiment of the Max-Planck-Institute in Heidelberg was included as well. This instrument discovered interstellar grains deep inside the heliosphere. A most modern cosmic-ray experiment built by an American-European team headed by John Simpson was also among the experiments accepted for the European spacecraft. (They all have been operating successfully on Ulysses.)

In the late 1970s, then, all was well, with an exciting mission design and an excellent payload. Then disaster struck. Reagan was elected President of the United States in 1980, and as one of his first acts in office he slashed the NASA budget, and NASA unilaterally, and without consultation with ESA, cancelled the US spacecraft for Ulysses, and with it the opportunity to fly the international instruments it was to carry. NASA did continue to support its contribution to the instruments on the ESA spacecraft, and was to provide the launch and tracking. Those US experimenters on the ESA spacecraft were fortunate. Those European experimenters on the NASA spacecraft, and their US colleagues, were not!

There was much outcry. Not only for the loss of science, but also for the way it was done. There was a Memorandum of Understanding between NASA and ESA governing Ulysses, but like all MOUs it had an escape clause for the signatories, which NASA exercised. It was an unfortunate time in NASA for this event to occur. During the change in Presidential Administrations, there was no NASA Administrator in place who might have successfully reversed the cancellation. And so it occurred, and influenced and clouded NASA-ESA relationships for years to come. Not to be burned like this again, ESA has insisted on more formality in its agreements than breakable MOUs for major cooperations with the US such as the International Space Station, and it has sought to maintain a vibrant Science Programme that is not dependent on the US.

The clean interface established early in the programme, with a separate ESA spacecraft, allowed the programme to continue, and since the US was still providing the launch and tracking, ESA honoured its commitments to the instrumentation that was being provided by the US for the ESA spacecraft, and the programme proceeded. However, like all spacecraft of that time, Ulysses was to be launched on the Shuttle, but the development of the Shuttle was delayed, and then the “Challenger” accident occurred, which delayed the launch still further. Not until October of 1990, sixteen years after that first meeting in ESTEC to design an exciting, cooperative mission between NASA and ESA, did Ulysses actually begin its journey to explore the heliosphere in three dimensions.

Lessons Learned for Cooperative Science Missions

There are lessons to be learned from the saga of Ulysses. First, the best science missions develop as grassroots efforts, where working scientists recognize an important scientific problem to be studied and conceive of a clever mission to pursue it. Second, important scientific problems are not the exclusive province of either Americans or Europeans, or other nations, and they are best pursued, whenever feasible, as an international adventure. Third, the scientific problem has to be of such importance that it remains central to the pursuit of a major science discipline for many decades, until the mission to study it is in fact realized. Fourth, there will be inevitable political obstacles that need to be overcome or worked around, which requires clever mission designs, and continuous vigilance and perseverance on the part of the science advocates of the mission.

When Johannes Geiss conceived of ISSI's role in the international space-science effort in the early 1990s, these lessons of Ulysses had to be in his mind. If missions are to be based on recognition of an important scientific problem, there has to be agreement on what the state of that problem is – what is known and what is left, or indeed required to be discovered. The many ISSI books that summarize the state of knowledge of broad problems in space science very admirably serve this purpose. Important and lasting scientific problems require continuous nurturing, and the opportunity to discuss them over many years and many workshops is important. Scientists develop consensus on scientific knowledge and on missions to pursue, not in isolation but by interacting in person, and the many ISSI workshops and team meetings serve this purpose well. Perhaps most important is the international flavour of ISSI, and its full recognition that knowledge resides throughout the World. ISSI is founded on the principle that the best science and the best missions need to have access to the broad international capability.

It was perhaps fitting also that one of the significant conversations that led to the formation of ISSI occurred at the launch of Ulysses in 1990. Roget Bonnet, then Director of ESA's Science Programme and now Executive Director of ISSI, and I, then the Associate Administrator of NASA, had the opportunity at the launch to discuss Johannes Geiss's concept of ISSI. It was fitting that the lessons of Ulysses and the spirit of international cooperation that it engendered led to our both believing that this was a concept worth pursuing and endorsing.

The Scientific Discoveries of Ulysses

Each of the scientists who participated in Ulysses, and for that matter worked in solar and heliospheric physics, will have their own list of the most significant Ulysses discoveries. Below are the ones on my personal list.

We should note in passing that all of the significant discoveries that can be expected to be on anyone's list have been a subject of discussion at an ISSI workshop, where their significance has been evaluated and their impact on broader science questions appreciated. At the very first ISSI workshop[3] on *The Heliosphere in the Local Interstellar Medium*, which was a seminal event in our understanding of the interactions between our star the Sun and interstellar space, a central theme was interstellar neutral gas in the heliosphere, which was observed in detail by Ulysses, both as neutral particles and for the first time as pickup ions. In the workshop[4] on *Cosmic Rays in the Heliosphere*, the observations from Ulysses were also central. The behaviour of cosmic rays in the heliosphere can only be understood as a three-dimensional problem, and without Ulysses little progress in this field would have been made. The workshop[5] on *Corotating Interaction Regions* (CIR) was built around the Ulysses observations. Particles are accelerated in CIRs, stream-stream interaction regions in the solar wind that occur at low heliographic latitudes. Yet the accelerated particles were observed by Ulysses at high heliographic latitudes, indicating that a fundamental rethinking of the configuration of the magnetic field in the solar wind is required. In *Solar Composition and its Evolution – from Core to Corona*,[6] the Ulysses observations of the composition of the solar wind, indeed the first such observations, were of primary importance.

There are three items that are high on my personal list of significant Ulysses discoveries. The first, which is highlighted above, is the composition of plasma particles in the heliosphere. It is important to note how primitive our measurements of the composition of the solar wind were prior to Ulysses. The Swiss foil experiments conducted by Johannes Geiss during the Apollo missions to the Moon in the 1960s gave us detailed information on the isotopic composition of the solar wind, but then only for a very brief period of time[7]. Instruments of the time were routinely measuring the solar wind, separating particles only by their mass/charge ratio, with the result that sensitivity and resolution were relatively low; only a limited number of elements and charge-states could be uniquely determined. The SWICS instrument on Ulysses, designed by George Gloeckler, provided the first detailed and continuous observations of the composition of the solar wind, separating charge from mass, and of course performing these observations away from the Sun en route to Jupiter and over the wide range of latitudes and solar conditions observed[2] by Ulysses.

The discoveries are many. By routinely determining the composition of the solar wind, the observations provided by SWICS unlock the power of composition measurements to understand fundamental solar processes. The elemental composition of the solar wind is biased according to the First Ionization Potential, or FIP, of its elements[8]. A FIP bias must be established very close to the Sun, where the particles are just being ionized, and thus the FIP bias can be used to identify the regions on the Sun from where the solar wind originates. The charge-states of the solar wind are frozen-in in the corona of the Sun, and reveal the conditions under which the solar wind is being accelerated; there is a strong anti-correlation between the solar-wind speed and the coronal electron temperature as determined from solar-wind charge states[9]. Coronal Mass Ejections, large eruptions of material from the Sun, contain plasma particles with unique charge states, and thus the composition measurements of SWICS on Ulysses provide a powerful identifier of remnant CMEs in the heliosphere[10].

Then there are the pickup ions. Prior to Ulysses, interstellar neutral gas was known to flow through the heliosphere. It was expected that the neutral gas would be ionized by charge-exchange with the solar wind and photo-ionization, and once ionized, it should be picked up by the solar wind and convected outward. This pickup-ion population should have a profound effect on the outer heliosphere. It is the dominant energy input into the solar wind, such that if this population were to remain separate from the core solar wind it would be the dominant internal pressure force in the solar wind beyond about the orbit of Saturn. Pickup ions were also predicted to be the source of Anomalous Cosmic Rays (ACRs), a component of the cosmic rays with an unusual composition that resembles the composition of interstellar neutral gas[11]. For this to occur there must be a major acceleration process in the outer heliosphere, presumably at the termination shock of the solar wind, since pickup ions are formed with energies of ~1 keV/nucleon, but ACRs occur at tens of MeV/nucleon. Prior to Ulysses, however, the main species of the pickup ions had never been observed. There was evidence for pickup helium, since this species can penetrate as neutral to within the orbit of Earth, before being ionized and convected outward. However, the dominant pickup ion species, hydrogen, and such interesting species as oxygen, do not penetrate to within the orbit of Earth, and could only be observed and measured in detail by SWICS on Ulysses, for the first time, en route to Jupiter[12].

There is now an entire branch of science built around pickup ions. New sources have been discovered. Solar wind particles appear to become embedded in dust grains near the Sun, and are released to form a pickup-ion population known as the inner source[13,14]. Comets emit neutrals that form pickup ions, and have an extended tail that can be observed, from which the composition of the comet can be determined[15]. Interstellar neutral gas is a measure of the local interstellar

medium and thus the composition of the Galaxy in the present epoch, as opposed to when the Solar System was formed 4.5 billion years ago[16]. The isotope helium-3 was measured in the pickup ions by Ulysses, and provides an important constraint on the evolution of baryonic matter in the Universe. All this from the pioneering measurements of SWICS on Ulysses.

The third item on my personal list of significant Ulysses discoveries is a much-improved understanding of the reversal in polarity of the large-scale magnetic field of the Sun, a fundamental solar problem.

Consider first solar-minimum conditions. At this time, the magnetic field at the poles of the Sun opens into the heliosphere and allows the escape of fast solar wind. The magnetic field is relatively strong in this region. As it opens into the corona, it comes into pressure equilibrium, with uniform field strength, and points radially outward as it is dragged outward with the solar wind. The polarity of the magnetic field at each solar pole is of course opposite. Once expanded into the heliosphere, the two regions of opposite polarity are separated by a single current sheet, which during solar-minimum conditions lies near the equatorial plane of the Sun.

At the next solar minimum, 11 years hence, the polarities in the two polar regions are reversed. The question is, how does this occur? The original theories for the field reversal of the Sun have new magnetic flux rising through the solar surface with the onset of solar activity, and then migrating to the solar poles such that the migrating flux has opposite polarity to that of the nearest solar pole[17,18]. Flux annihilation occurs and the old polar flux is replaced by magnetic flux of opposite polarity. However, magnetic flux emerges in the form of closed magnetic loops. The magnetic field at the solar poles at successive solar minima opens into the heliosphere. How was the closed flux turned into open flux, and what should we have seen in the heliosphere?

Ulysses in fact observed a remarkably simple picture. That single current sheet separating the two regions of opposite magnetic polarity, which lies near the equatorial plane during solar minimum, appears to be preserved throughout the solar cycle. It simply rotates through 180 degrees to accomplish the field reversal. A single preserved current sheet has profound implications[19,20]. Since the open magnetic flux can be eliminated from the Sun only by reconnection at the current sheet, and little reconnection seems to be occurring, there appears at some level to be a constancy to the open magnetic flux of the Sun. The field reversal, the formation and dissipation of coronal holes, all these will need to be accomplished by moving the open flux about on the Sun, presumably by moving it along the solar surface.

One final personal note. Over the last several years, I have written a number of papers about how the open magnetic flux of the Sun should behave and move about on the solar surface[21]. The progenitor for this theoretical work was a paper[22] explaining how motions of the open flux could distort the configuration of the heliospheric magnetic field, allowing energetic particles to propagate from CIRs at low latitudes to high latitudes where they were observed by Ulysses (the main subject of the ISSI workshop[5] on *Corotating Interaction Regions*). The principal calculations for this progenitor theory were done on a plane ride to ISSI, for an entirely different workshop, where I was able to share my ideas with my colleagues, in off-line conversations, and get the feedback I needed to proceed – which is after all what ISSI is all about.

References

1. A. Balogh, R.G. Marsden & E.J. Smith (Eds.), The Heliosphere Near Solar Minimum, Praxis Publishing Ltd., Chichester, UK, 2001.
2. G. Gloeckler *et al*., *Astron. Astrophys. Suppl. Ser.,* **92**, 267, 1992.
3. R. von Steiger, R. Lallement & M.A. Lee (Eds.), The Heliosphere in the Local Interstellar Medium, SSSI Vol. 1, Kluwer Academic Publ., Dordrecht, 1996, and *Space Science Rev.,* **78**, Nos. 1-2, 1996.
4. L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger & K.-P. Wenzel (Eds.), Cosmic Rays in the Heliosphere, SSSI Vol. 3, Kluwer Academic Publ., Dordrecht, 1998, and *Space Science Rev.,* **83**, Nos. 1-2, 1998.
5. A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach & H. Kunow (Eds.), Corotating Interaction Regions, SSSI Vol. 7, Kluwer Academic Publ., Dordrecht, 1999, and *Space Science Rev.,* **89**, Nos. 1-2, 1999.
6. C. Fröhlich, M.C.E. Huber, S.K. Solanki & R. von Steiger (Eds.), Solar Composition and its Evolution – from the Core to Corona, SSSI Vol. 5, Kluwer Academic Publ., Dordrecht, 1999, and *Space Science Rev*., **85**, Nos. 1-2, 1998.
7. See the recent review article by J. Geiss *et al*., *Space Science Rev*., **110**, 307, 2004.
8. J. Geiss *et al*., *Science,* **268**, 1033, 1995a.
9. G. Gloeckler, T.H. Zurbuchen & J. Geiss, *J. Geophys. Res.,* **108**, 1158, doi:10.1029/2002JA009286, 2003.
10. See, for example, S.T. Lepri *et al*., *J. Geophys. Res.,* **106**, 29231, 2001.
11. L.A. Fisk, B. Kozlovski & R. Ramaty, *Astrophys. J. Lett.,* **190**, L35, 1974.
12. G. Gloeckler *et al*., *Science,* **261**, 70, 1993.
13. J. Geiss, G. Gloeckler, L.A. Fisk & R. von Steiger, *J. Geophys. Res.,* **100**, 23373, 1995b.
14. G. Gloeckler, L.A. Fisk, J. Geiss, N.A. Schwadron & T.H. Zurbuchen, *J. Geophys. Res*., **105**, 7459, 2000a.
15. G. Gloeckler *et al*., *Nature,* **404**, 576, 2000b.
16. G. Gloeckler & J. Geiss, *Nature,* **381**, 210, 1996.
17. H.W. Babcock, *Astrophys. J.,* **133**, 572, 1961.
18. R.B. Leighton, *Astrophys. J.,* **140**, 1547, 1964.
19. E.J. Smith, A. Balogh, R.J. Forsyth & D.J. McComas, *Geophys. Res. Lett.,* **28**, 4159, 2001.
20. G.H. Jones, A. Balogh & E.J. Smith, *Geophys. Res. Lett.,* **30**, 8028, doi:10.1029/2003GL017 204, 2003.
21. See, for example, L.A. Fisk & N.A. Schwadron, *Astrophys. J.,* **560**, 425, 2001.
22. L.A. Fisk, *J. Geophys. Res.,* **101**, 15547, 1996.

The Astrophysical Relevance of Space Plasma Physics

R.A. Treumann[a,b] and R.Z. Sagdeev[c]

[a]*Theory Department, Max-Planck-Institute for Extraterrestrial Physics, Garching, Germany*
[b]*Department of Geosciences, Ludwig-Maximilians University, Munich, Germany*
[c]*Department of Physics, University of Maryland, College Park, USA*

Introduction

Plasmas are abundant in the Universe. They constitute more than 90% of the baryonic matter, from the hot dilute gas in clusters of galaxies, to the halos of galaxies, supernova remnants, accretion disks, stellar winds, stellar atmospheres, and the magnetospheres of neutron stars. We know about them from observation of the radiation they emit, and we interpret these observations with the help of atomic and nuclear data obtained in the laboratory[1]. Unfortunately, in the laboratory the production of plasmas with properties similar to most natural plasmas is impossible. It is thus a fortunate circumstance that Earth's environment is filled with a dilute, high-temperature plasma[2,3] which allows one to study its properties in situ. This fact has turned near-Earth space into an accessible plasma-physics laboratory that can serve the astrophysical needs.

Among the first discoveries of space plasma physics were the solar wind and the magnetosphere of Earth. The solar wind is the continuous plasma stream emitted radially out from the Sun. It compresses Earth's magnetic field and confines it to the magnetosphere. The solar wind is supersonic. Hence, as it flows around the magnetosphere a bow shock builds up in front of the magnetosphere. At much larger distance from the Sun, the solar wind interacts with the interstellar gas forming the heliosphere, with its outer boundary, the heliopause, and the termination shock standing in the heliosphere at some distance before the heliopause. Figure 1 illustrates these three important Solar System plasma regions. Physical processes and properties of their plasma populations are presented elsewhere in this book[4].

Stellar winds have been modelled after the solar wind, and the interaction of supersonic stellar winds with the interstellar gas has been modelled after the

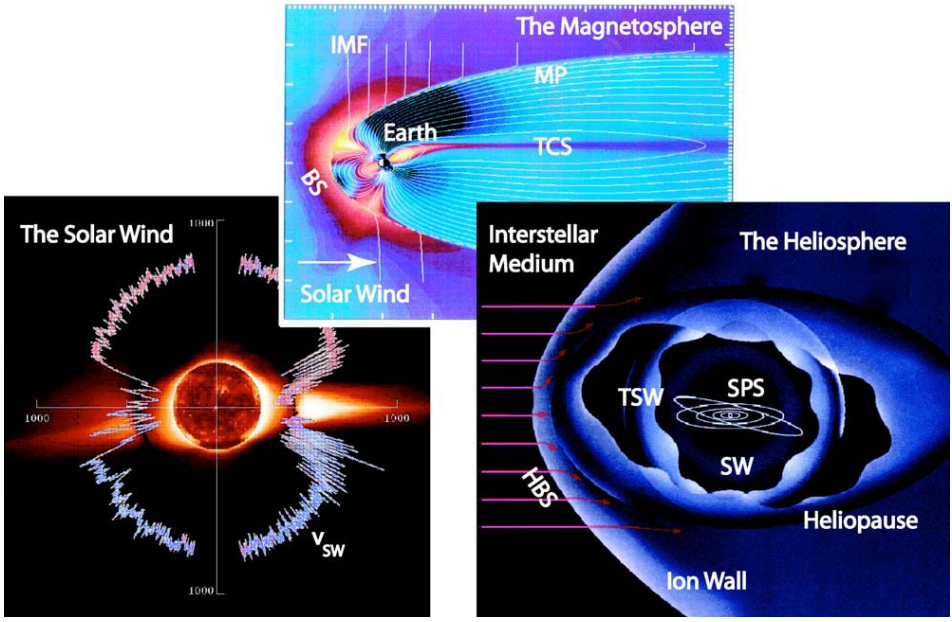


Figure 1. A combined view of (*left*) the Sun with its corona and expanding solar wind, (*right*) the entire heliosphere, and (*top insert*) Earth's magnetosphere. The scale increases from left to right. The solar planetary system (SPS) appears as the small system of elliptic planetary orbits. *Left*: The solar wind is represented by its velocity vectors v_{SW}, in dependence on solar latitude. Bright is the dense low velocity solar wind in the ecliptic plane. *Right*: The bubble of the heliosphere in the interstellar medium produced by the solar wind. Its boundary is the Heliopause; HBS is the heliospheric bow shock. A standing termination shock (TSW) evolves inside the heliosphere. *Top insert*: The magnetosphere with its bow shock (BS) and magnetopause (MP) boundary. MT is its long magnetospheric tail containing a thin tail current sheet (TCS). The interplanetary magnetic field (IMF) penetrates the magnetopause due to reconnection. Colour indicates plasma density and temperature; both are large between the bow shock and the magnetosphere and in the TCS.

heliosphere. Similarly, the magnetosphere served as a model for the magnetospheres of the magnetized planets[5-6], of pulsars[7], and even of black holes[8].

Physical dimensions in the solar wind and Earth's magnetosphere are typically much smaller than the mean free path – the distance a particle travels before colliding with another particle. The plasma is then said to be "collisionless". On the other hand, in the large astronomical objects – the galaxies and clusters of galaxies – the physical dimensions are typically larger than the mean free path of the plasma constituents. This has been taken as justification for a description of astrophysical plasmas as fluids. However, in the plasma the validity of fluid dynamics is limited not only by the mean free path, but also by the ion gyration radius, which limits the motion of the charged plasma components in the pres-

ence of a magnetic field. Since in most natural plasmas, the mean free path is much larger than the gyration radius, it is characteristic for natural plasmas that important processes take place on small scales, much smaller scales than the mean free path. Indeed, a multitude of in-situ measurements performed during the past four decades in the plasma of Earth's environment have demonstrated that the dynamics of hot dilute plasmas are governed by collisionless processes. Thus, not only in the heliosphere, but also in a large variety of astrophysical object shock waves, magnetopauses and many other plasma phenomena do actually evolve under collisionless conditions. The plasma in Earth's environment offers the unique opportunity to study these phenomena in-situ.

In the following we discuss three collisionless plasma processes and their relevance for astrophysics: the merging of magnetic flux tubes, i.e. "reconnection", the generation of electric potential drops along the magnetic field, and the formation of "collisionless shocks".

Collisionless Reconnection: The Microscopic View

The left part of Figure 2 shows the simplified mechanism of reconnection: two oppositely directed magnetic flux tubes of about equal magnetic field strength approach each other at some velocity. At the point where they come into contact, the flux tubes rearrange and stretch; a process in which the plasma tied to the flux tubes is accelerated into two oppositely directed jets that emanate from the site of first contact. The existence of this mechanism in the collisionless space plasma has been proved[9] by in-situ observation of plasma jetting at the Earth's magnetopause.

The mechanism of collisionless reconnection

Reconnection occurs where collisionless plasmas containing oppositely directed magnetic fields approach each other[10]. Its effect is the release of the magnetic energy that was stored in the sheared magnetic field. In a collisionless plasma, this energy cannot be dissipated by friction and heating. Instead, the energy is transformed into the kinetic energy of the jets, and part of it goes into the acceleration of a small number of particles to high energies[11]. In addition, reconnection reorders the plasma in such a way that two initially separated plasmas can mix after reconnection has taken place by moving along the newly formed magnetic flux tubes, as shown in Figure 2.

According to this description, reconnection appears a simple process. However, the devil is in the detail. Reconnection requires that plasma and magnetic field decouple locally, otherwise the magnetic fields could not rearrange. But in col-

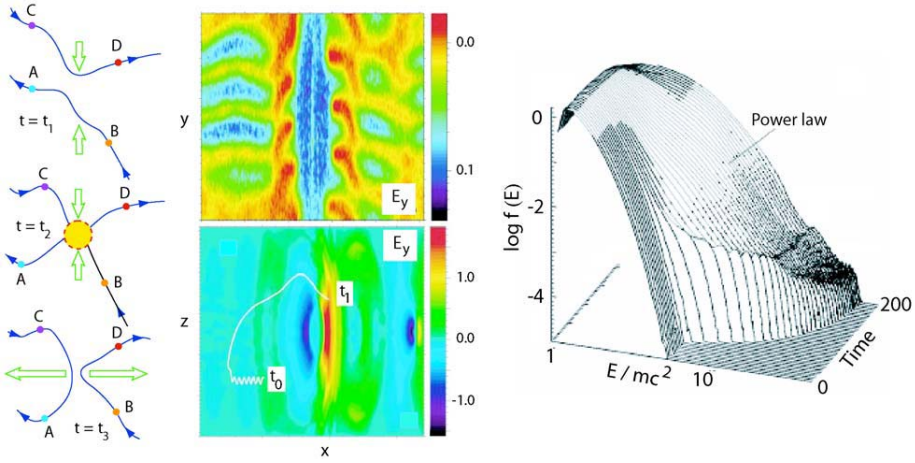


Figure 2. *Left*: Schematic representation of the reconnection process. Two magnetic flux tubes of opposite orientation in slow approach from above and below contact, reconnect (yellow), and escape to the right and left. Two fast-diverging plasma jets (green) are generated. Plasma elements A-D reorder with respect to their field lines (from Refs. 15 and 16). *Centre top*: The finite extension of the central reconnection region along the current, as seen in the electric field. The current is flowing downward. Blue and green indicate the direction of downward fields, yellow and red upward fields. To the left and right of the reconnection region, the horizontal structure shows the wavy kinking of the current sheet. *Centre below*: Side view of the same reconnection region. The current is out of the plane. Blue and red indicate upward and downward directed reconnection electric fields. The important observation is the white curve, which is the orbit of one particular arbitrary electron. Initially the electron performs a small-amplitude oscillation around the magnetic field. On encountering the moving reconnection site, it is accelerated and the gyro radius of its orbit increases. *Right*: The time evolution of the particle energy distribution function. Time increases from foreground to background. It is counted in plasma periods (inverse plasma frequencies). It can be seen that the initial Maxwell-distribution evolves into a final high-energy power-law distribution. The average electron energy increases by more than a factor of 10 in the 200 plasma periods when the final state is reached. For a plasma period of 0.001 seconds (corresponding to a density of 0.01 per cubic centimetre), this time is just 0.2 seconds (from Ref. 14).

lisionless plasma it is strictly forbidden for the particles to leave their magnetic flux tubes. Charged particles are "frozen" into the magnetic field. Reconnection should not then be possible.

Narrow current sheets as the site of reconnection

The way out of this dilemma is to realize that the transition regions between the approaching flux tubes are narrow current sheets like those in the magnetopause and tail of the magnetosphere shown in Figure 1. Their width is comparable to the ion gyration radius, roughly 100 km at the magnetopause and 1000 km in the magnetospheric tail. In these narrow sheets the ions behave as if there is no magnetic field and the motion of ions and magnetic flux tubes decouple. The multi-

spacecraft Cluster mission[12], one of the cornerstones of the ESA "Horizon 2000" Science Programme[13] initiated in 1984 by Roger Bonnet (then Director of the ESA Science Programme) has contributed substantially to the proof that the current sheets involved in reconnection are indeed narrow.

Galactic and extragalactic reconnection

In the Galaxy, the magnetic field strength is ~10^{-6} gauss, and the ion gyro radius of the galactic matter is 100–10 000 km. These scales are miniscule compared to galactic dimensions. Hence, under most conditions reconnection will be collisionless. What happens physically in reconnection can be studied in-situ solely in near-Earth space. Such studies must be accompanied by numerical simulations of the self-consistent motion of particles. In order to be applicable to galactic and extragalactic conditions, these have to be relativistic. Such simulations[14] have shown that charged particles are accelerated to ultra-relativistic energies in reconnection, and in a few hundred seconds develop power-law energy distributions like those observed in cosmic rays.

Individual reconnection sites cannot be resolved by observation from a distance. Very many of them can be placed into the volume of galactic and extragalactic radiation sources. Integrated over the volume, their effect is visible in the synchrotron radiation that is emitted by the accelerated electrons. The electron energies obtained in acceleration are high enough to explain, for instance, the synchrotron emission from variable galactic radio sources. In fact, only a small fraction of the volume needs to be filled in order to reproduce the radiation intensities emitted.

Parallel Electric Fields: The Auroral Paradigm

Celestial objects attract our attention through beautiful pictures. In contrast, the plasma phenomena in space are invisible. The only exception is the polar light, the "aurora", that is observed in the high-latitude upper atmosphere during disturbances of the geomagnetic field with its colourful, highly structured and variable appearance (see Fig. 3). In times past, humans attributed the aurora to fights of the goddesses in the sky. Later Christian generations saw in the aurora the shining of the candles held up by the Saints during processions in heaven[17].

The main problem with the aurora

In the 1950s it became clear that the auroral light is emitted at 100–400 km altitudes by electron beams with energies of kilo electron volts precipitating along the magnetic field into the upper atmosphere. There they hit the atmospheric atoms and stimulate them to emit light. The atomic excitation and optical emis-

Figure 3. A ground-based recording of an aurora. The folded braid of the intense light emission comes from altitudes between 100 and 400 km. The auroral bands are horizontally extended in latitudinal direction over long distances. In latitudinal direction, they are very narrow and consist of narrow rays. This narrow striation fine structure is along the straight but inclined magnetic field lines, indicating the very narrow filamentation of the magnetic-field-aligned auroral currents and the acceleration process actingat altitudes high above the aurora.

sion mechanisms in aurorae are very well understood and will not be addressed here. But what is the origin of the energetic electron beams?

Microscopic electric double layers

In the early 1970s, Hannes Alfvén[18] and others became convinced that electrical potential differences of kilovolts along Earth's magnetic field lines must exist when aurorae occur. These potential differences are caused by the auroral electric current flowing far above the visible aurora along the magnetic field into (and out of) the ionosphere, but why are there no current short-circuits with electric potential drops so large?

This puzzle was partially resolved when high-time-resolution plasma data became available from above the auroral region[19] at altitudes of 2500-7000 km. "Microscopic electric double layers" were detected, that is isolated stationary electric-field structures ~1 km in extent along the magnetic field in narrow ~10 km wide magnetic flux tubes. The width of these flux tubes is much larger

than the electron or ion gyration radii, and these microscopic double layers are thus nearly one-dimensional. They are organized in chains with the field-aligned distances between them a few times their individual extents. A large number of them fit into an auroral magnetic flux tube over the ~4000 km altitude range. Their individual small potential drops add up to the large auroral potential of several kilovolts that accelerates electrons and ions in opposite directions. This acceleration has been confirmed by in-situ observations[20] in the aurora. The narrow flux tubes are directly related to the ray-like striations in the optical aurora in Figure 3.

Aurora-like phenomena in the cosmos

Meanwhile, it has been realized that auroral phenomena are quite common in the Universe and we extrapolate from the *in-situ* studies in the near-Earth plasma that the microscopic electric double layers play a similar role to that in the Aurora Borealis. Aurorae have been detected on the magnetized planets of the Solar System (see Fig. 4). Similar phenomena occur in the solar atmosphere during flares and they have been generalized to magnetized stars and flare stars. One also expects that magnetized extrasolar planets exhibit aurorae[21], often much more violent than on the planets in the Solar System. In the strong converging magnetic fields in pulsar magnetospheres or in astrophysical jets that emanate from the centre of Active Galactic Nuclei (AGN), field-aligned electric currents will generate very large numbers of these remotely unobservable microscopic double layers. Their potential drops add up to enormous potentials, of the order of the energy of the engine that produces the jets. Such electric fields necessarily accelerate particles up to relativistic cosmic-ray energies.

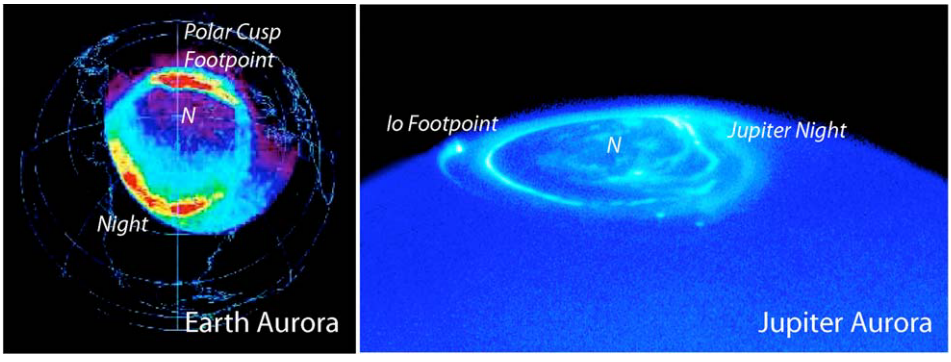


Figure 4. *Left*: A satellite recording of Earth's optical aurora showing the auroral oval around the magnetic pole and the two regions of simultaneous high auroral activity on the nightside and at the foot point of the dayside polar cusp. *Right*: A Hubble Space Telescope optical observation of the auroral oval on Jupiter. Similar to Earth the auroral oval encircles the entire polar region. On the nightside, it consists of several active rings. The bright "hot spot" to the left is the foot point of the magnetic flux tube that is locked at the tectonically active Jupiter moon Io.

Coherent radiation from aurora-like systems

As a by-product, the accelerated particles emit radiation in the strong magnetic field[22]. Earth and Jupiter emit radio waves of enormous intensity at kilometre and hectometre wavelengths, respectively, from their auroral regions. The output is many orders of magnitude more intense than incoherent synchrotron emission from the energetic particles could ever provide. In fact, the radiation mechanism resembles the well-known laser and maser effects. Analogous to lasers, the microscopic double layer electric field "pumps" the velocity distribution of the electrons in the plasma into an "excited state", which in this case is a deformed electron distribution function. In close similarity to the laser, the pumped-up electrons release their excess energy in concert in the form of coherent radiation.

A simple radiation mechanism like this should be realized in many strongly magnetized astrophysical systems like blazer jets, AGNs, and neutron-star mag-

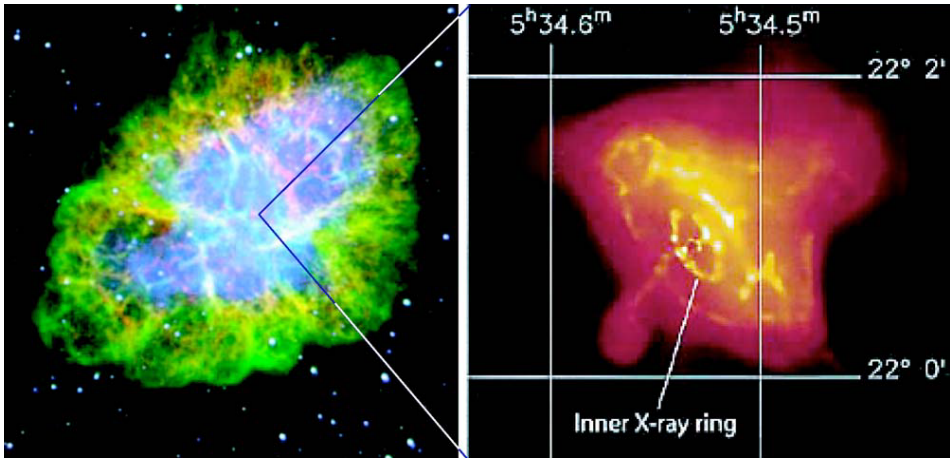


Figure 5. The remnant of the 1054 AD supernova observed by Chinese astronomers (the so-called Crab nebula) in optical (left) and X-ray (right) emissions. The optical (Hubble Space Telescope) view shows the entire Crab consisting of a rapidly expanding and highly turbulent supernova wind, which forms a whole network of shock waves. The bluish colouring towards the centre of Crab indicates the increasingly high temperatures of the gas therein. The X-ray image (from the Chandra spacecraft) on the right shows the most central part of Crab around the (invisible) Crab pulsar. Two narrow radiation jets escape along the rotation axis of the neutron star. In their outer parts, these jets are deformed by the interaction with the surrounding material. The invisible pulsar magnetosphere is surrounded by a magnetic torus of much larger radius containing a toroidal field, which maps into the torus-like radiation belt. The bright inner ring is the termination shock of the pulsar wind. Both the optical and X-ray radiation is (electron) synchrotron emission in the magnetic field of Crab. In the outer weak magnetic field regions, the emission is in the optical eV range, while in the strong-field inner part it is in the X-ray keV range.

netospheres. Radiation energies up to the energy range of gamma rays are expected there. The enormous amplification of the radiation in the presence of the microscopic double-layer electric fields implies that radiative cooling is very significant in systems exhibiting field-aligned potentials and relativistic particle acceleration. The decisive plasma processes proceed on microscopic scales far below the mean free path and many orders of magnitude below any resolvable astronomical scales. They are not accessible other than by analogy with the processes taking place in the aurora.

Collisionless Shocks: Lessons from the Bow Shock

Shocks as a general phenomenon in the cosmos

One encounters shocks everywhere in astrophysical systems: in supernova remnants, in colliding galaxies, in astrophysical jets, as termination shocks near the light cone in pulsar winds, and so on. As an example, Figure 5 (left) shows an optical picture of the Crab nebula, the remnant of a supernova explosion observed by Chinese astronomers in 1054 AD. It appears to consist of a network of intensely radiating shocks, which are produced in the interaction between the relativistic supernova wind and the surrounding galactic gas. The X-ray plot of the central part of Crab on the right shows the vicinity of the (invisible) Crab pulsar with its X-ray jets ejected along the axis of rotation and the ring-like termination shock at the light cylinder. All of these shocks are collisionless[23] in the sense that their widths are just a few ion gyro radii thick, which is much less than the mean free path. The fact that they are visible in electromagnetic radiation implies that they generate high-energy electrons which emit synchrotron radiation, and that magnetic fields are involved – a clear indication of the collisionless nature of the shocks. In contrast, gas-dynamic shocks are dominated by collisions. Their widths are comparable to the mean free path, and the radiation they emit is thermal.

Shocks accessible for direct observation include the bow shocks of the magnetized Solar System planets, comets, travelling interplanetary shocks, and the heliospheric termination shock shown in Figure 1.

Earth's bow shock as the best-investigated shock

Earth's bow shock[24] is the most accessible and best-studied of them all. It is generated in the interaction between the solar wind and the magnetosphere, in which the ions are supersonic while the electrons remain subsonic. Its width is of the order of one ion gyro radius, only ~1000 km thick. A summary of observational results in and near Earth's bow shock is given in Figure 6. The bow shock is a quasi-stationary phenomenon standing at sunward distances of 12 to

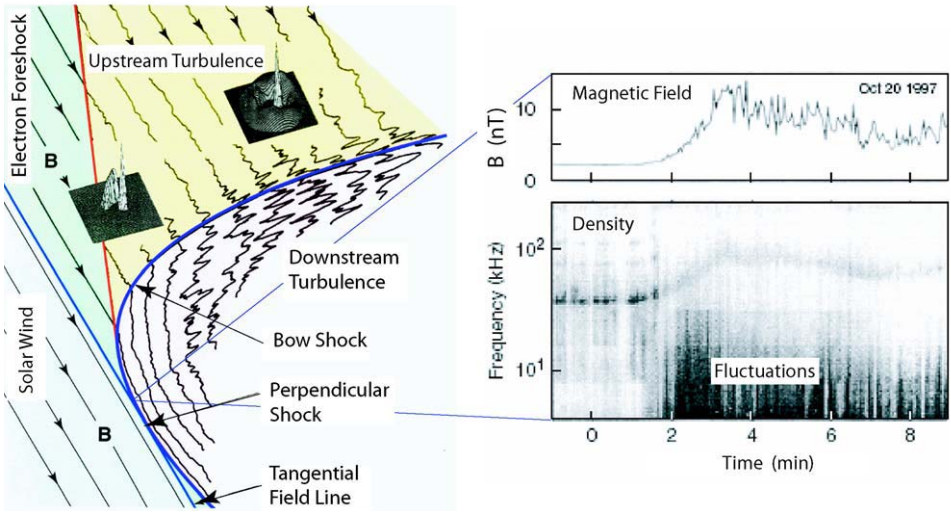


Figure 6. *Left*: Geometry of Earth's bow shock with the electron and ion foreshocks. The small inserts show the evolution of the shock-reflected ion beam from the narrow beam distribution into a ring of diffuse heated ions. The steep peak in the centre of each insert is the cold and fast inflowing supersonic plasma stream. The turbulent upstream magnetic field region is indicated (from Ref. 15, Chapter 62). Moreover, it can be seen that the bow shock is curved like a paraboloid. It consists of two parts: the region on the lower left where the magnetic field is tangential to the shock – this is traditionally called the quasi-perpendicular shock – and the remaining regions where the magnetic field has a large angle to the shock – the quasi-parallel shock. *Right*: The magnetic field amplitude across the lower left quasi-perpendicular part of the shock showing the field increase in the shock ramp (top) and the electric fluctuation spectrum across the shock (below). The plasma emission line at high frequencies maps the shock density profile. The intense low-frequency fluctuations in the shock ramp and foot are caused by a multitude of localized small-scale electric-field structures, so-called electron and ion holes (from Ref. 25).

14 Earth radii from Earth's centre. Depending on the direction of the magnetic field with respect to the collisionless shock, one can distinguish "quasi-perpendicular" and "quasi-parallel" shocks. The bow shock allows both types to be investigated because its shape is that of a bent parabolic shield in front of the magnetosphere.

The true shock transition region

Because of the absence of dissipation in collisionless plasmas, shock formation relies on the reflection of a substantial part of the inflowing plasma at the shock front. The larger the Mach number, the more particles will be reflected. At a Mach number smaller than 10, the bow shock reflects only a few percent of the solar wind. The reflected particles escape along the magnetic field upstream into the solar wind. They excite oscillations of the magnetic field in the region

upstream of the shock until this region becomes highly turbulent. The turbulence scatters the reflected particle beams and retards the solar wind stream. Thus the true shock transition is the entire turbulent region, not just the narrow shock front. The spatial extent of both together is much larger than the width of the shock front, while still remaining much smaller than the mean free path – in the case of the bow shock just some ten Earth radii.

Relativistic shocks in the cosmos

The shocks in the very-high-Mach-number relativistic plasmas in supernova remnants, galaxies and clusters of galaxies necessarily reflect a very large percentage of the inflowing plasma back upstream, and create an extended upstream turbulent transition region. This transition region appears broad enough to become visible in the synchrotron radiation emitted by the scattered relativistic electrons.

Shocks are also produced in encounters of plasma shells in collimated jets, which are ejected from pulsars and AGNs, and in relativistic winds. Such an interaction is similar to the case of travelling shocks in the solar wind which are observed in corotating interaction regions[26]. The shocks evolve locally as current sheets where the shells interact, and are accompanied by strong large-scale magnetic fields which confine the particles. The particles are forced to oscillate back and forth until they have been accelerated via the first-order Fermi mechanism to such high speeds that their gyro radii exceed the width of the shock. The resulting particle distributions will again exhibit power-law tails[27], which map to power-law spectra in the emitted synchrotron radiation. The centre of the Crab nebula (shown on the right in Fig. 5) is an example. Here the torus-like shock magnetic field formed in the relativistic wind and the pulsar jets appear in the X-ray radiation with energies of several kilo electron volts, which is the signature of shock-accelerated relativistic electrons.

Conclusions

These few examples of fundamental processes in collisionless plasmas allow us to conclude that several important astrophysical phenomena will be understood only when referring to microscopic scales far below the mean free path. Not unexpectedly, the interesting physics takes place on small spatial and temporal scales far below observational resolution. Fortunately, space plasma physics provides the unique opportunity to study processes of this kind in-situ and in real time in the accessible near-Earth space and the heliosphere. These are the only places in the entire Universe where humans can investigate processes in collisionless plasmas experimentally and in detail and infer their macroscopic effects.

References

1. See, e.g., the article of L. Colangeli in this book.
2. S. Chapman & V.C.A. Ferraro, *Terrestr. Magnetism Atmospher., Electricity* **36**, 77 and 171, 1931; *ditto*. **37**, 147, 1932 were the first to speculate about the occasional presence of plasma streams in interplanetary space. They suggested that these streams when encountering Earth's magnetic field generate large-amplitude geomagnetic variations.
3. L. Biermann, *Z. Astrophys*., **29**, 274, 1951, concluded from the persistent directions of cometary tails radially away from the Sun that the Sun emitted a continuous plasma wind blowing against the comets.
4. Cf. the articles of A. Balogh & V. Izmodenov, and B. Hultqvist *et al*. in this book.
5. T. Encrenaz, R. Kallenbach, T.C. Owen & C. Sotin (Eds.), The Outer Planets, SSSI Vol 19, Springer, Dordrecht, 2005, and *Space Sci. Rev*., **116**, Nos. 1-2, 2005, in particular Chapter 3: Aurorae and Magnetospheres.
6. M. Blanc, R. Kallenbach & N.V. Erkaev, in Ref. 5, p. 227; M. G. Kivelson, *ditto*., p. 299.
7. T. Gold, *Nature,* **218**, 731, 1968, *ditto*. **221**, 25, 1969; F. Pacini, *ditto*. **216**, 567, 1967, *ditto*. **219**, 145, 1968; R.A. Hoyle, J. Narlikar & J.A. Wheeler, *ditto*. **203**, 914, 1964.
8. R.D. Blandford & M.J. Rees, *Monthly Not. RAS,* **169**, 395, 1974; R.D. Blandford & R.L. Znajek, *ditto*. **179**, 433, 1077.
9. G. Paschmann *et al*., *Nature,* **282**, 243, 1979.
10. For a collection of reviews on this subject see B. Hultqvist and M. Øieroset (Eds.), Transport across the Boundaries of the Magnetosphere, SSSI Vol. 2, Kluwer Academic Publ., Dordrecht, 1997, and *Space Sci. Rev*., **80**, Nos. 1-2, 1997; B. Hultqvist, M. Øieroset, G. Paschmann & R. Treumann (Eds.), Magnetospheric Plasma Sources and Losses, SSSI Vol. 6, Kluwer Academic Publ., Dordrecht, 1999, and *Space Sci. Rev.,* **88**, Nos. 1-2, 1999.
11. Cf. Ref. 12, Part III; M. Fujimoto *et al*., *Geophys. Res. Lett*., **24**, 2893, 1997; T. Nagai *et al*., *J. Geophys. Res*., **106**, 25929, 2001; M. Øieroset *et al*., *Nature,* **412**, 414, 2001; F.S.Mozer *et al*., *Phys. Rev. Lett*., **89**, 015002, 2002; M. Øieroset *et al*., *Phys. Rev. Lett*., **89**, 195001, 2002.
12. G. Paschmann, S. Schwartz, P. Escoubet & S. Haaland (Eds.), Dayside Magnetospheric Boundaries: Cluster Results, SSSI Vol. 20, Kluwer Academic Publ., Dordrecht, 2005, and *Space Sci. Rev.* (in press) 2005.
13. Space Science Horizon 2000, ESA-SP 1070, ESA Publications Division, Noordwijk, Holland 1984; Horizon 2000 Plus, ESA-SP 1180, ESA Publications Division, Noordwijk, Holland 1995.
14. C.H. Jaroschek, R.A. Treumann, H. Lesch & M. Scholer, *Phys. Plasmas,* **11**, 1151, 2004; C.H. Jaroschek, H. Lesch & R.A.Treumann, *Astrophys. J*., **605**, L9, 2004.
15. J.A.M. Bleeker, J. Geiss & M.C.E. Huber (Eds.), The Century of Space Science, Kluwer Academic Publ., Dordrecht, 2003.
16. Cf. Ref. 15, Vol. 2, Chapter 62, p. 1495.
17. S.-I. Akasofu, Majestic Lights - The Aurora Borealis, University of Alaska Press, Fairbanks, 1988.

18. H. Alfvén & C.-G Fälthammar, Cosmical Electrodynamics, 2nd Edition, Clarendon Press, Oxford, 1963; H. Alfvén, *Space Sci. Rev.*, **7**, 140, 1967; L.P. Block, *ditto*, 198, 1987.
19. See Chapter 4 in: G.Paschmann, S. Haaland & R. Treumann (Eds.), Auroral Plasma Physics, SSSI Vol. 15, Kluwer Academic Publ., Dordrecht, 2002, and *Space Sci. Rev.,* **103**, Nos. 1-4, 2002; R.E. Ergun *et al*., *Geophys. Res. Lett*., **25**, 2025, 1998.
20. Cf. Ref. 19, C.W. Carlson *et al*., *Geophys. Res. Lett*., **25**, 2017, 1998.
21. P. Zarka, R.A. Treumann, B.P. Ryabov & V.B. Ryabov, *Astrophys. Space Sci.,* **277**, 293, 2001.
22. Suggested first by R.E. Ergun *et al*., *Astrophys. J*., **538**, 456, 2000; for a review of the astrophysical implications of Maser and Laser emissions, see R.A. Treumann, *Rev. Astrophys. Astron*., submitted, 2005; first estimates for Blazar emissions are given in M.C. Begelman, R.E. Ergun & M.J. Rees, astro-ph/0502151v1-7Feb2005.
23. See Ref. 12, Part II.
24. For the first theoretical attempt to describe collisionless quasi-laminar shocks see R.Z. Sagdeev, *Rev. Plasma Phys*. **4**, 23, 1966. This was followed by a theory on magnetized shocks by C.F. Kennel & R.Z. Sagdeev, *J. Geophys. Res*. **72**, 3303, 1967; a more recent discussion of simulations of collisionless shocks is found in B. Lembège *et al*., *Space Sci. Rev*. **110**, 161, 2004, resulting from an ISSI Team effort.
25. S.D. Bale *et al*., in Ref. 12, Part II; cf. also S.D. Bale *et al*., *Astrophys. J*., **575**, L25, 2002.
26. See the collection of papers in: A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach & H. Kunow (Eds.), Co-rotating Interaction Regions, SSSI Vol. 7, Kluwer Academic Publ., Dordrecht, 1999, and *Space Sci. Rev*., **83**, Nos. 1-4, 1999.
27. C.B. Hededal, T. Haugbolle, J.T. Frederiksen & Å. Nordlund, *Astrophys. J*., **617**, L107, 2004; C.H. Jaroschek, H. Lesch & R.A. Treumann, *Phys. Plasmas,* **11**, 1151, 2004; *ditto*., *Astrophys. J*., **605**, L9, 2004.
28. We acknowledge very much the constructive criticisms and contributions of Johannes Geiss.

The Role of Laboratory Experiments in Characterizing Cosmic Materials

L. Colangeli

INAF, Astronomical Observatory of Capodimonte, Naples, Italy

Why Laboratory Experiments?

Astronomical observations are the classical way to investigate the properties of materials populating circumstellar and interstellar media. The advent of new space-borne observatories (such as the Hubble Space Telescope, the Infrared Space Observatory and Spitzer, and the future Planck-Herschel mission), together with the continuous progress of ground-based telescope capabilities (e.g. the 8 m class telescopes, such as the VLT), have provided major tools to increase the quantity and quality (in terms, for instance, of sensitivity and spectral resolution) of information about cosmic dust and ices. In the field of Solar System exploration, most valuable data come from *remote observations*, obtained from instrumentation onboard spacecraft getting close to the target: planets or small bodies. This has been the case, for instance, for various NASA missions and the first European planetary mission, Mars Express, orbiting Mars and for flyby missions to comets, such as Giotto to 1P/Halley. There are currently many new and exciting appointments with several other targets: the Cassini-Huygens spacecraft is already exploring the Saturn system and Titan in particular, the Rosetta mission is on its way to a rendezvous with comet 46P/Churyumov-Gerasimenko in 2014, Venus Express will study Venus, while the NASA Messenger and, later, the ESA BepiColombo missions will make a detailed study of Mercury. But the future of Solar System exploration is oriented towards complementing remote with *in-situ measurements*, especially in cases such as Mars, where direct data are required to solve specific questions, e.g. those related to the search for exo-biological signatures. One of the most ambitious goals of future Solar System exploration is the *return of samples* from Mars and other primordial bodies for detailed laboratory analyses.

The previous scenario clearly indicates that sets of new and exciting data are and will be continuously available to constrain models of the formation and evolution of cosmic matter, from interstellar medium to planetary systems, similar to that hosting our planet Earth. In this framework, it is essential to study the physical, chemical and geological properties of materials populating/constituting various space environments.

Laboratory astrophysics is an interdisciplinary research field that has gained a more and more prominent role in modern astrophysics. In fact, it is nowadays well-recognised that laboratory experiments are essential for a thorough interpretation of astronomical data. The experimental approach for studying cosmic materials is based on the production of *analogues* with physical and chemical properties suitable for reproducing, as far as possible, cosmic matter. The next step consists of analysis of the samples using a combination of various techniques. Spectroscopy is the most used tool for performing direct comparisons with astronomical data. However, the characterisation of the physical and chemical properties of compounds responsible for observed features relies on the correlation of spectroscopic and other laboratory results from the analysed samples. It is worth stressing that an interdisciplinary approach is required in this case, where competences from such fields as astrophysics, geology, geophysics, solid-state physics and optics must be combined. Experiments have recently been devoted to addressing the effects on materials of *processes* active in space, such as thermal annealing, UV irradiation, ion bombardment and gas-solid interactions. In fact, it is nowadays clear that matter in space experiences a continuous (cyclic) evolution. Thus, an integrated and self-consistent approach to modelling evolution in space must be applied to identify materials capable of reproducing observations and, at the same time, be compatible with conditions (e.g. thermal environment, UV-radiation and ion-bombardment doses, solid-gas processing) typical of different environments.

The role of laboratory experiments in the characterization of cosmic material has been the subject of activities of an international team of scientists that has operated at ISSI in the period 1999 - 2000. One of the tangible outputs of the team activities has been a paper which appeared in *Astronomy and Astrophysics Reviews* in 2003[1].

Astrophysical Materials

Three main classes of solids are relevant in processes occurring in space: ices, carbon-based materials and silicates. Their chemical composition has been identified, mainly thanks to absorption or emission spectroscopy in the infrared range, where fundamental vibrations of molecules fall.

Table 1 summarises interstellar ice molecules identified by their infrared absorption features[2]. Water ice is the most abundant interstellar ice, at 1×10^{-4} abundance with respect to the total H column abundance[3], followed by CO_2. The abundance of solid CO_2 is 15-20 % relative to H_2O ice, towards field stars and low-mass protostellar objects, whereas the range is 15-40 % towards high-mass

Molecule	Infrared bands (μm)
H_2O	2.96, 3.07, 3.2 – 3.7, 4.5, 6.0, 12, 44
CO_2 / $^{13}CO_2$	2.70, 2.78, 4.27, 15.2 / 4.38
CO / ^{13}CO	4.67 / 4.78
OCS	4.92
H_2CO	5.83
HCOOH	5.83, 7.25
CH_4	3.32, 7.67
CH_3OH	2.27, 3.54, 3.85, 3.94, 4.1, 6.85, 8.9, 9.7
SO_2 ?	7.60
NH_3 ?	2.96, 3.2 - 3.7, 3.47, 9.01
$HCOO^-$?	7.41
OCN^- ?	4.62

Table 1. Interstellar ices and their absorption fingerprints[2]

protostars[4]. Solid methanol is abundant towards high-mass protostars (up to 35% relative to H_2O ice), while it is almost absent (< 3% relative to H_2O) towards field stars and low-mass protostars. Up to now, the detected icy species are relatively simple molecules formed only by H, C, N, O and S. But are all other atoms embedded in the refractory dust component, or is there a chance that minor fractions of them are incorporated in the icy component?

Carbon-based materials identified or potentially present in space span from molecular species to solid grains, and cover a wide range of physical, chemical and structural properties. Polycyclic Aromatic Hydrocarbon (PAH) molecules are considered the prime carriers of the infrared bands observed in emission from dusty regions exposed to intense ultraviolet radiation[5]. Bands at 3.3, 6.2, 7.7, 8.6 and 11.2 μm, accompanied by other minor bands, are observed from many different astronomical objects[6]. Carbon grains residing in interstellar clouds with different densities are responsible for the wide UV-extinction bump falling at 217 nm[7]. Today, amorphous grains composed of aromatic units seem the most likely candidates for interpreting this feature[8]. C-H aliphatic molecular resonances in hydrogenated carbon grains are considered the origin of the 3.4 μm (3.38, 3.41, and 3.48 μm triplet) absorption band observed in the diffuse interstellar medium, but lacking in dense clouds spectra[9].

Silicates are detected in the spectra of a variety of dusty sources[10] by the typical infrared bands at 10 and 20 μm. These bands are due to Si-O stretching and O-Si-O bending modes in SiO_4 tetrahedra, the building blocks of silicate com-

pounds. The classes of silicates of major relevance in the astrophysical context are olivines, $(Mg_xFe_{1-x})_2SiO_4$, with end-members forsterite (x = 1) and fayalite (x = 1), and pyroxenes, $(Mg_xFe_{1-x})SiO_3$, with end-members enstatite (x = 1) and ferrosilite (x = 0). Other cations (e.g. Ca, Mg, Al) can be included in the chemical structure[11]. The status of silicate dust is mainly amorphous in the interstellar medium, but a non-negligible crystalline component is detected around young, evolved and post-AGB stars, in planetary nebulae, in massive stars and in comets[1].

Species other than ices, carbon and silicates are present in space to a minor extent, such as metal oxides, carbides, sulphides and carbonates[12,13].

The Laboratory Experiments

Strong progress has been made in recent years by several laboratories operating around the World in covering the complementary aspects of production, characterisation and processing of analogues of cosmic compounds.

The simplest way to produce samples for laboratory analyses is to use terrestrial natural minerals, which can be studied as bulk rocks or ground up and selected in size. This approach is applicable to simulate planetary surfaces and regolith, provided that the chemical compositions of the selected minerals are of interest due to their similarities to other planetary materials. As an example, a selection of natural terrestrial minerals and rocks potentially relevant as Martian analogues is reported in Table 2.

Class of materials	Representative members
Carbonates	Calcite, Dolomite, Siderite, Magnesite, Aragonite
Clays	Kaolinite, Montmorillonite, Smectite, Nontronite, Schwertmanite, Ferrihydrite
Feldspars	Albite, Anorthite, Labradorite, Orthoclases
Hydrous Carbonates	Artinite, Gaylussite, Hydromagnesite, Dypingite
Igneous Rocks	Andesite, Basalt, Palagonite
Iron Oxides	Hematite, Ilmenite, Chromite, Magnetite, Lepidocrocite, Goethite
Nitrates	Niter, Nitratine, Nitrocalcite, Nitromagnesite
Olivines	Fayalite, Forsterite
Phosphates	Merillite, Whitlockite, Apatite
Pyroxenes	Enstatite, Augite, Pigeonite
Sulphates	Gypsum, Jarosite, Kiéserite, Anhydrite
Others	Quartz, Maghemite, Pyrite, amorphous Fe-rich clay

Table 2. Selection of terrestrial mineral analogues of Martian soil

The previous approach is not applicable to interstellar dust analogues, both because relevant dimensions (typically micron/sub-micron) cannot easily be obtained by grinding of rocks and because the chemical composition and structure of terrestrial samples are often not suitable to reproduce cosmic grains. In this case, production methods are based on vaporisation and subsequent condensation of pure materials. A wide variety of carbon- and silicon-based *smokes* are obtained by vapour condensation. This technique has provided a powerful tool to investigate the processes of dust formation and evolution[1]. Laser bombardment of homogeneous targets or mixtures of different targets, arc discharges between carbon or graphite electrodes, and laser pyrolysis in a gas flow are used to vaporise materials. Cooling of the gas-phase mixture gives the formation of molecular clusters, which grow to solid particles. Alternative sol-gel chemical processes can be used to produce a variety of cosmic analogues. Besides olivines and pyroxenes, a wide variety of silicates can be obtained by changing the relative abundance of cations (e.g. Mg^{2+}, Fe^{2+}, Al^{3+}, Ca^{2+}) in the original reaction mixture. The selection of formation conditions allows the tuning of the composition and structure of the products.

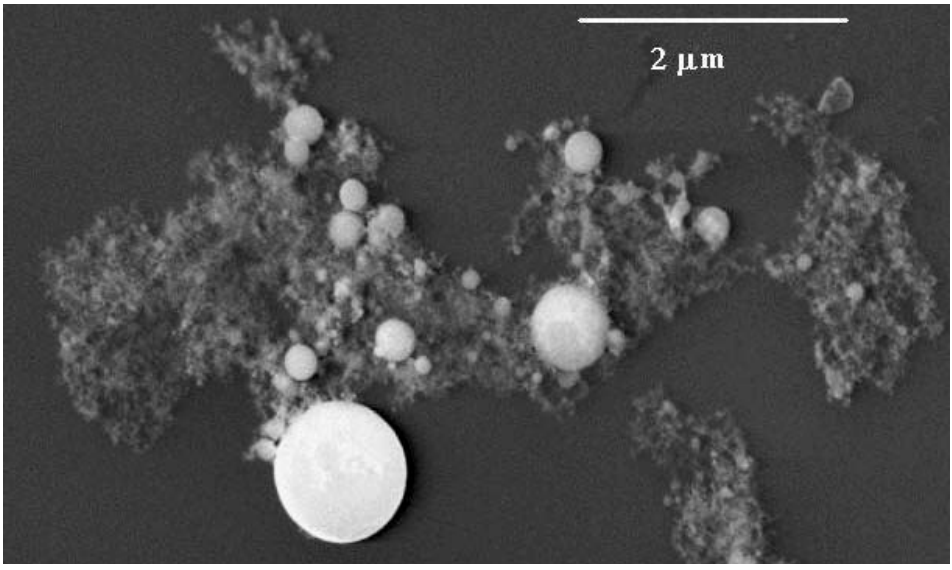


Figure 1. SEM image of amorphous fayalite grains produced by laser evaporation

A thorough characterisation of the samples requires the use of several analytical techniques[1,14]. The morphology and structural order of solid samples at nanometre and sub-nanometre scales are studied by scanning, transmission, and analytical electron microscopy (Fig. 1). X-ray absorption spectroscopy (XAS), extended X-ray absorption fine structure (EXAFS), and X-ray absorption near

edge structure (XANES) techniques are used to determine the sample's crystal structure. Elemental composition can be determined by analysis of dispersed X-rays. Raman spectroscopy is also used to analyse the structure of (especially carbon–based) materials. However, spectroscopy remains the most powerful tool for investigating the different aspects of material properties, depending on wavelengths of light. In fact, electronic transitions of solids fall in the vacuum ultraviolet, while the electronic gap is in the visible region, allowing the identification of the conduction properties of materials. Molecular vibration resonances generally fall in the mid-infrared range, while material structure and morphology drive the spectral behaviour in the far-IR region.

Another key step in terms of laboratory experiments is the study of the reactivity of materials to processes active in space. Thermal annealing occurs from stellar outflows (up to 1000 K) to the pre-solar nebula (up to about 100 K) environments. By varying maximum temperature and time of processing in laboratory experiments[15-17], physical quantities are derived, such as the activation energy, E_a, describing thermal conditions required to produce the amorphous-to-crystalline transition (Fig. 2). UV irradiation, with a typical dose of 3×10^{23} eV cm^{-2} during the $\sim 3 \times 10^7$ years of residence in the interstellar medium, certainly influences cosmic-grain chemical and structural properties, while ion bombardment processes grains at different doses in the interstellar medium (3×10^3 eV mol^{-1}), pre-cometary phase (10^6 eV mol^{-1}), comets (6×10^2 eV mol^{-1}) and interplanetary medium (10 - 100 eV mol^{-1} @ 100 keV, 10^5 - 10^6 eV mol^{-1} @ 1 keV). Experiments to reproduce both effects on carbon grains have been performed in the laboratory[18,19]. Finally, interaction of dust with gas (mainly hydrogen) is a very important process that has been tested in the laboratory[20].

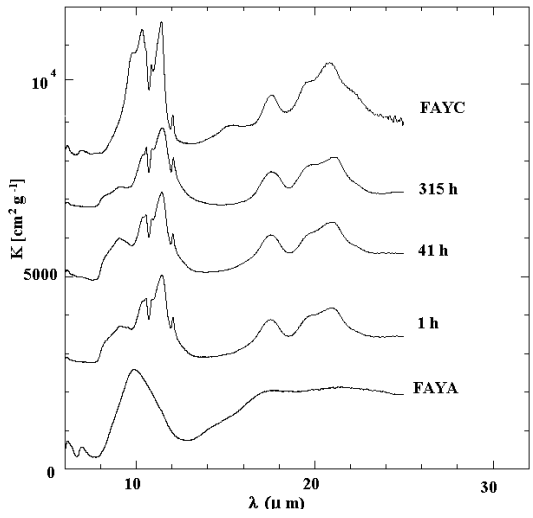


Figure 2. IR absorption coefficient of amorphous fayalite (FAYA) grains after thermal annealing at 800 °C for different times[17]. The spectrum of crystalline fayalite (FAYC) grains is also shown.

Application to Key Astrophysical Problems

The key role of laboratory experiments is clearly demonstrated by several astrophysical questions, where possible interpretations have been obtained only by means of experimental results. Here we will discuss just two examples.

The presence of crystalline silicates in comets is nowadays well demonstrated by IR sharp emission features observed, in particular, thanks to ISO on comet Hale-Bopp C/1995 O1[21]. The major peaks are well fitted by laboratory data for crystalline Mg-rich olivine (forsterite) and pyroxene (enstatite)[22,23] (Fig. 3). On the other hand, as mentioned above, silicates are mainly amorphous in the interstellar medium. Thus, an efficient crystallisation mechanism is required in the early stages of proto-solar nebula evolution. Moreover, according to laboratory results, 10^6 years of annealing above ~800 K are required for amorphous-to-crystalline transformations to occur. The temperature in the outer nebula, where comets should have been formed, was too low for this to happen. Thus, some subsequent thermal processing at high temperatures is required. Turbulent radial mixing in the solar nebula[24] and/or annealing of dust by nebular shocks[25] have been invoked as possible mechanisms. Several uncertainties still remain, which could receive clarification from laboratory experiments addressing, for example,

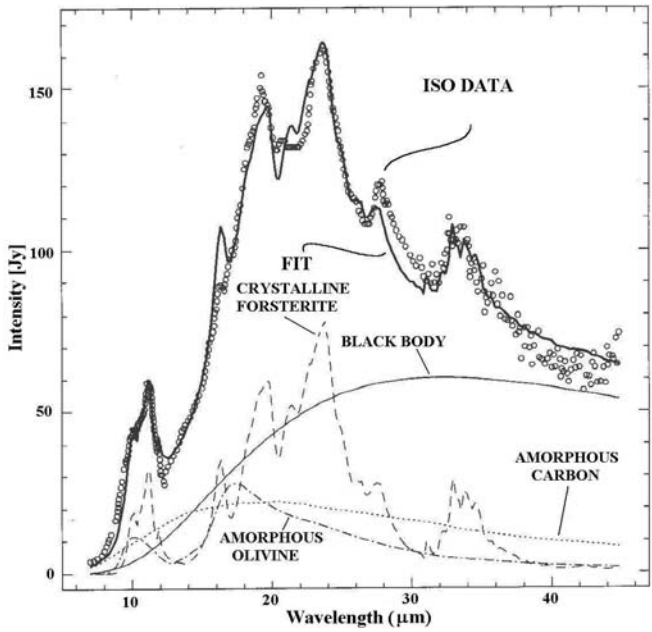


Figure 3. Fit to Hale-Bopp ISO data[21] with the optical properties of amorphous and crystalline materials measured in the laboratory[22].

flash-heating at 1100 K of micrometre-sized particles, which has been proposed to happen for precursors of meteoritic chondrules[26]: a few minutes could be sufficient to crystallise amorphous silicates.

An interesting and topical problem regarding carbon evolution in space concerns the 3.38, 3.41, and 3.48 μm absorption bands mentioned above. While they are observed in the diffuse interstellar medium and in the proto-planetary nebula CRL 618[9,27], they are lacking in the dense interstellar medium. The most plausible interpretation is based on the following laboratory results: the interaction of amorphous carbon grains with H atoms (hydrogenation) produces the appearance of a neat aliphatic 3.4 μm band[20], while the band disappears when hydrogenated carbon grains are exposed to UV irradiation (Lyman emission)[18], even in the presence of an ice coating on the grains[28], or to ion bombardment[29]. When formation/destruction rates measured in the laboratory are compared with typical doses in space, it turns out that the aliphatic C-H bond formation by H atoms reacting on carbon grains prevails over destruction by UV irradiation, under typical diffuse interstellar conditions. Carbon grains in dense clouds are coated by an ice mantle, so that the carbon core is shielded from interaction with H atoms, while C-H bonds can be destroyed by penetrating UV photons and/or ions. This scenario is fully compatible with presently available observational results.

The Future

The importance of laboratory work is clearly demonstrated by the examples reported above. Despite the important results obtained so far, further experiments are needed to contribute to the interpretation of observations. The variety of different materials of potential relevance for space applications is so wide that only long-term and systematic studies can provide the results required to interpret observational data. With the advent of more sophisticated observational tools (e.g. greater spectral range and better resolution), laboratory experiments must also be improved to cope with the new requirements. Moreover, models of the evolution of cosmic matter in different environments, from the interstellar medium to planetary surfaces, require further guidelines concerning the role of volatile and refractory materials at different stages and after different processes.

The genesis of silicates around stars and their amorphous-to-crystalline (and, possibly, inverse) transformation is still a matter of uncertainty as regards, for example, the possible existence of metastable states (produced, for example, by ion irradiation) from which transitions to a crystalline state require very little energy. This is an area where laboratory experiments can tell us much. As far as carbon is concerned, the relations between solid and molecular phases still

require in-depth investigation, whereby the laboratory will certainly play a role in understanding, for example, how large molecules arrange to form grains, and/or what kind of molecules are produced by destructive processes of solid particles. Last but not least, regarding ices, a general question is their formation mechanism, which may proceed either via grain surface chemistry or by accretion of simple molecules with subsequent processing by cosmic rays, UV photolysis, or a combination of both. Only grain-surface-reaction laboratory experiments may elucidate which molecules are formed by the reaction of atoms and molecules on grain surfaces. Important key molecules to be studied are CO_2 (formed by the reaction CO+O) and CH_3OH (formed by hydrogenation of CO). Crucial parameters can be derived from experiments, such as reaction rates and activation barriers.

In conclusion, the key role of laboratory experiments in the space sciences is reaffirmed by the very many areas in which experimental results are needed. The now-recognised importance of the laboratory work is well-evidenced by the organisation of working groups aiming at dedicated laboratory studies to prepare for the interpretation of results from major observation projects (e.g. Planck, Herschel) and space missions (e.g. Rosetta, Cassini, Mars missions) aimed at observing/visiting cosmic bodies[30].

References

1. L. Colangeli *et al*., *Astron. Astrophys. Rev*., **11**, 97, 2003.
2. A.C.A. Boogert & P. Ehrenfreund, in A.N. Witt, G.C. Clayton and B.T. Draine (Eds.), Astrophysics of dust, ASP Conference Series Vol. 309, p. 547, 2004.
3. A.G.G.M. Tielens, A.T. Tokunaga, T.R. Geballe & F. Baas, *Astrophys. J.,* **381**, 181, 1991.
4. P.A. Gerakines *et al*., *Astrophys. J.,* **522**, 357, 1999.
5. D.M. Hudgins & L.J. Allamandola, in Ref. 1, p. 665, 2004.
6. P. Cox & M.F. Kessler (Eds.), The Universe as Seen by ISO, ESA-SP **427**, Vol. I, II, The Netherlands, 1999.
7. E.L. Fitzpatrick, in Ref. 1, p. 33, 2004.
8. V. Mennella, L. Colangeli, E. Bussoletti, P. Palumbo & A. Rotundi, *Astrophys. J.,* **507**, L177, 1998.
9. Y.J. Pendleton, in Ref. 1, p. 573, 2004.
10. J.S. Mathis, *Ann. Rev. Astron. Astrophys*., **28**, 37, 1990.
11. W.A. Deer, J. Zussman & R.A. Howie, Rock-forming minerals, Vol. 1-5, The Geological Society, London 1997.
12. T. Henning & H. Mutschke, *Spectrochim. Acta A,* **57**, 815, 2001.
13. T. Henning & M. Schnaiter, in P. Ehrenfreund, H. Kochan, C. Krafft & V. Pirronello (Eds.), Laboratory astrophysics and space research, Kluwer, Dordrecht, p. 249, 1998.

14. A. Rotundi *et al*., *Met. Planet. Sci*., **37**, 1623, 2002.
15. V. Mennella *et al*., *Astrophys. J.,* **444**, 288, 1995.
16. S.L. Hallenbeck, J.A. Nuth III & R.N. Nelson, *Astrophys. J.,* **535**, 247, 2000.
17. J.R. Brucato, V. Mennella, L. Colangeli, A. Rotundi & P. Palumbo, *Planet. Space Sci*., **50***,* 829, 2002.
18. V. Mennella *et al*., *Astrophys. J.,* **464**, L191, 1996.
19. V. Mennella *et al*., *Astrophys. J.,* **481**, 545, 1997.
20. V. Mennella, J.R. Brucato, L. Colangeli & P. Palumbo, *Astrophys. J.,* **524**, L72, 1999.
21. J. Crovisier *et al., Science,* **275***,* 1904, 1997.
22. J.R. Brucato, L. Colangeli, V. Mennella, P. Palumbo & E. Bussoletti, *Planet. Space Sci.,* **47***,* 773, 1999.
23. D.H. Wooden *et al*., *Astrophys. J.,* **517**, 1034, 1999.
24. D. Bockelée-Morvan, D. Gautier, F. Hersant, J.M. Huré & F. Robert, *Astron. Astrophys.,* **384**, 1107, 2002.
25. D.E. Harker & S.J. Desch, *Astrophys. J.,* **565**, L109, 2002.
26. F.J.M. Rietmeijer, in J.J. Papike (Ed.), Planetary Materials, *Rev. Mineralogy,* **36***,* 2-1, 1998.
27. J.E. Chiar, Y.J. Pendleton, T.R. Geballe & A.G.G.M. Tielens, *Astrophys. J.,* **507**, 281, 1998.
28. V. Mennella *et al*., *Astron. Astrophys*., **367**, 355, 2001.
29. V. Mennella *et al., Astrophys. J.,* **481**, 545, 1997.
30. The experimental work at INAF – Osservatorio Astronomico di Capodimonte is supported by ASI (Agenzia Spaziale Italiana), MIUR (Ministero Universita e Ricerca) and INAF (Istituto Nazionale di Astrofisica).

Part B

Evolution of Matter in the Universe

J. Geiss[a] and G. Gloeckler[b]

[a]*International Space Science Institute, Bern, Switzerland*
[b]*Department of Physics and IPST, University of Maryland, College Park, Maryland, USA*

Matter and energy content of the Universe control its geometry and expansion. In the early Universe, the density was dominated by relativistic particles. At a cosmic time of a second these were photons, neutrinos and electron pairs. Their influence on the expansion has become negligible in the present epoch, and now baryons, non-baryonic "dark matter" and "dark energy" dominate the large-scale dynamics and geometry of the Universe (Fig. 1). Baryons are the well-known constituents of ordinary matter. For the existence of the other two components, we have only indirect, but increasingly compelling evidence.

Although the influence of baryons on the overall dynamics and geometry of the present Universe is relatively minor, their physical properties are unique. Among the major forms of matter and energy that populate the present Universe, only baryonic matter participates in all the physical forces known to us: the strong forces (transmitted by gluons), the electromagnetic forces (transmitted by photons), the weak forces (transmitted by the $W^{\pm}$ and Z^0 bosons), and gravity (transmitted by gravitons). These four physical interactions enable baryons to self-organize, form a multitude of microscopic and macroscopic structures and, indeed, create all the variety and beauty that we observe in the World.

The Expanding Universe

When in 1916, Albert Einstein formulated the theory of General Relativity, space and time became objects of physical enquiry, along with matter and radiation[1]. For the first time this allowed development of scientific cosmologies that made predictions that could be compared with observations. Thus General Relativity provides the framework for cosmological models, but matter and energy content determine which model is valid for our Universe. A variety of observations made during the last fifty years show that we live in a Universe that expanded out of an extremely dense and hot phase[2]. We depict, in Figure 1, the matter and energy components prevailing in the cosmos at three selected times. Not only did the cosmic density decrease by many orders of magnitude during

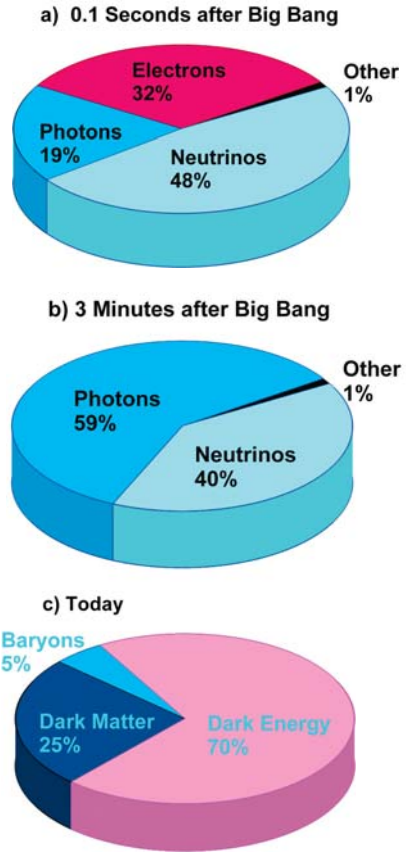


Figure 1. The dominant matter and energy components in the Universe at three selected times.

A tenth of a second after the Big Bang, relativistic photons and leptons dominate. Dark matter and baryonic matter (protons and neutrons) contribute much less than 1 percent of the density.

At a cosmic time of 3 minutes the pairs of positive and negative electrons have been annihilated. Controlled by the rules of thermodynamics, a part of the liberated energy heats the photon gas; the other part accelerates the expansion. A very small surplus of negative electrons remains and balances the charge of the protons.

In the present Universe, the photons have cooled down to 2.73 K and the neutrinos probably to ~2 K. The energy density of these light particles has become so low that they do not influence the dynamics of the Universe. Dark energy, dark matter and baryonic matter control cosmic expansion.

the time-span covered by Figure 1, but also the mixture of the matter and energy components changed drastically.

The abundances in Figures 1a and 1b are quantitatively based on results of elementary particle physics and established thermodynamic rules. In the present Universe (Fig. 1c) the density of baryonic matter is well-established. Less well established are as yet the densities of dark matter and dark energy, but progress is expected from ongoing and future research.

The Epoch of the Quark-Gluon Plasma

The elementary particles of baryonic matter are quarks. There are six kinds of quarks and six kinds of antiquarks. The nature of the strong interaction does not allow quarks to occur in isolation[3], but they can exist as mesons (quark-antiquark pairs), as baryons (three quarks), and as antibaryons (three antiquarks).

The quark-gluon plasma epoch is the earliest phase of the evolving Universe for which we can investigate microscopic processes in the laboratory. At that time the assemblage of quarks, antiquarks, gluons and other elementary particles behaved somewhat like a liquid. This epoch ended when, at a cosmic time of ~100 microseconds, the expanding and cooling Universe was approaching a density of 5×10^{16} kg/m^3 (or 50 million tons per cubic centimetre) and a temperature of 10^{12} K. The quark-gluon plasma became unstable and separated, forming mesons, baryons and antibaryons. Mesons decayed while baryons and antibaryons annihilated each other, all within microseconds. A very tiny fraction of the baryons was spared, but this was enough for populating all the galaxies in the Universe.

Symmetry Breaking in the Very Early Universe

The survival of some baryonic matter at the end of the quark-gluon plasma epoch remains an unresolved problem of cosmology. An excess of baryons over antibaryons could result from a difference in the behaviour of matter and antimatter. Andrei Sakharov, winner of the Nobel Prize for Peace in 1975, has listed observations that could account for the excess of baryons. The violation of the time-reversal invariance, found in the decay of neutral K-mesons is an example. A necessary condition is that protons should decay, however slowly, into mesons. No such proton instability has been found so far, but experiments revealed that the lifetime of the proton far exceeds the age of the Universe. This ascertains that even on a cosmic time scale the number of baryons will not decrease by spontaneous decay.

Baryon-antibaryon annihilation is a strong interaction process. Therefore, in a homogenous Universe only a totally insignificant amount of antibaryons should have left the Big Bang. Indeed, even though the identification of primordial antimatter is complicated by interactions of cosmic rays with matter that produce proton-antiproton pairs, experiments as well as observations have not given any indication of the presence of primordial antibaryons. In fact, the fraction of antiprotons found in cosmic rays is fully compatible with such a secondary origin.

Primordial Nucleosynthesis (0.1 s to 3 min in cosmic time)

During the epoch of primordial nucleosynthesis, lasting from ~100 milliseconds to ~3 minutes in cosmic time, the physics is well-known, so that we can make quantitative predictions for microscopic and macroscopic processes.

R.V. Wagoner, W.A. Fowler and F. Hoyle[4] formulated the theory of Standard Big Bang Nucleosynthesis (SBBN) in 1967. Based on Einstein's General Relativity, SBBN assumes a homogeneous and isotropic Universe during the epoch of nucleosynthesis, and neglects degeneracies of leptons. When, with the LEP collider at CERN, it was shown that there exist three neutrino flavours ($N_\nu = 3$), this was included in the SBBN theory. Recent results of neutrino oscillation experiments assure us that the rest masses of all these neutrinos are very low, low enough to be negligible during the nucleosynthesis epoch. Thus, the baryonic density remains *the only* important free parameter in the SBBN theory.

The sequence of events during the epoch of primordial nucleosynthesis is as follows: At a cosmic age of 10 milliseconds, the temperature had decreased to 10^{11} K. Mesons and heavier leptons had virtually all decayed, and only protons and neutrons, the lightest variety of baryons, remained. As a result, energy density and expansion rate were completely dominated by relativistic particles, i.e. photons, neutrinos and electrons[5], with protons and neutrons being only minor constituents (Fig. 1a). Since neutrons are heavier than protons, the neutron/proton ratio decreased with decreasing temperature through the weak interaction until, at a cosmic time of ~1 second and a temperature of ~10^{10} K, the weak interaction became ineffective, and the neutron/proton ratio was frozen-in at a value of one fifth. Afterwards, beta decay of the neutrons slowly decreased this ratio further until all neutrons were bound in stable nuclei.

Nucleosynthesis, i.e. the fusion of protons and neutrons into deuterium and heavier nuclei, effectively began when the temperature had decreased to 10^9 K at a cosmic time of ~100 seconds, and it was completed 200 seconds later. Since all the nuclei of atomic mass A = 5 and A = 8 are extremely short-lived, the production could not go beyond the isotopes of the lightest three elements (Fig. 2). Of these, only deuterium (D or ^{2}H), the heavy isotope of

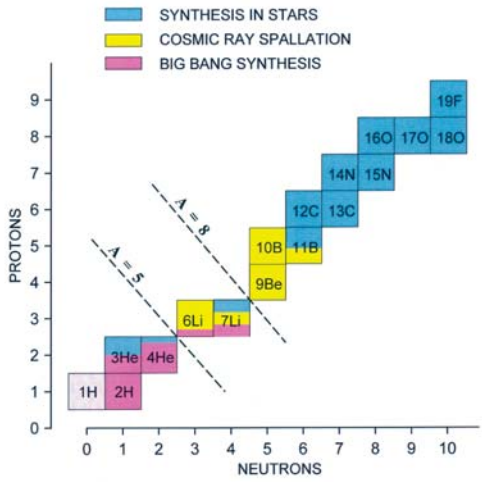


Figure 2. Origin of nuclei[6]. Because all nuclei having an atomic mass 5 or 8 are extremely short lived, the nucleosynthesis in the early Universe is limited to the lightest three elements. All nuclei with atomic mass A above 11 are produced in stars. For species with mixed origin, such as ^{7}Li, the relative proportions change with time and location.

hydrogen, was created exclusively (>99%) during the first few minutes in the life of the Universe.

The Universal Density of Baryonic Matter

The predicted Big Bang production of the isotopes of hydrogen and helium is shown in Figure 3. ^{1}H and ^{4}He represent more than 99.9% of the total mass. D and ^{3}He are rare and, as Figure 3 shows, their yields depend inversely on baryonic density. This is analogous to chemical reactions, where the yields of intermediate products decrease with increasing supply of reacting partners. Since the early 1970s, deuterium abundance measurements in the solar wind, meteorites, Jupiter and the Galactic interstellar gas were used to derive the primordial abundance of deuterium. Values of D/H in the range 3-5 × 10^{-5} were obtained from which a universal baryonic density of 3-6 × 10^{-28} kg/m^3 was calculated[7], corresponding to about 0.2 atoms per cubic metre. A general consensus existed on these values[7] until, in 1994, deuterium was measured by absorption of radiation from distant quasars in intervening clouds of gas. Since the investigated clouds are extremely old and virtually free of heavier elements (i.e. they have nearly "zero metallicity"), their deuterium abundance should be close to primordial. The problem was that widely varying D/H ratios were reported, ranging from 3×10^{-5} to 2×10^{-4}. These results reopened a broad discussion not only on the usefulness of galactic data for deriving primordial abundances, but also on the reliability of the SBBN theory for calculating the universal baryonic density.

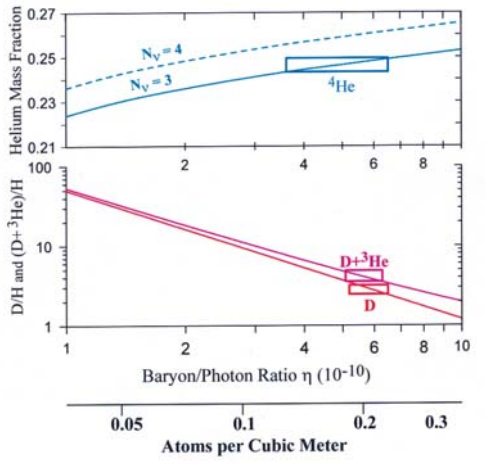


Figure 3. Predicted (solid lines) and observed (boxes) values of the primordial helium mass fraction (top) and the ratios D/H and (D+ ^{3}He)/H (bottom), as a function of the baryon/photon ratio and of the baryonic density. The solid line labelled $N_\nu = 3$ corresponds to the SBBN prediction (three neutrino flavours). The dashed line labelled $N_\nu = 4$ is calculated for the hypothetical case of four neutrino flavours (see text).

This was the situation when in May 1997 ISSI convened the workshop on *"Primordial Nuclei and their Galactic Evolution"*[8]. It became clear at this workshop that the low D/H ratios in distant clouds reported by D. Tytler and

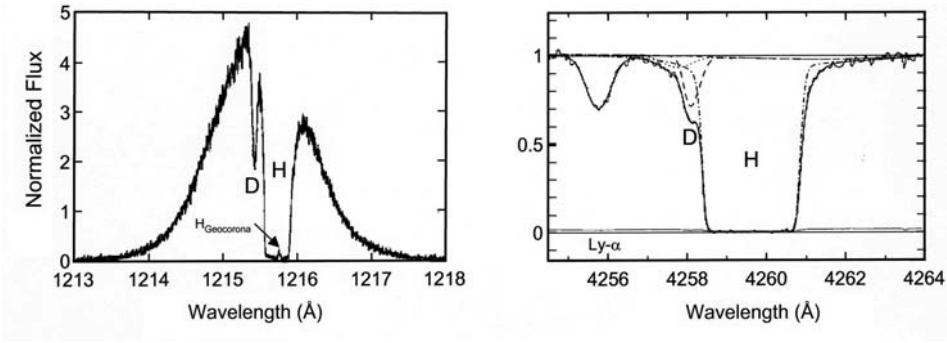


Figure 4. Deuterium abundances from Lyman-alpha absorption spectra.
Left: Absorption in the Local Interstellar Cloud (LIC) of the extreme-ultraviolet radiation from a nearby stellar source, as observed with the Hubble Space Telescope[11]. A deuterium/hydrogen ratio of 1.6×10^{-5} was derived from the hydrogen and deuterium Lyman-α absorption lines observed in spectra of several nearby stars.
Right: Absorption of radiation from quasars by a very distant cloud observed with the Keck 10-metre telescope on Mauna Kea, Hawaii[9]. This cloud is flying away from us at 85% of the speed of light. As a consequence, the Lyman-α lines of H and D are shifted from far-ultraviolet to visible wavelengths and can be observed from the ground. From several such distant clouds, a primordial deuterium/hydrogen ratio of 3×10^{-5} was derived[10].

associates[9], and not the high values found by other authors, could be reconciled with the ^{3}He and deuterium abundances in the Solar System and the present-day Galaxy.

The ^{3}He and deuterium abundances[11,12] in the Protosolar Cloud and the Local Interstellar Cloud (Figs. 4 and 5) demonstrated that the principal effect of stellar processing is the conversion of deuterium into ^{3}He with the sum, D+^{3}He remaining nearly constant[13] (Fig. 6). This was supported by new theoretical work[14] showing that ^{3}He from incomplete hydrogen burning does not have a large effect on the chemical evolution in the Galaxy.

The best current estimates of the primordial D/H and (D+^{3}He)/H ratios are compared in Figure 3 with the theoretically predicted values. It is evident that both

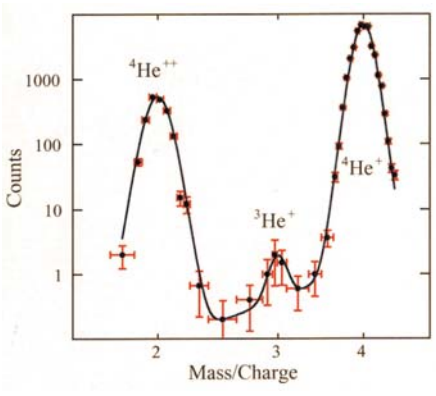


Figure 5. Neutral helium of the Local Interstellar Cloud (LIC) penetrates deep into the heliosphere where it can be directly investigated by spacecraft. The mass spectrum shown here was obtained with the Solar Wind Ion Composition Spectrometer (SWICS) on Ulysses[12]. The LIC is the only present-day galactic sample for which both the deuterium and ^{3}He abundances have been determined.

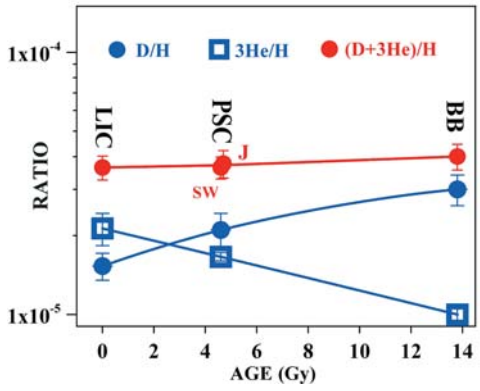


Figure 6. The abundance ratios, relative to hydrogen, of deuterium (D/H), the light helium isotope (^{3}He/H) and (D+^{3}He)/H in the Local Interstellar Cloud (LIC), the Protosolar Cloud (PSC), and very distant clouds that approximately represent matter released from the Big Bang (BB)[13]. Deuterium is exclusively produced in the Big Bang (Fig. 2), and converted thereafter into ^{3}He in stars. The net effect on these two species by other nuclear processes is found to be relatively small, so that throughout galactic history the (D+^{3}He)/H ratio remained nearly constant. Note that (D+^{3}He)/H in the PSC was derived independently from solar wind (SW) and Jupiter (J) data[13,15].

ratios give a universal baryon/photon ratio of $(5.8\pm0.6)\times10^{-10}$ and a present-day universal density of baryonic matter of $\sigma_B = (4.1\pm0.4)\times10^{-31}$ g/cm^3 or about 0.2 atoms per cubic metre. The baryon/photon ratio is one of the fundamental numbers of cosmology. So far, it is known only empirically. Any theory about the earliest phases of the Big Bang will have to predict a value that is compatible with the number derived from deuterium and ^{3}He.

Since the sum of D and ^{3}He is nearly independent of galactic evolution, the primordial baryonic density can be derived from this sum with little, if any, extrapolation. As Figure 6 shows, (D+^{3}He)/H in the two galactic samples and in the distant low-metallicity clouds are nearly the same. Thus, at the time of primordial nucleosynthesis, the baryonic densities in the far-away regions of these clouds and in our part of the Universe were the same, which is evidence for a homogenous Universe at the time of primordial nucleosynthesis.

Deuterium as a Tracer of Natural Processes

From that time on, deuterium in the Universe has been continuously decreasing, as stars are destroying, but not producing it. This "one-way" behaviour makes deuterium a unique tracer for physical and chemical processes in nature[16,17]. Deuterium in our bodies is authentic Big Bang stuff. Its relatively high abundance of 2-3 grams in each of us is due to chemical enrichment in the cold molecular cloud from which the Solar System formed[16,18].

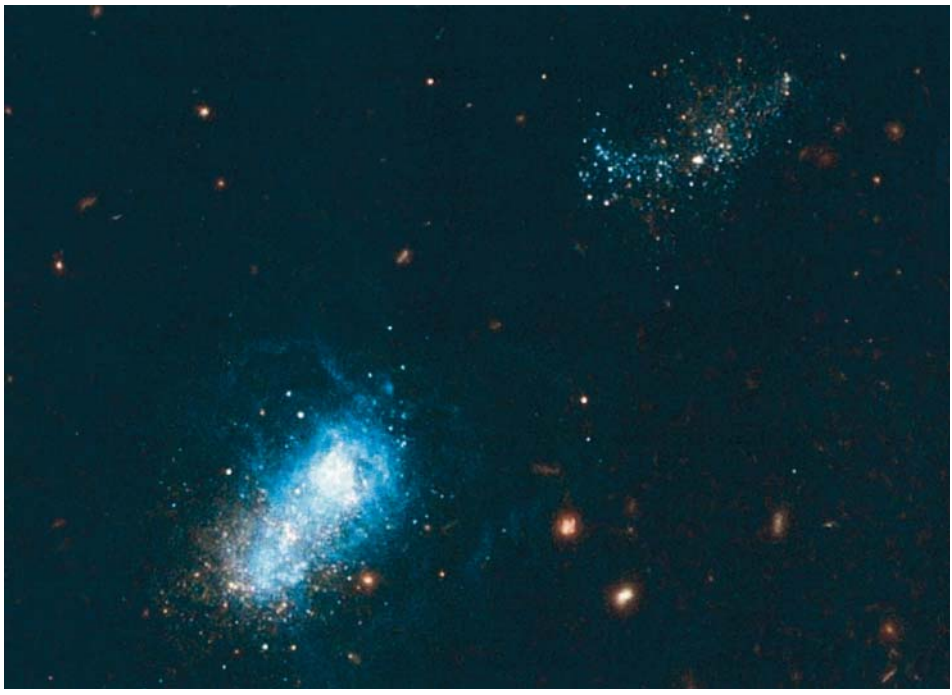


Figure 7. The blue-dwarf galaxy Izw18 at a distance of 40 million light years (observation by Y. Izotov & T. Thuan with the Hubble Space Telescope). Since the matter in this and other blue dwarfs is nearly unprocessed by stellar nucleosynthesis, they are suitable for deriving the composition of matter as it comes out of the Big Bang[19].

How Much Helium from the Big Bang, How Much from Stars?

The primordial abundance of ^{4}He is best obtained by extrapolating the helium abundance measured in H II regions to "zero metallicity", i.e. to vanishing O/H or N/H ratios. For many years, several authors using this method consistently derived a primordial helium mass fraction near 23%. Then, at the 1997 ISSI workshop, Trinh Xuan Thuan and Yuri Izotov reported on their observations of H II regions in blue dwarf galaxies (Fig. 7), and showed that the primordial mass fraction had to be increased to 24.5%[19], a value that has since been adopted. This may seem like a small change but, as the following discussion will show, it is significant for analyzing the physical processes in the early Universe.

The Big Bang production of ^{4}He depends only weakly on the baryonic density (Fig. 3). Thus, by using the baryon/photon ratio as determined above, ^{4}He can be used for testing the validity of the SBBN theory, or, to express it more generally, the validity of the laws of physics under the extreme conditions that were prevalent in the early Universe. To mention two examples:

The measured primordial abundance of ^{4}He is not compatible with the curve labelled $N_\nu = 4$ in Figure 3. This shows that, at a cosmic time of one second, the Universe was not populated by relativistic particles, other than photons, electron pairs and three neutrino flavours. This exclusion holds for light particles of any kind, provided they are covered by Einstein's equivalence principle and had interacted to attain a similar temperature to the known particles.

The agreement between the predicted and observed primordial helium abundances shows that at a cosmic time of one second the relative strengths of the strong, weak and gravitational forces were the same to within a few percent as those measured in laboratories on Earth. This is a remarkable invariance considering that, at a cosmic time of one second, the total density was 10^{35} times higher than it is in the present Universe.

Stellar Production of Carbon and Heavier Elements

The gap in the sequence of stable nuclei at atomic masses 5 and 8 (see Fig. 2) is overcome by the 3-alpha nuclear reaction, producing ^{12}C, the major isotope of carbon[20]. Since this reaction involves three partners, a high density is required for it to become effective. This condition is only fulfilled when stars have evolved into red giants, with high enough central densities and with temperatures of ~100 million degrees. Once ^{12}C is present in a star, the synthesis continues to heavier elements as the star contracts further and increases its core temperature. The fusion of lighter nuclei into heavier ones continues up to the group of elements around iron, which possesses the minimum free energy. Elements beyond the iron group are produced by slow neutron capture in red giants, and by the "r-process", an extremely rapid capture of neutrons during super novae explosions. The relative proportion of thorium, uranium and plutonium isotopes in the Solar System proves that "r-process" synthesis was not restricted to some violent early epoch, but has been going on throughout galactic history[21].

Dark Matter....

In 1937 Fritz Zwicky discovered that the visible mass of the galaxies in the Coma Cluster (Fig. 8: left) and other such clusters was far from sufficient to keep them gravitationally bound, and he concluded that these clusters were held together by a surplus of "dark matter" that astronomers could not readily account for.

During the last decades of the 20th century, it became increasingly clear that the Universe harbours more gravitational attraction than could possibly be accounted for by the 0.2 atoms/m^3 of matter that was derived from D and ^{3}He abundances

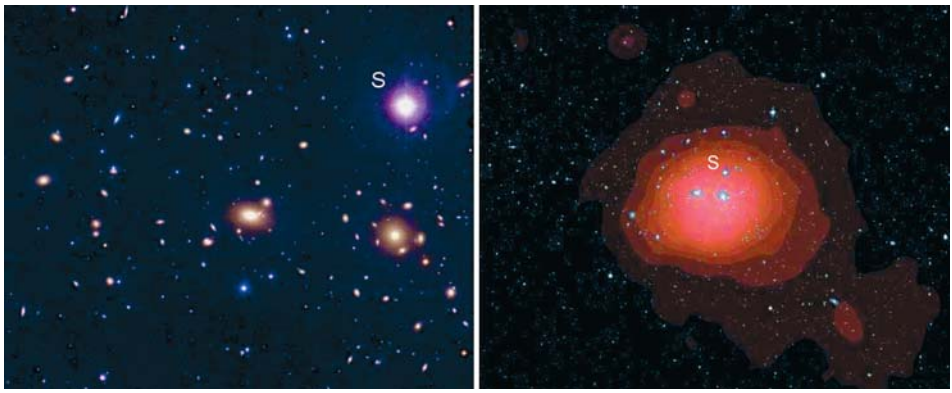


Figure 8. Left: The central region of the Coma Cluster. This cluster of galaxies is 300 million light years away from us, and consists of 2000 galaxies of various sizes. The two brightest of them in the centre are much more massive than the Milky Way or the Andromeda Galaxy (S is a nearby star). Right: X-ray image of the coma cluster from the Rosat All-Sky Survey (H. Böhringer, MPE)[22]. The optical image (from the Palomar Sky Survey) is superposed. The density of the hot intergalactic medium emitting the X-rays is not sufficient to account for the gravitational potential that holds this cluster of galaxies together; non-baryonic dark matter must be contributing the largest share. The fractions of matter in the Coma Cluster and other large clusters are typically ~5% in the galaxies, ~20% in the X-ray luminous gas, and ~75% in non-baryonic dark matter[22, 23].

(Figs. 3 and 6). Not only the clusters of galaxies, but also the galaxies themselves are embedded in potential wells that are mainly caused by non-baryonic dark matter (see Figs. 8: right and 9).

At the ISSI workshop on *"Matter in the Universe"* held in March 2001, astronomical observations at various wavelengths including gamma-rays, and X-rays were presented, along with results from gravitational lensing[25]. They all showed that non-baryonic matter contributes most of the gravitational forces on the scale of clusters of galaxies and even galaxies[22, 26]. Indeed, these clusters would fly apart, and galaxies - including our own - would tend to disintegrate without the presence of an unknown form of matter (Figs. 8: right and 9). The fluctuations in the Cosmic Microwave Background (CMB) radiation demonstrated on a cosmic scale that, in addition to baryonic matter, a non-baryonic form of matter must have come out of the Big Bang[27].

....and Dark Energy

In recent years, observations of type IA supernovae explosions indicated that the expansion of the Universe has been speeding-up during the past several billion years. This effect is attributed to a "dark energy" with an equation of state that combines positive energy with negative pressure[28-30]. In a medium with negative pressure (p), the energy density (ε) decreases less rapidly, because the medium

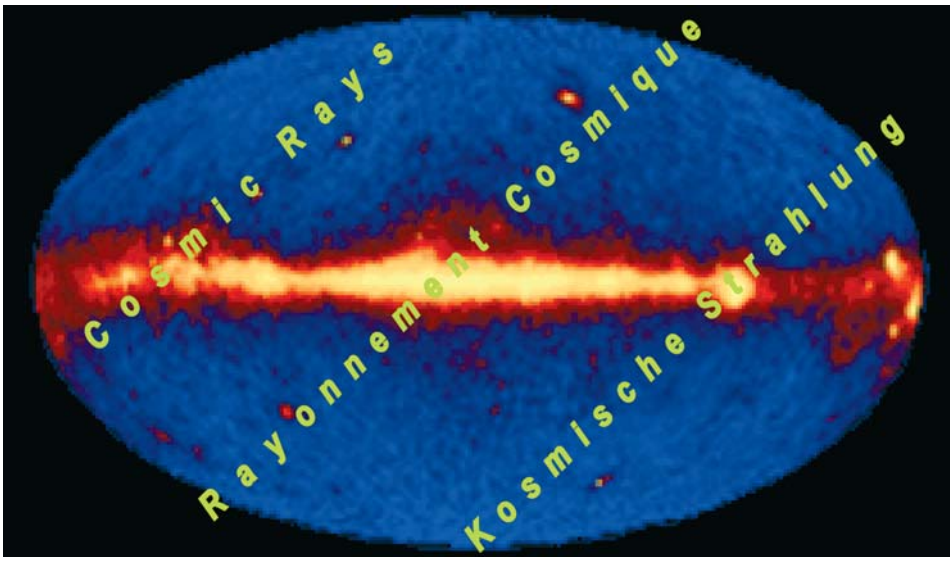


Figure 9. Cosmic rays produce gamma-rays whenever they hit baryonic matter. This map, from NASA's Compton Gamma-Ray Observatory, shows that gamma-ray sources are essentially confined to the galactic disk[24]. (Bright spots outside the disk are AGNs and a quasar, located far away from our Galaxy). Indirect evidence shows that the cosmic rays are not confined just to the galactic disk, but fill a large halo. The absence of gamma-ray sources outside the disk implies that there is very little gas or dust of baryonic matter in the halo.

receives, but does not expend work. When p/ε is below $-1/3$, the negative pressure overcomes the gravitational attraction, accelerating the expansion[31].

In the years following the 2001 ISSI workshop, new and refined measurements have firmed up the above conclusions[30,32], without fundamentally changing those presented at the workshop[25,26,33,34]. Particularly, the CMB observations obtained with the Wilkinson Microwave Anisotropy Probe (WMAP, Fig. 10) confirmed the earlier results and improved quantitative predictions[35].

The relative proportions of the major forms of energy in the Universe are shown in Figure 1c. The share of baryonic matter is modest, but its density as derived from deuterium and ^{3}He is very robust. If, for example, all the dark matter were baryonic, the abundance of deuterium would be 80 times lower, and neither the D absorption lines in Figure 4, nor the ^{3}He peak in Figure 5 would be noticeable. Whereas dark and baryonic matter decelerate the expansion of the Universe, dark energy tends to accelerate it [31]. Since the densities of dark matter and baryonic matter decrease more quickly than the density of dark energy, deceleration of the expanding Universe turned into acceleration several billion years ago, and so it will continue to expand into the distant future. That is our present understanding.

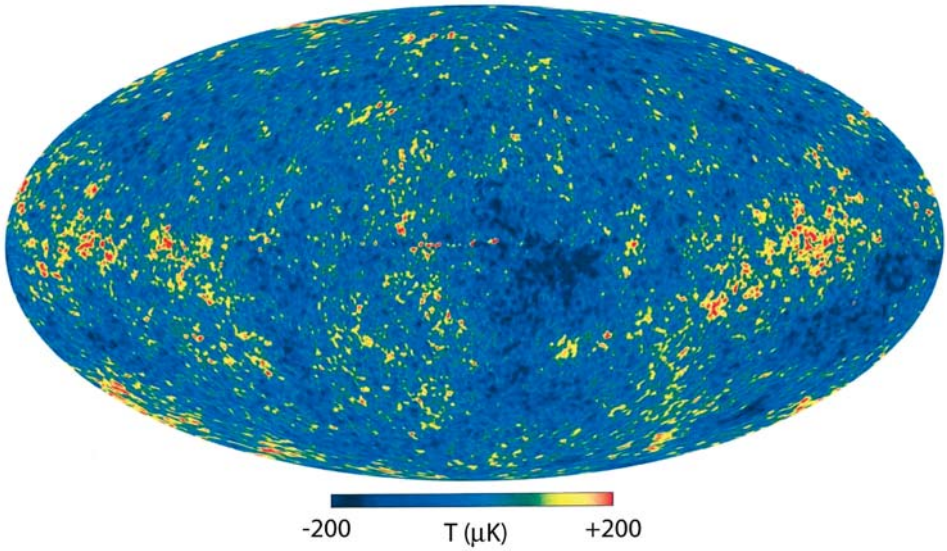


Figure 10. Full-sky map of the Cosmic Microwave Background (CMB) radiation observed with NASA's Wilkinson Microwave Anisotropy Probe (WMAP). The foreground contamination was suppressed[35]. The extremely small (± 200 micro-degrees) CMB-temperature fluctuations developed in the early universe. These fluctuations give the most direct evidence for the pivotal role of Dark Matter in structure building[27,35].

Building Cosmic Structure

Dark matter, not being affected by electromagnetic interactions, decoupled from the photon gas very early and initiated the growth of cosmic structure long before baryons could have done this. When at a cosmic time greater than 100,000 years baryons decoupled from photons, the baryons were drawn into already existing blobs of dark matter and began to form the structures we

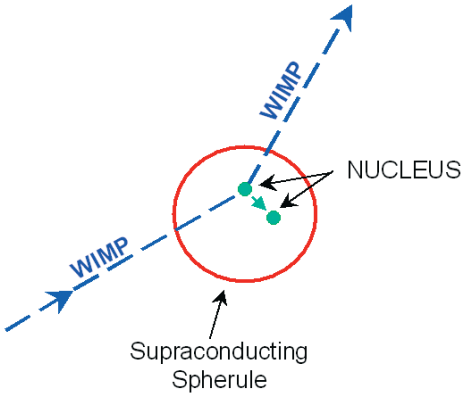


Figure 11. Method of detection of Weakly Interacting Massive Particles (WIMPs) in supraconducting spherules exposed to an external magnetic field[33]. When a WIMP has one of its rare collisions, it transfers momentum and energy to the supraconducting spherule causing a phase transition from the supraconducting to the normal state, which will be detected in a pickup coil.

Figure 12. The Large Magellanic Cloud is a dwarf galaxy that is bound to our Galaxy at a distance of about 150 000 light years. (Cederberg Observatory, South Africa)

observe. In places of strong enough concentration, baryonic matter, contracting under its own weight, formed stars that then produced carbon and heavier elements, essential ingredients of complex molecules and crystals. These highly organized systems of baryonic matter are the crucial building blocks of comets, solid planets and life.

The Nature of Dark Matter

The particles of dark matter have not yet been identified. Weakly Interacting Massive Particles (WIMPs), but also virtually non-interacting light particles (axions), are being considered. Experiments to detect WIMPs produced by accelerators or natural WIMPs are underway (Fig. 11). Such measurements could provide information on the mass and interaction properties of the dark-matter particles and on their temperature in the solar neighbourhood. These properties of the dark matter allow predictions regarding the evolution of medium-scale structures, such as the number of dwarf galaxies in relation to fully grown galaxies, the amount of baryonic matter falling quasi-continuously into our own Galaxy, or the concentration of matter towards the centre of galaxies and clusters. Such predictions are important for comparison with observations.

The direct detection of dark-matter particles is, in principle, made easier by its concentration in the galactic potential well. For the neighbourhood of the Sun, an energy density of 0.3 GeV/cm^3 of dark matter is derived from the width and depth of this well[33]. This corresponds to 580 grams of dark matter for the whole volume of Earth. Comparison with the mass of the Earth of 6×10^{24} kilograms shows just how much more locally concentrated and structured baryonic matter is in the Universe, a direct consequence of the strong and electromagnetic forces.

Dark Matter and Chemical Evolution in the Local Group

Dark matter must have a strong influence on the dynamics and evolution of the fifty dwarf galaxies in the Local Group, as well as on the Galaxy and the

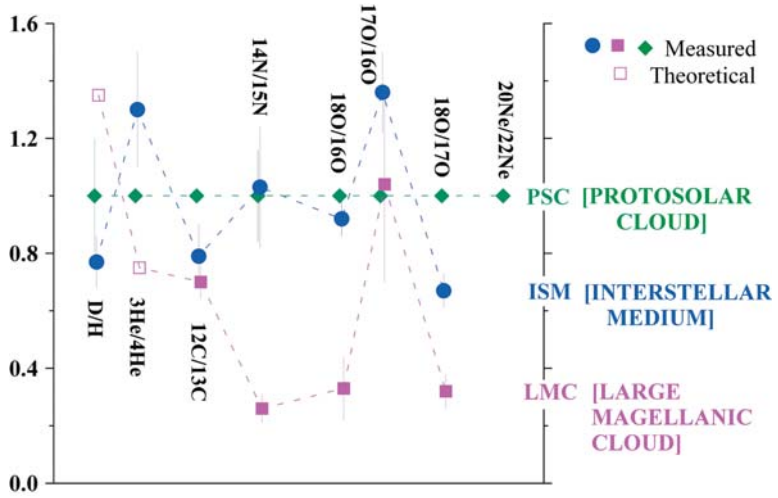


Figure 13. Isotopic ratios of C, N and O in the interstellar medium (ISM) and in the Large Magellanic Cloud (LMC) normalized to the protosolar ratios[37]. The huge differences testify to the potential of isotopic measurements for investigating the dynamical and chemical history of galaxies and the Local Group of galaxies.

Andromeda Nebula, the Group's dominant members. The strength and geometry of the gravitational potential in the Local Group have not only determined the trajectories, but also the structural and chemical evolution of Group members. The differences between elemental and isotopic composition in the Galaxy and the Magellanic Clouds are striking (Figs. 12 and 13), and attempts are being made to pinpoint the causes of these differences. Some observations suggest that infall of matter from dwarf galaxies is making a significant contribution to the chemical evolution of our own Galaxy[36]. If successful, such studies might also shed light on the nature of dark matter and the gravitational field that has shaped the history of the Local Group.

Is the Universe Flat?

Carl Friedrich Gauss was the first to question and to test the flatness of natural space. In 1823 he measured the angles in a triangle of 2800 km^2 between three landmarks in the region around Göttingen, using his least-squares method for "data reduction". This was a pioneering experiment, but as we now know, the deviation from 180 degrees for the sum of the three angles is far too small to be measurable. Nevertheless, during the early decades of the 19th century Gauss and others[38] developed the non-Euclidean geometries that are adequate for describing curved space in the cosmos on any scale. High matter densities cause strong

warping of space. Gauss' experiment, if performed above the surface of a neutron star of 1.5 solar masses, would have given 183 degrees[39] for a triangle of ten square kilometres.

The total energy density in the Universe, i.e. the contributions from dark energy, dark matter and baryonic matter taken together, is found to be close to the critical density, implying that on a very large scale, space is Euclidean or flat, at least approximately. This does not mean, of course, that the pre-Einstein physics would provide an adequate description of cosmology. Whereas, Newton's physics *presupposes* space to be Euclidean, Einstein's General Relativity provides a natural link between matter, energy and the curvature of space, and based on present observation it *concludes* that space is rather flat on the largest scale. Near strong mass concentrations, however, space is significantly curved and non-Euclidean[40].

References and Notes

1 During the last decades, Einstein's theory of General Relativity has been extensively applied and tested at short distances near the Sun and in the Solar System, in the strong gravity fields of pulsars, and at cosmological distances and timescales. No conflict between theory and observation was ever confirmed.

2. About 1923 Aleksandr Friedmann of St. Petersburg showed that a variety of expanding universes could be derived from Einstein's General Relativity theory.

3. The theory of the strong interaction, the "Quantum Chromodynamics" was formulated in 1973; see D.J. Gross & F. Wilczek, *Phys. Rev. Lett.*, **30,** 1343, 1973; H.D. Politzer, *Phys. Rev. Lett.* **30**, 1346, 1973; H. Fritzsch, M. Gell-Mann & H. Leutwyler, *Phys. Lett.,* **B 47,** 365, 1973.

4. R.V. Wagoner, W.A. Fowler & F. Hoyle, *Astrophysical J.,* **447**, 680, 1967.

5. Due to the extremely weak interaction of gravitons with other particles, their temperature should be much smaller in the nucleosynthesis epoch than that of photons and neutrinos, so that their energy density can be neglected, see e.g. E.W. Kolb & M.S. Turner, The Early Universe, Frontiers in Physics, Reading MA, Addison-Wesley, 1990.

6. J. Geiss & R. von Steiger, in *Fundamental Physics in Space*, Proc. Alpbach Summer School 1997, *ESA SP-420*, 99, 1997.

7. See J. Geiss & H. Reeves, *Astron. Astrophys.,* **18**, 126, 1972; H. Reeves, J. Audouze, W.A. Fowler, D.N. Schramm, *Astrophys. J.,* **179**, 909, 1973; M. Riordan & D.N. Schramm, The Shadows of Creation, W.H. Freeman & Co., New York, 1991.

8. N. Prantzos, M. Tosi & R. von Steiger (Eds.), Primordial Nuclei and their Galactic Evolution, SSSI Vol. 4, Kluwer Academic Publ., Dordrecht, 1998, and *Space Science Rev.,* **84**, 1-326, 1998.

9. D. Tytler *et al.*, *Nature,* **381,** 207, 1996; S. Burles & B. Tytler, in Ref. 8, p. 65

10. J.M. O'Meira, D. Tytler, D. Kirkman *et al.*, *Astrophys. J.,* **552**, 718, 2001.

11. J.L. Linsky, in Ref. 8, p. 285.

12. G. Gloeckler & J. Geiss, *Nature,* **386**, 210, 1996; G. Gloeckler & J. Geiss, in Ref. 8, p. 275.

13. After J. Geiss & G. Gloeckler, in Ref. 17, p. 3, with references to the origin of the data.
14. C. Charbonnel**,** in Ref. 8, p. 199; M. Tosi, in Ref. 8, p. 207.
15. P.R. Mahaffy, T.M. Donahue, S.K. Atreya, T.C. Owen & H.B. Niemann, in Ref. 8, 251.
16. J. Geiss & H. Reeves, *Astron. Astrophys.,* **93**, 189, 1981.
17. R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T. Owen & F. Robert, (Eds.), Solar System History from Isotopic Signatures of Volatile Elements,SSSI **1**6, and *Space Science Rev.,* **106**, Nos. 1-4, 2003.
18. T. Owen & T. Encrenaz, in Ref. 17, p. 121.
19. T.X. Thuan & Y.I. Izotov, in Ref. 8, pp. 83.
20. E.M. Burbidge, G.R. Burbidge, W.A. Fowler & F. Hoyle, *Rev. Modern Phys.,* **29**, 547, 1957.
21. F.-K. Thielemann, P. Hauser, E. Kolbe *et al*., in Ref. 25, p. 277.
22. H. Böhringer, in Ref.25, p. 49.
23. A.E. Evrard, MNRAS, **292**, 289, 1997.
24. All-sky gamma-ray map (E>100MeV) by the Energetic Gamma Ray Experiment Telescope on board the NASA Compton Gamma Ray Observatory.
25. Ph. Jetzer, K. Pretzl & R. von Steiger (Eds.), Matter in the Universe, SSSI **14** and *Space Science Rev*., **100,**1-319, 2002.
26. S. Schindler, in Ref. 25, p. 299.
27. R. Rebolo, in Ref. 25, p. 15.
28. S. Perlmutter *et al., Astrophys. J.,* **517**, 565, 1998.
29. C. Wetterich, in Ref. 25, p.195.
30. R.P. Kirshner, *Science,* **300**, 1914, 2003.
31. An accelerating expansion is not new in cosmology. It is an essential feature of the Steady-State Theory developed in 1948 by Hermann Bondi, Thomas Gold and Fred Hoyle, and of the inflationary epoch proposed to having occurred in the very early universe.
32. J.P. Ostriker & P. Steinhardt, *Science,* **300**, 1909, 2003.
33. K. Pretzl, in Ref. 25, p. 209.
34. H. Reeves, in Ref. 25, p. 312.
35. C.L. Bennett *et al*., *Astrophys. J. Suppl. Series*, **148**, 1, 2003.
36 . J. Geiss, G. Gloeckler & C. Charbonnel, *Astrophys. J.,* **578**, 562, 2002.
37. Y. Chin, in New Views of the Magellanic Clouds, IAU Symposium, **190**, 279, 1999 and references given therein.
38. Following the publication of Euclid's "The Elements" more than 2000 years ago, mathematicians debated whether the "Parallel Axiom" was needed, could be proven, or even whether a geometry could be constructed without it. Consistent non-Euclidian geometries were only developed during the 19th century by Carl Friedrich Gauss at Göttingen, the Hungarian Janos Bolyai, the Russian Nicolay Ivanovich Lobachevsky, and Gauss' student Bernhard Riemann.
39. This value is based on an equation derived by Petr Hajicek of the University of Bern.
40. We thank Heinrich Leutwyler and Rudolf Treumann for critically reading the manuscript and for valuable discussions and advice. We are indebted to Hans Böhringer, David Tytler, Klaus Pretzl, Trinh Xuan Thuan and Thomas Schildknecht for advice and help with figures. This work was in part supported by NASA/JPL contract 955460 and NASA/Caltech grant NAG5-6912.

Cosmic Rays in the Galaxy and the Heliosphere

R.A. Mewaldt[a] and G.M. Mason[b]

[a]*California Institute of Technology, Pasadena, California, USA*
[b]*University of Maryland, College Park, Maryland, USA*

Introduction

Charged particles are accelerated to speeds approaching that of light throughout the Universe. Since the discovery of the mysterious "cosmic rays" in 1910 by the Austrian physicist V.F. Hess, we now recognize several sources of high-energy particles observed at Earth: (1) the Galaxy – Galactic Cosmic Rays, (2) the Sun – Solar Energetic Particles, and (3) Anomalous Cosmic Rays believed to be accelerated at the boundary of the Solar System or heliosphere. It appears that most high-energy particles observed in the heliosphere and Galaxy were energized by a mechanism first proposed by Enrico Fermi at shock waves produced by explosive events on the Sun, at planetary bow shocks, at the solar-wind termination shock at the outer edge of the heliosphere, and at shock waves produced by supernova explosions. The launch of a new generation of instruments during the past decade has made it possible to test models of shock-acceleration processes with high-resolution composition data in a number of environments. These new data indicate that shock acceleration is a selective process – not all particles are injected and accelerated with the same efficiency. We discuss several new results on cosmic rays in the Galaxy and the heliosphere, focusing on some that bear on the question of what material is accelerated to become cosmic rays. These topics were all drawn from presentations and discussions that occurred at three ISSI workshops summarized in the Space Science Series of ISSI, Volumes 3, 10, and 13[1-3].

Cosmic Rays in the Galaxy

During the 1990s high-resolution measurements of the isotopic composition of galactic cosmic rays (GCRs) with improved statistical accuracy became available, beginning with the Ulysses mission and culminating on ACE. A comparison of the abundances of various heavy isotopes in cosmic-ray source material and Solar System material is shown in Figure 1. The surprising level of agreement between these two samples of matter, spanning four orders of magnitude

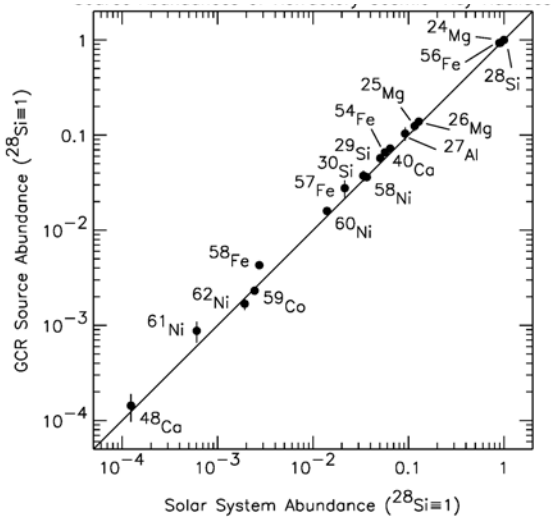


Figure 1. Comparison of cosmic-ray source abundances[4,7] with corresponding Solar System values[8]. Points along the solid line correspond to equal abundances, relative to ^{56}Fe, in the two samples of matter. Only non-volatile elements are included in this comparison.

in abundance, implies that cosmic-ray source material must originate from a stellar mixture very similar to that which contributed to the interstellar matter from which the Solar System condensed[4]. In addition, since the cosmic-ray lifetime in the Galaxy is only ~15 million years as a result of various loss processes[5], cosmic rays must be a recent sample of galactic material, and it appears that there has been relatively little evolution in the composition of interstellar matter over the 4.6 billion years since the Solar System formed[4]. (For a discussion of galactic evolution over the last 5 billion years, see the recent paper by Geiss *et al*.[6])

The one exception to the agreement in Figure 1 is ^{58}Fe, which is overabundant in cosmic-ray source material by a factor of ~1.75. The other well-known anomaly in the cosmic-ray isotopic composition is ^{22}Ne; the ^{22}Ne/^{20}Ne ratio is a factor of ~5 greater in the cosmic-ray source than in the solar wind[9]. Both ^{22}Ne and ^{58}Fe are the products of He burning. Some years ago, Casse and Paul[10] proposed that a significant fraction of heavy cosmic rays originate in Wolf-Rayet (W-R) stars – massive (>25 $M_\odot$) stars with intense stellar winds of several thousand km/s (see Fig. 2a). As a result of mass loss, the H envelopes of these stars have been stripped off and the winds of one class of these stars are comprised of He-burning products emerging from their surfaces. Casse and Paul pointed out that if W-R He-burning products were mixed with material of normal composition in the correct proportion, one could account for both the ^{22}Ne anomaly and the C/O ratio in cosmic rays (C/O ≈ 0.5 in the Sun, while C/O ≈ 1 in cosmic rays). This

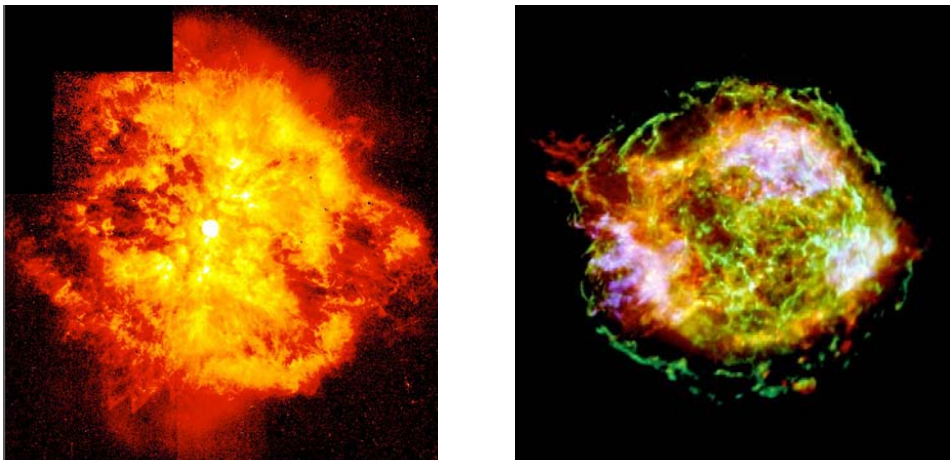


Figure 2. (*a*) Image of Wolf-Rayet star WR124 surrounded by nebula M1-67[11]. Such stars are thought to produce a significant fraction of the heavy elements in cosmic rays. (*b*) The supernova remnant in the constellation Cassiopeia, Cas A[12], is ~10 light years in diameter. The bright outer ring (green) marks the location of the shock wave generated by the supernova explosion.

model also predicted smaller enhancements for other species[13-14], including ^{58}Fe. A comparison of the predictions of the W-R model[9] with Fe-group measurements from ACE, indicates reasonable agreement (see Fig. 3). More recent calculations by these workers also show improved agreement[15] with additional isotope ratios including Mg and Si.

The correspondence between the W-R composition and cosmic-ray composition anomalies is very suggestive, but it is also necessary to show that there is enough W-R material available, and to understand why this material should be selected for acceleration to cosmic-ray energies rather than other samples of interstellar material. These, and related questions sparked lively debate[14,16,17] at the ISSI workshop on the Astrophysics of Cosmic Rays in 2000. It is generally believed that most cosmic rays are accelerated by supernova shock waves[18] (see Fig. 2b). Supernovae (SN) can provide the 3 x 10^{40} ergs/s needed to sustain cosmic rays in the Galaxy. Shock acceleration also accounts for the observed energy spectrum of primary cosmic rays. However, measurements of the electron-capture isotopes ^{57}Co and ^{59}Ni (and their daughters) in cosmic rays imply that there is a delay of at least 10^5 years between the nucleosynthesis and acceleration of cosmic-ray material[19], implying that supernovae do not accelerate their own ejecta.

Casse & Paul[10] suggested that W-R stellar wind material might be pre-accelerated by processes analogous to those in corotating interaction regions (CIRs) in the heliosphere, and then further accelerated at the termination shock surround-

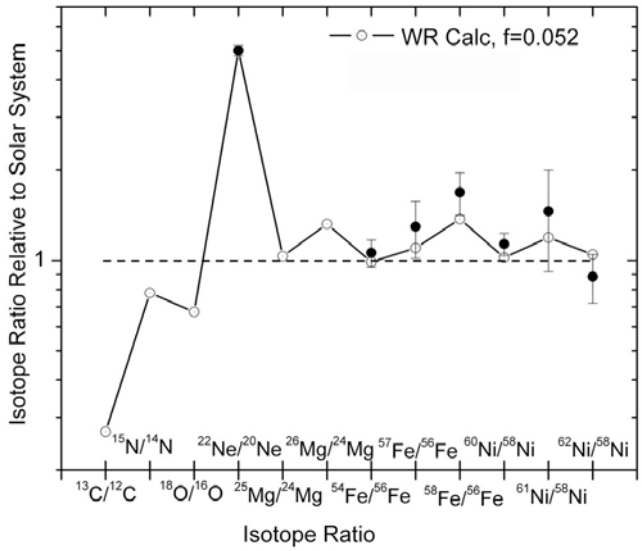


Figure 3. Isotopic abundances for Fe and Ni isotopes[4] are compared with a model of cosmic-ray composition due to Meynet *et al*.[14] in which 5.2% of Wolf-Rayet material is mixed with interstellar material with a Solar System composition. The calculation was normalized to the $^{22}Ne/^{20}Ne$ ratio in cosmic-ray source material (figure from Ref. 8). The dashed line shows the expected ratios with no contribution from WR material.

ing a W-R star. Maeder and Meynet[20] proposed a model in which cosmic rays originate from W-R stars and supernovae in the inner galaxy, where the relative density of W-R stars is greater. The model of Meyer *et al*.[21], in which the refractory (non-volatile) elements in cosmic rays originate from material sputtered from accelerated dust grains, also invoked the acceleration of W-R stellar winds by the subsequent explosion of the W-R star to explain the cosmic-ray $^{22}Ne/^{20}Ne$ ratio.

It is now known that most supernovae occur in super-bubbles that are created by correlated explosions of a number of core-collapse supernovae. Higdon and Lingenfelter[22] argue that the bulk of galactic cosmic rays are accelerated in the interiors of super-bubbles that are enriched by the metals synthesized in the supernova, and also in ^{22}Ne and other He-burning products from the W-R winds of the more massive SN progenitors. This model is consistent with the decay of ^{59}Ni in cosmic rays (see above) if SN ejecta reside in the super-bubble for at least 10^5 years before being accelerated by shocks from subsequent SN within the same super-bubble. An attractive aspect of the super-bubble model is that W-R winds are co-located with the SN shocks believed to be responsible for cosmic-ray acceleration.

Cosmic Rays in the Heliosphere

Soon after the discovery of anomalous enhancements in the low-energy spectra of He, N, O and other ions some 30 years ago, it was proposed that these anomalous cosmic rays (ACRs) originate from interstellar neutral atoms that drift into the heliosphere[23], are ionized, and then convected to the solar-wind termination shock where they are accelerated[24] to energies of 10 to 50 MeV/nucleon. This theory accounts for the anomalous abundances of species that have substantial neutral abundances in the local interstellar medium (LISM), including H, He, N, O, Ne and Ar.

Quite aside from the dynamic heliospheric processes that produce the anomalous cosmic rays, the modulation of the ACR and GCR intensities by the solar wind (by processes collectively termed "solar modulation") has a major effect on the observed intensities at Earth's orbit. Figure 4 shows the intensities of high-energy (cut-off energy 2.99 GeV) cosmic rays over several solar cycles. The alternating appearance of "more sharply pointed" versus "flatter" solar minima peaks is due to particle drifts whose effects alternate as the solar magnetic field undergoes a 22-year cycle[26]. More complete models of the heliosphere are now being used to help interpret these observations. For example, it is now evident that a significant fraction of GCR modulation takes place in the heliosheath[27], well beyond the present location of the Voyager spacecraft.

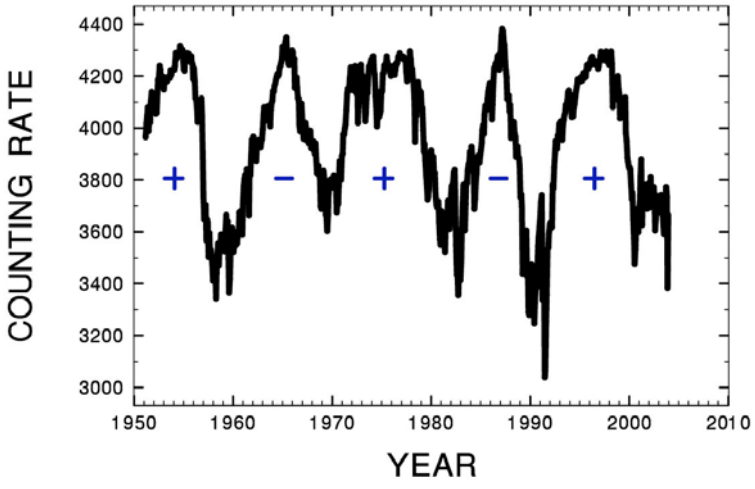


Figure 4. Intensity of high-energy cosmic rays measured by the Climax neutron monitor over almost five solar cycles (adapted from Ref. 25). The solar magnetic field is predominantly positive in the northern hemisphere for those solar cycles marked with a plus sign and predominantly negative in the northern hemisphere for those cycles marked with a minus sign.

Although many details remain to be worked out (see ISSI SSS Vol. 3), current heliospheric models have established a framework that allows interpretation of the spatial and temporal variations of GCR and ACR energy spectra[28,29]. For the GCRs, the primary effect of modulation is adiabatic energy loss that prevents low-energy particles from reaching 1 AU[30]. As a result, little is known of the spectrum of low-energy (<300 MeV/nuc) cosmic rays outside the heliosphere. Variations in the amount of adiabatic energy loss over the solar cycle also lead to changes in cosmic-ray isotope ratios that have recently been detected[31] by high-resolution spectrometers on ACE.

ACR observations are much more sensitive to modulation by the solar wind than are GCRs, due mainly to the lower energies of the ACRs. At 1 AU, the intensity of ACR oxygen, for example, varies by more than a factor of 100 from solar maximum to solar minimum[32]; much greater than the factor of ~4 solar-cycle variation for ~200 MeV/nucleon GCRs. In current models, the location of maximum ACR production varies with the polarity of the solar cycle: for the cycles marked with a plus sign in Figure 4, the solar magnetic field is predominantly positive in the northern hemisphere and particles drift along the shock to the poles and then downward into the heliosphere; for the cycles marked with a minus sign, the solar fields are reversed and the maximum ACR intensity is at the equator[33]. These changes cause the predicted latitude gradients of ACRs to change sign from one cycle to the next, consistent with observations in the outer heliosphere. (Since positive and negative cosmic rays drift in opposite directions in the heliosphere, there are also solar-cycle variations in the observed ratio of cosmic-ray electrons to ions, and in the ratio of cosmic-ray positrons to electrons, that validate this picture[34].) Progress such as this on cosmic-ray transport in the heliosphere, in addition to developments in shock acceleration theory, has made it possible to construct models of ACR origin, acceleration and transport, and to relate the observed composition to sources in the local interstellar medium[29].

Beginning with the solar-minimum period in the mid-1990s evidence was presented for additional ACR species, including elements such as Mg, Si, and S that are expected to be mainly ionized in the ISM[35, 36, 29]. These observations suggested that ions of other origins might be accelerated at the termination shock. The ACR abundances observed by Voyager are shown in Figure 5, along with those of possible seed populations that might contribute. The neutral interstellar source can account for the observed abundances of H, He, N, O, Ne, Ar, and possibly C (somewhat <1% of C in the LISM is expected to be neutral), but Solar System sources are apparently required to account for the observed low-energy increases in the rare ACR ions Na, Mg, Si, and S.

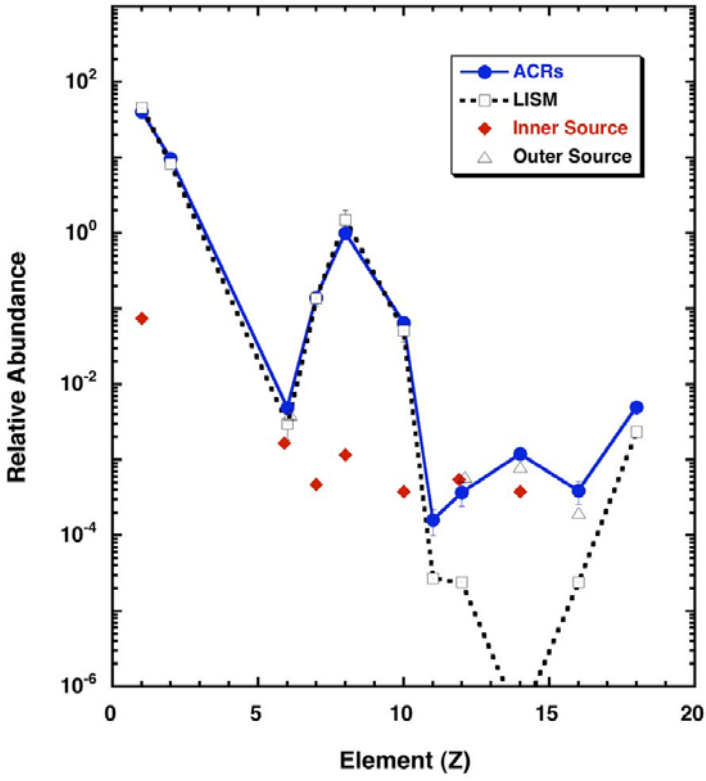


Figure 5. Comparison of the intensity of anomalous cosmic-ray species from Voyager[29] with the relative abundance of suggested seed populations, including interstellar neutrals (LISM; derived from a combination of pickup-ion[37] and local interstellar[38] abundances), inner-source pickup ions[39], and outer-source pickup ions[40]. This comparison was adapted in large part from data in Cummings, Stone and Steenberg[29] by assuming identical acceleration efficiencies for ions with $Z \geq 6$, but using the derived acceleration efficiencies for H and He. Note that interstellar gas[23] is the main source of ACR H, He, N, O, Ne, and Ar, while Na, Mg, Si, and S appear to be of Solar System origin. Carbon may include significant contributions from both interstellar and Solar System sources. (We are not aware of any observations or predictions of pickup-ion Na.)

It would certainly not be surprising if some solar wind were accelerated at the termination shock. The solar wind's composition is consistent with the observations of the rare ACR ions, and only a small fraction (~10^{-4}) need be accelerated to make a significant contribution[41]. However, the shape of the energy spectra observed by Voyager favours a source of singly-charged rather than multiply-charged ions[29]. Another possibility is the so-called "inner source" of pickup ions[37], singly-charged ions that apparently result from solar-wind interactions with dust grains near the Sun. The inner source also appears to have about the right composition, but the near-Sun origin of the inner source already results in rapid cooling of these ions by 1 AU, and they are apparently not as efficiently

accelerated in interplanetary and corotating shocks as are interstellar pickup ions. Recently an "outer source" of pickup ions has been proposed[40], made up of pickup ions originating in the Kuiper Belt. Although there are not yet measurements of these ions, they undoubtedly exist at some level, and should not suffer as much cooling given their origin in the outer heliosphere. It remains to be seen if they have the correct composition and are present in sufficient numbers. Although data from the Voyagers in the vicinity of the termination shock may shed light on these possible sources, there is clearly a need for charge-state and composition measurements of suprathermal ions in the outer heliosphere.

Heliospheric and Galactic Cosmic-Ray Parallels

Galactic cosmic rays below 10^{16} eV are presently believed to be accelerated by supernova shock waves travelling through the interstellar medium. This process results in a power-law spectrum of particles that is further modified as higher energy particles escape from the Galaxy more easily than at lower energy. In the heliosphere, the energetic particle spectra are comprised of particles from a wide range of solar, interplanetary and galactic sources, some of which are highly variable. It was therefore somewhat surprising that the 3-yr integrated fluence of heliospheric particles, extending over six decades in energy, exhibits a reasonably high degree of organization that has several parallels with the GCR spectrum.

The measured fluence of oxygen nuclei from solar wind to galactic cosmic ray energies[42] is shown in Figure 6. The GCR all-particle spectrum[43] is shown in Figure 7. As indicated in Figure 6, this fluence spectrum results from a superposition of a wide range of acceleration processes taking place in various locations within the heliosphere. The peak at ~0.0008 MeV/nucleon corresponds to slow-speed solar wind with a mean velocity of ~400 km/s. The contribution of occasional higher speed streams, sometimes extending to ~1000 km/s, is also evident. Beyond 10 keV/nucleon, a long suprathermal tail extends with a power-law slope of -2 all the way to ~10 MeV/nucleon. At the low-energy end (~10 keV/nucleon), such tails are known to be continuously present on all solar-wind species, even in the absence of interplanetary shocks or visible solar activity[39]. However, at higher energies, from ~30 keV/nucleon to ~1 MeV/nucleon, most of the fluence comes from a superposition of many solar events.

Near ~10 MeV/nucleon there is a gradual "knee" and the oxygen spectrum (as well as those of other species[42]) briefly steepens. Over this 3-year period solar particles contribute much more than anomalous cosmic rays. Above ~100 MeV/nucleon, the modulated fluence of galactic cosmic rays begins to dominate, continuing on for many more decades in energy (see Fig. 7).

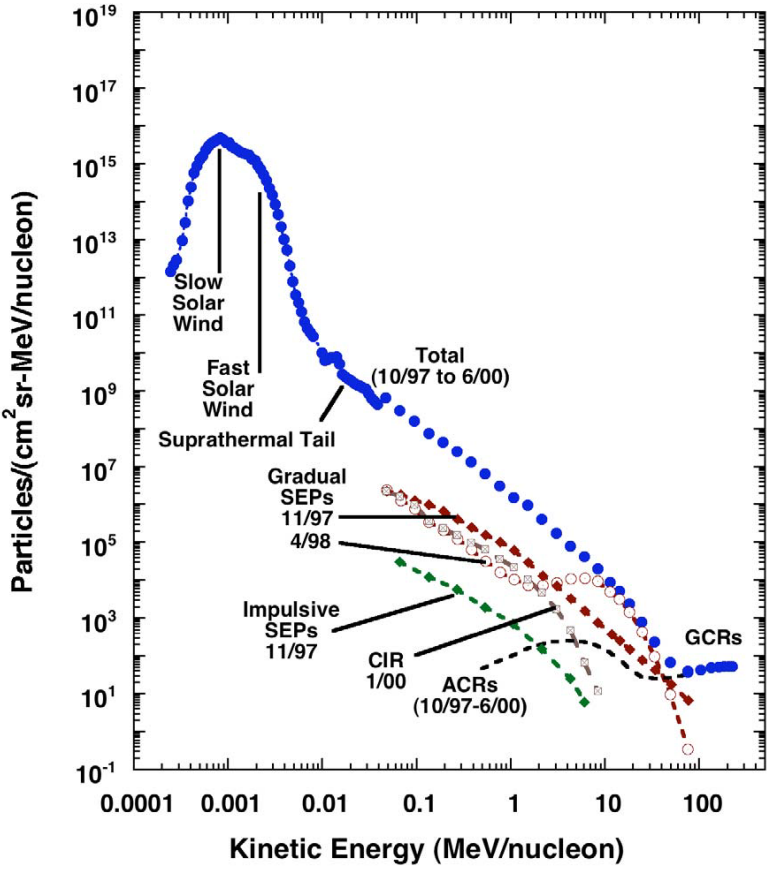


Figure 6. Fluence of oxygen nuclei measured from September 1997 through June 2000, a period that included both solar-minimum and solar-maximum conditions (adapted from Ref. 42). Also illustrated are examples of the contributions that several solar and interplanetary events made to the fluence. Note the region from ~10 keV/nucleon to ~10 MeV/nucleon, which has an E^{-2} spectrum. Both He and Fe were found to have very similar fluence spectra.

There are several similarities between the heliospheric spectra in Figure 6 and the GCR spectra in Figure 7. It is interesting that over almost three decades in energy, heliospheric species (including eight elements[42] and also ^{3}He) have power-law spectra with a slope of -2. This occurs in spite of the fact that the individual events that comprise these spectra often have non-power-law spectra, with variations of as much as a factor of 100 for abundance ratios such as Fe/O. It is likely that at much higher energies GCRs also include contributions from several sources that may have differing spectra and compositions, yet the pri-

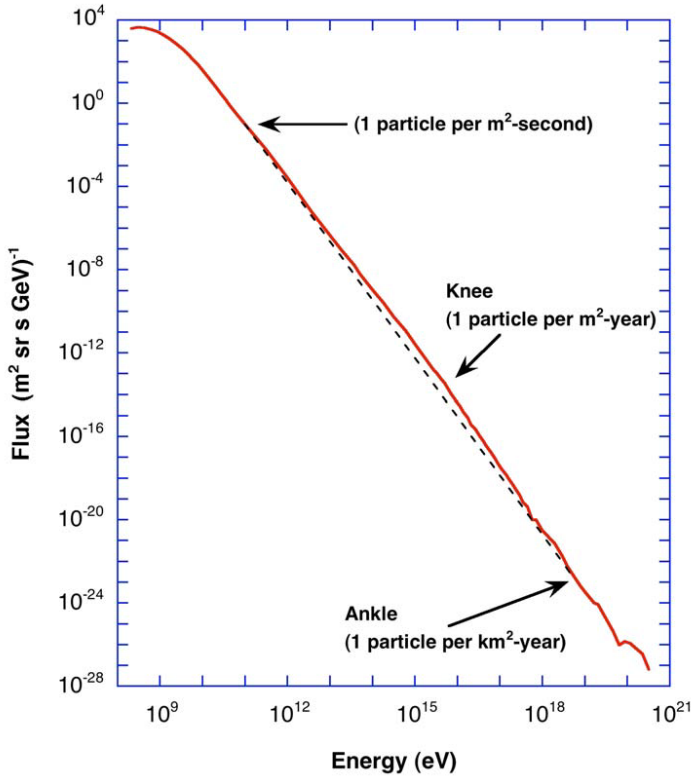


Figure 7. The spectrum of galactic cosmic rays has a slope of approximately -2.75 from ~10^{10} eV to near the "knee" region at ~10^{16} eV, after which it steepens to -3.1 (adapted from Ref. 43). The "ankle" region beyond 10^{18} eV is thought to represent an extra-galactic component. After taking into account the energy dependence of cosmic-ray escape from the Galaxy, the cosmic-ray spectrum accelerated by supernova shocks is found to have a power-law index[43] of ~ -2.1 to -2.3.

mary GCR species from ~10^{10} to ~10^{14} eV/nucleon appear to be consistent with a single common source spectrum that is slightly steeper[43] than -2.

In both instances the power-law regions end with a "knee". In the heliosphere, such knees are observed in many of the largest SEP events at a (species-dependent) location corresponding to the maximum rigidity to which the CME-driven shock can efficiently accelerate particles. Similarly, the "knee" in the all-particle GCR spectrum is believed to represent the maximum energy to which supernova shocks can efficiently accelerate cosmic rays[44]. In the heliosphere, there is an ankle in the spectrum above which the GCR component begins to dominate. In the GCR spectrum there is an ankle at ~10^{18} – 10^{19} eV, where it is believed that an extra-galactic component takes over[45].

It is thought that shock-acceleration processes make the dominant contributions to the GCR spectra from 10^9 to $>10^{15}$ eV/nucleon, and to the solar-maximum heliospheric spectra from ~10^4 to ~5 x 10^7 eV/nucleon. During the last solar maximum, improved instrumentation on missions like ACE, SOHO, Ulysses, and Wind has provided new insights into the seed particles that are accelerated by CME-driven shocks. It appears that most of the particles accelerated to high energy come not from the solar wind, but from a pool of suprathermal particles originating from a variety of sources, including small SEP events associated with solar flares that have a composition enriched in heavy nuclei like Fe, CIRs, and the suprathermal tails on the solar wind[46]. These suprathermal particles are apparently injected more easily into the acceleration process.

Similarly, at the termination shock there is a variety of seed populations available including solar wind and several types of pickup ions and ions accelerated in CIRs. Probably, some fraction of all of these are further accelerated at the termination shock, but interstellar pickup ions are the only component that is unambiguously identified, and they also seem to be efficiently accelerated in the inner heliosphere[37]. Perhaps similar selection processes are at work in cosmic-ray sources. In the present picture of cosmic-ray acceleration in super-bubbles[22], it is a mixture of seed populations that is accelerated, including SN ejecta from previous supernova, W-R wind, interstellar gas, and sputtered particles from high-speed dust grains. If processes like those in the heliosphere are at work, there may well be suprathermal tails formed on W-R winds that facilitate the acceleration of these ions by subsequent supernova shocks.

Future Outlook

The topics discussed above continue to be areas of active experimental and theoretical investigation. Improved models for W-R contributions to cosmic rays have recently investigated the importance of stellar mass and rotation, and the agreement between theory and observation continues to improve, including comparisons of additional isotope ratios[15]. Measurements of trans-iron cosmic rays with $30 \leq Z \leq 96$ can further test whether cosmic rays originate in super-bubbles[47]. Calculations of cosmic-ray modulation have invoked evermore sophisticated models to understand how the three-dimensional heliosphere modulates cosmic rays over the solar cycle, and to investigate what effects cosmic rays have on the heliosphere[48]. Over the next few years, data from the Voyagers will undoubtedly shed light on the nature of particle acceleration processes at the termination shock, but it may still be difficult to isolate contributions from the competing seed populations without charge-state and composition measurements of suprathermal ions in the outer heliosphere. However, while working towards an Interstellar Probe mis-

sion, there remain ample opportunities for both experimental and theoretical progress on the sources, radial evolution, and acceleration of suprathermal ions closer to the Sun. Finally, discovery of the origin of the ubiquitous suprathermal tails on all solar-wind species may be a key to understanding particle injection into shock-acceleration processes on a wide range of scales.

In summary, the interdisciplinary nature of the ISSI workshops on these topics has led to new insights and avenues of investigation into how the complex interplay between solar/stellar winds, neutral particles, magnetic fields, dust and explosive events on the Sun and in the Galaxy, results in the acceleration of a small fraction of the particles to very high energy.

References

1. L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger & K.-P. Wenzel (Eds.), "Cosmic Rays in the Heliosphere", Space Science Series of ISSI, Vol. 3, Kluwer, Dordrecht 1998, and *Space Sci. Rev.*, **83**, Nos. 1-2, 1998.
2. J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger & R. Kallenbach (Eds.), "Cosmic Rays and Earth", Space Science Series of ISSI, Vol. 10, Kluwer, Dordrecht 2000, and *Space Sci. Rev.*, **93**, Nos. 1-2, 2000.
3. R. Diehl, E. Parizot, R. Kallenbach & R. von Steiger (Eds.), "The Astrophysics of Cosmic Rays", Space Science Series of ISSI, Vol. 13, Kluwer, Dordrecht 2001, and *Space Sci. Rev.*, **99**, Nos. 1-4, 2000.
4. M.E. Wiedenbeck *et al.*, in Ref. 3, p. 15, 2001.
5. See, e.g., R.A. Mewaldt *et al.*, in Ref. 3, p. 27, 2001; N.E. Yanasak *et al.*, *Astrophys. J.,* **563**, 768, 2001.
6. J. Geiss, G. Gloeckler & C. Charbonnel, *Astrophys. J.*, **578**, 862, 2002.
7. M.E. Wiedenbeck *et al.*, "Refractory Nuclides in the Cosmic-Ray Source", Proc. 28th Internat. Cosmic Ray Conf., Vol. 4, 1899, 2003.
8. E. Anders & N. Grevesse, *Geochim. Cosmochem. Acta,* **53**, 197, 1989.
9. W.R. Binns *et al.*, in "Solar and Galactic Composition", AIP Conf Proc. 598, R.F. Wimmer-Schweingruber (Ed.), p. 257, 2001.
10. M. Casse & J.A. Paul, *Astrophys. J.*, **258**, 860, 1982.
11. Hubble Space Telescope image by Y. Grosdidier *et al.*, STScI and NASA.
12. Chandra X-ray image by U. Hwang *et al.*, *Astrophys. J.,* **615**, L117, 2004.
13. N. Prantzos, M. Arnould, J.P. Arcoragi & M. Casse, "Neutron-Rich Nuclei in Cosmic Rays and Wolf-Rayet Stars", Proc. 19th Internat. Cosmic Ray Conf., La Jolla, Vol. 3, p. 167, 1985.
14. G.M. Meynet, M. Arnould, G. Paulis & A. Maeder, in Ref. 3, p. 73, 2001.
15. W.R. Binns *et al.*, *Astrophys. J.,* to be submitted, 2005.
16. L. O'C. Drury *et al.*, in Ref. 3, p. 329, 2001.
17. D. Muller *et al.*, in Ref. 3, p. 353, 2001.

18. W.I. Axford, in "Acceleration of Cosmic Rays by Shock Waves", Proc. 17th Internat. Cosmic Ray Conf., Vol. 12, p.155, 1981.
19. M.E. Wiedenbeck *et al*., *Astrophys. J.,* **523**, L61, 1999.
20. A. Maeder & G. Meynet, *Astron. Astrophys*., **278**, 406, 1993.
21. J.P. Meyer, L. O'C Drury & D. Ellison, *Astrophys. J.,* **487**, 182, 1997.
22. J.C. Higdon & R.E. Lingenfelter, *Astrophys. J.,* **590**, 822, 2003.
23. L.A. Fisk, B. Kozlovsky & R. Ramaty, *Astrophys. J.,* **190**, L35, 1974.
24. M.E. Pesses, D. Eichler & J. R. Jokipii, *Astrophys. J*. **246**, L85, 1981.
25. M.A. Shea & D.F. Smart, in Ref. 2, p. 229, 2000.
26. J. Kota & J.R. Jokipii, *Astrophys. J.,* **265**, 573, 1983, and references therein.
27. F.B. McDonald, B. Heikkila, N. Lal & E.C. Stone, *J. Geophys. Res.,* **105**, No. A1, 1, 2000.
28. C.D. Steenberg, H. Moraal & F.B. McDonald, in Ref. 1, p. 269, 1998.
29. A.C. Cummings, E.C. Stone & C.D. Steenberg, *Astrophys. J.,* **578**, 194, 2002.
30. M.S. Potgieter, in Ref. 1, p. 147, 1998.
31. S.M. Niebur *et al*., *J. Geophys. Res.,* **108**, No. A10, 8033, doi:10.1029/2003JA009876, 2003.
32. R.A. Mewaldt, B. Klecker & A.C. Cummings, in Ref. 1, p. 261, 1998.
33. J.R. Jokipii & J. Giacalone, in Ref. 1, p. 123, 1998.
34. P. Evenson, in Ref. 1, p. 63, 1998.
35. B. Klecker, R.A. Mewaldt, M. Oetliker & R.A. Leske, in Ref. 1, p. 299, 1998.
36. D.V. Reames, *Astrophys. J.,* **518**, 473, 1999.
37. G. Gloeckler & J. Geiss, in "Solar and Galactic Composition", AIP Conference Proc. No. 598, R.F. Wimmer-Schweingruber (Ed.), AIP, Melville, NY, p. 281, 2000.
38. J.D. Slavin & P.C. Frisch, *Astrophys. J.,* **565**, 364, 2002.
39. G. Gloeckler, L.A. Fisk, T.H. Zurbuchen & N.A. Schwadron, in "Acceleration and Transport of Energetic Particles Observed in the Heliosphere", AIP Conference Proc. No. 528, R.A. Mewaldt *et al*. (Eds.) AIP, Melville, NY, p. 221, 2000.
40. N.A. Schwadron, M. Combi, W. Huebner & D.J. Comas, *Geophys. Res. Lett*., **29**, 20, doi:10.1029/2002GL015829, 2002.
41. R.A. Mewaldt, *Adv. Space Res*., **23**, 541, 1999.
42. R.A. Mewaldt *et al*., in "Solar and Galactic Composition", AIP Conf. Proc. No. 598, R.F. Wimmer-Schweingruber (Ed.), AIP, Melville, NY, p. 165, 2001.
43. S.P. Swordy, in Ref. 3, p. 85, 2001.
44. P.O. Lagage & C.J. Cesarsky, *Astron. Astrophys*., **125**, 249, 1983.
45. D. Muller, in Ref. 3, p. 105, 2001.
46. G.M. Mason, in Ref. 3, p. 119, 2001.
47. R.E. Lingenfelter, J.C. Higdon, K.-L. Katz & B. Pfeiffer, *Astrophys. J*., **591**, 228, 2003.
48. V. Florinski, G.P. Zank & N.V. Pogorelov, *J. Geophys. Res*., **108**, No A6, 1228, doi:10.1029/2002JA009695, SSH 1-1, 2003.
49. We are grateful to ISSI for the hospitality and opportunities they provided during our visits to Bern. We thank M.E. Wiedenbeck, W.R. Binns. M.A. Shea and S.P. Swordy for assistance with Figures 1, 3, 4, and 7, respectively. This work was supported by NASA under grants NAG5-6912 and NAG5-12929 at Caltech, and by NASA Grant 44A1055749 at the University of Maryland.

The Long-Term Variability of the Cosmic Radiation Intensity at Earth as Recorded by the Cosmogenic Nuclides

K.G. McCracken[a], J. Beer[b] and F.B. McDonald[a]

[a]*IPST, University of Maryland, USA*
[b]*Swiss Federal Institute for Environmental Science and Technology*

Introduction

Long-term instrumental measurements of the 1-100GeV/nucleon cosmic radiation intensity commenced in the 1930s using ionization chambers[1], and in 1951 using neutron monitors[2]. Together, they showed that the cosmic-ray intensity at Earth varies in response to short-term (<1 year) and the 11-year changes in solar activity. B. Peters[3] predicted in 1955 that the production of the cosmogenic isotopes in the Earth's atmosphere, and their subsequent storage in terrestrial archives, had provided a record of the cosmic-ray intensity prior to the commencement of instrumental measurements. The most abundant cosmogenic nuclides, ^{10}Be and ^{14}C, with half-lives of 1.5×10^9 and 5730 years, respectively, are well-suited to this purpose, and by the mid-1990s there were comprehensive archives of both nuclides extending over the past 50,000 years (^{14}C in tree rings and marine and lake sediments) and several over 100,000 years (^{10}Be in polar ice cores and sediments).

During the ISSI workshop "Cosmic Rays and Earth"[4] on the worldwide neutron monitor network, in 1999, it was emphasised that the cosmogenic radionuclides constitute a natural form of neutron monitor[5]. This and other work[6] was instrumental in establishing an analytical base that now allows the ^{10}Be and ^{14}C data to be used to investigate the temporal variability of the cosmic radiation in the past. Other workers[7,8] provided further insight into the geomagnetic and other effects in the cosmogenic data. In particular, it was shown that the ^{10}Be data provide a measurement of cosmic rays of lower energy than in the case of the neutron monitor (^{10}Be peak response ≅ 1.8 GeV/nucleon, compared to 6 GeV/nucleon for a high-latitude, sea-level neutron monitor).

Based on these several advances, the cosmogenic data are now used to investigate the levels of solar activity, the state of the heliosphere, and the manner in which the interplanetary magnetic field has changed over the past millenium[9-11], and they are being used as a proxy for the interplanetary field and in climate studies by others[12]. Recognising this widespread interest, a workshop "^{10}Be, ^{14}C, the Sun, and the Heliosphere" was held at ISSI in 2003 and brought together several key scientific communities (specialists in cosmogenic radionuclides, cosmic-ray modulation, cosmic-ray and solar theory). This workshop further stimulated investigations into the temporal variability of the cosmic-ray intensity at Earth in the pre-instrumental era, as outlined in the following.

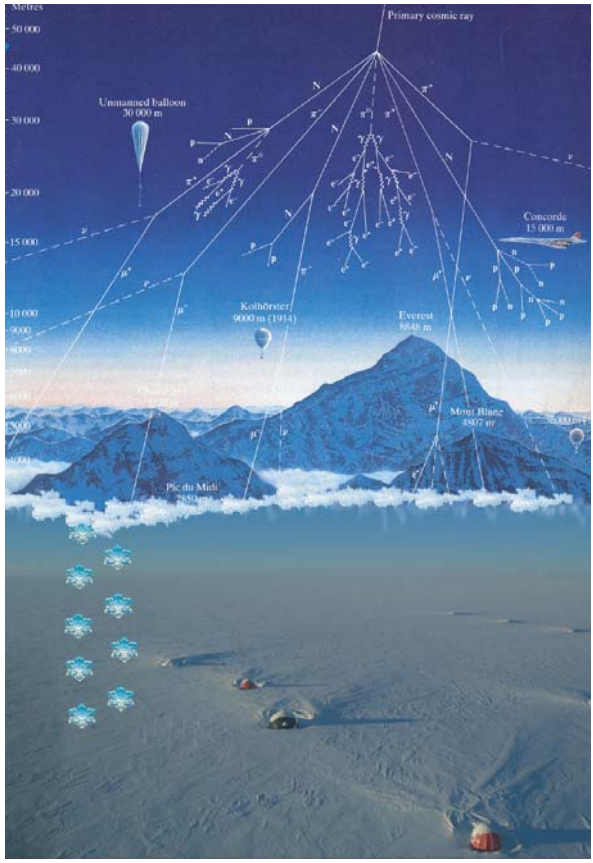


Figure 1. Illustrating the nucleonic cascade and atmospheric processes that generate the cosmogenic nuclides observed on Earth. The "primary" cosmic ray that has reached Earth from the Galaxy interacts with atmospheric nuclei to produce nucleonic and electromagnetic cascades. The neutrons and protons then interact with other atmospheric nuclei to yield the cosmogenic nuclides ^{10}Be (half life 1.51×10^6 years); ^{14}C (5730 years) and others. The ^{10}Be is precipitated to Earth with snow in the polar regions, and is then compacted into ice, which is then sampled by a drilling system. The ^{14}C remains in the atmosphere as CO_2 and participates in the biological carbon cycle.

The Cosmogenic Nuclides as Cosmic-Ray Detectors

The underlying principle is the same for both the instrumental, and cosmogenic methods of recording the cosmic-ray intensity. In the instrumental cosmic-ray detectors (e.g. ionization chamber, neutron monitor, etc.), the secondary cosmic-ray mesons or nucleons produce electronic responses in the detector, and these are summed to yield the total cosmic-ray flux over an appropriate period of time. In the case of the cosmogenic nuclides[5], the cosmic rays undergo nuclear interactions with the nuclei of the gases in the Earth's atmosphere, yielding radionuclides such as ^{10}Be and ^{14}C that are not otherwise present on Earth (Fig. 1). Since no on-line recording of the interaction is possible (as in an instrumental detector), a memory is needed which reliably stores this information and which can be read out centuries or millennia later. This recording function is provided by the deposition of the cosmogenic ^{10}Be in snow in the polar caps and in ocean sediments, and by the uptake of cosmogenic ^{14}C in biological materials such as tree rings. In principle, the concentration of the cosmogenic nuclide then yields the cosmic-ray intensity, at the time in the past determined by an independent means. In the case of ^{10}Be in ice, the time scale is established by several means (e.g. the annual variations in the isotope ^{18}O and dust trapped in the ice; volcanic time markers, and ice flow models); in the case of ^{14}C sequestered in trees, by counting the annual rings.

Figure 2. Illustrating the first stages of the measurement of ^{10}Be concentration in ice. A small drilling system used to recover a short (~300 m) core in Greenland is shown. The ice core was then protected against contamination and, following sample preparation, analysed using an accelerometer mass spectrometer.

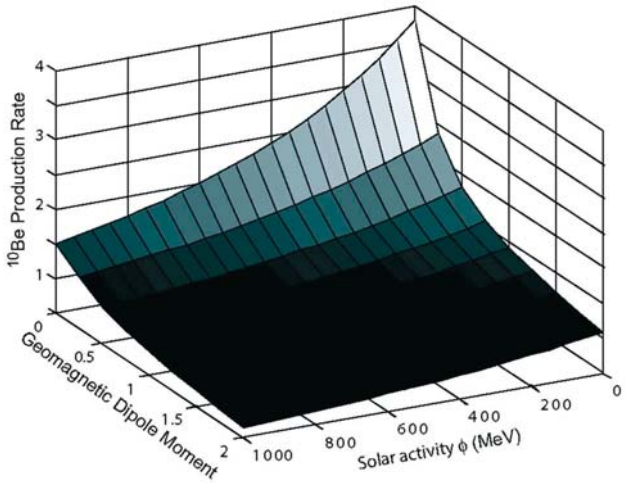


Figure 3. Calculated relative mean global production rate of ^{10}Be in the atmosphere as a function of solar activity expressed in terms of the modulation potential Φ, and the dipole moment of the geomagnetic field, relative to its present value. Φ = 0 MV corresponds to a very quiet Sun, Φ = 1000 MV to a very active one. The production rates are expressed relative to 0.018 ^{10}Be atoms cm^{-2} s^{-1} (from Ref. 6).

Figure 2 illustrates the measurement procedure used for ^{10}Be. First, a drilling machine obtains an ice core that can range from <200 m to >2000 m in length. The ice core is then divided into lengths corresponding to 1-8 years of snow accumulation, and the ^{10}Be extracted by chemical means from each sample[5]. The concentration of the ^{10}Be is then determined by using accelerator mass-spectrometry, and expressed as the number of ^{10}Be atoms per gram of ice.

The instrumental measurements since 1936 have shown that the cosmic-ray intensity is strongly affected ("modulated") by the strength and other properties of the interplanetary magnetic field and the solar wind. A quasi-theoretical quantity, the cosmic-ray modulation potential, Φ, has been defined that provides a useful quantisation of these effects[13]. Since the commencement of neutron monitor measurements in 1951, Φ has varied between 500 MV near sunspot minimum and 1200 MV near sunspot maximum. The geomagnetic field deflects the charged cosmic rays and thus prevents lower energy cosmic rays from reaching the top of the temperate and equatorial atmosphere. Consequently, the substantial variations in the strength of the geomagnetic field (±30-40%) in the recent past have had a strong effect upon the cosmic-ray intensity at Earth as summarised[6] by Figure 3. This shows that the ^{10}Be mean global production will vary by a factor of up to 10, for $0 < \Phi < 1000$ MV, and for the variability in the geomagnetic field determined from archaeomagnetic and paleomagnetic data (relative to the present dipole moment of the geomagnetic field).

The cosmogenic ^{14}C data exhibit a "memory" effect due to the long-term storage of the ^{14}C in the oceans and biosphere (the carbon cycle), and as a consequence they do not provide a direct measure of the cosmic radiation intensity at a specific time in the past[14]. In effect, the carbon cycle acts as a low-pass filter, whose response varies approximately as the reciprocal of frequency. After allowance for this frequency response, the ^{14}C data provides a sensitive record of the periodicities in the cosmic radiation in the past. A mathematical model of the carbon cycle can also be used to "invert" the observed ^{14}C data, yielding estimates of the rate of production of ^{14}C (and hence the cosmic-ray flux) as a function of time.

The cosmogenic ^{10}Be data have a much shorter memory effect, because the ^{10}Be is removed from the atmosphere after a mean residence time of about 1 year. As a consequence, they can be used directly as measurements of the cosmic-ray intensities in the past. Together, the ^{10}Be and ^{14}C data provide two independent measurements of the cosmic-ray intensity over extended periods of time. Different atmospheric processes are involved in storing the ^{10}Be in polar ice, and the ^{14}C in biological material, and intercomparison of the two methods allows solar and geomagnetic effects to be distinguished from atmospheric transport and climatic processes. In summary, the two cosmogenic nuclides, ^{10}Be and ^{14}C, exhibit a number of different characteristics, and when used together provide a quantitative understanding of the time dependence of the cosmic radiation in the past.

There are several external factors that introduce systematic and random errors into the cosmogenic measurements of the cosmic-ray intensity[6,7], and as a consequence the errors in the cosmogenic nuclide data are considerably larger than in modern instrumental cosmic-ray data. However, the primary role of the cosmogenic data is the investigation of the cosmic-ray intensity changes over time scales ranging from decades to millennia and above. Taking the case of 22-year averages of the ^{10}Be data, it has been shown that the standard deviation of individual 22-year averages is ~4.4%[15], while the decadal and century-scale variations due to solar processes are in the range 30 - 40% (see Fig. 4). Thus, while there are considerable errors in the data compared to their modern instrumental counterparts, the ^{10}Be data provides a very good "signal to noise" ratio for studying the long-term variability of the cosmic-ray intensity.

Against that background, we will now outline some of the advances that have occurred in recent years in the use of cosmogenic data to investigate the historic variations in the cosmic-ray intensity at Earth. In large part, these advances were refined and stimulated by the inter-community workshop "^{10}Be, ^{14}C, the Sun, and the Heliosphere" at ISSI in 2003.

The Cosmic-Ray Intensity during the Last Millennium

The top two panels of Figure 4 display the ^{10}Be data from Dye 3, Greenland, and the South Pole, for the interval 850 - 1950 AD. To emphasise the long-term secular changes, these data are 22-year running averages, spaced every 7 years[15]. While there are differences in the detailed variations in Greenland and Antarctica, it is clear that there is general agreement between the century-scale

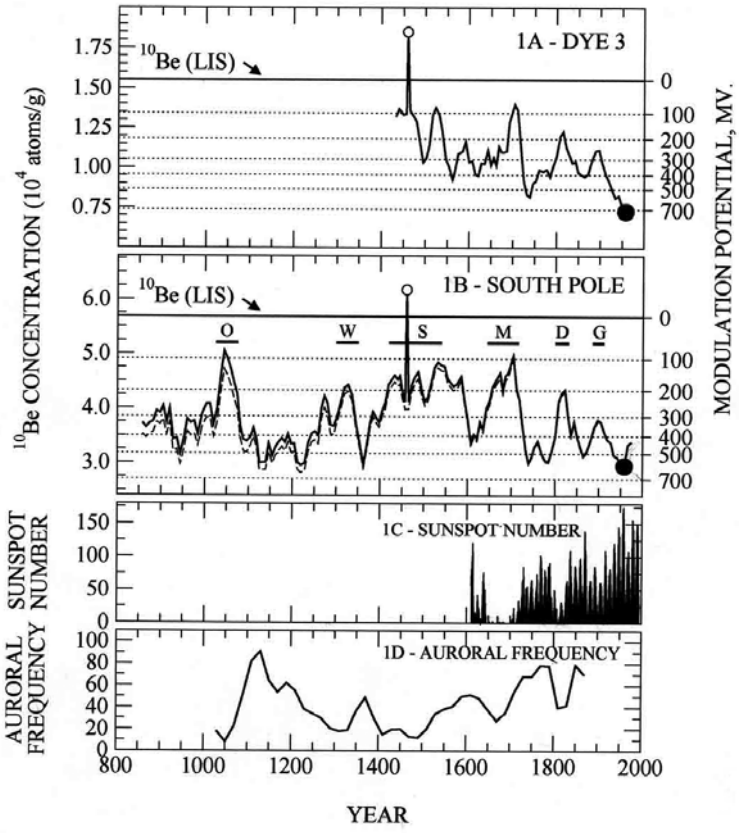


Figure 4. The temporal variation in the concentration of cosmogenic ^{10}Be observed in the Arctic ("Dye 3"), and the Antarctic ("South Pole"), for the interval 850 - 1950 AD, after compensation for the secular changes in the geomagnetic field. The concurrent sunspot and auroral record given in the lower panels show the manner in which solar activity has changed throughout the interval, as discussed in the text. The periods of low solar activity and their counterparts in the ^{10}Be records are identified as follows: O = Oort minimum; W = Wolf minimum; S = Spoerer minimum; M = Maunder minimum; D = Dalton minimum; and G = the Gleissberg minimum of ~1895. The lines labelled ^{10}Be(LIS) represent the estimated ^{10}Be concentrations in the absence of any solar modulation[15]. The right-hand scale gives the "modulation potential", Φ, as discussed in the text. The high points near 1460 are discussed in the caption to Figure 5 (from Ref. 15).

changes in the two records. As discussed above, we will now regard the ^{10}Be data as a measurement of the cosmic-ray intensity at Earth in the vicinity of 2 GeV/nucleon.

The historical sunspot record, and the mid-latitude auroral record are also displayed in Figure 4. As first inferred from ^{14}C data[16], the cosmic radiation intensity was high during the Maunder (1645 - 1715) and Dalton (1810 - 30) "grand minima" in the sunspot record, while it was ~40% lower during the intervening period of high solar activity. The rate of occurrence of mid-latitude aurora in the bottom panel of Figure 4 reflects the level of solar activity; modern experience shows that low frequencies of occurrence correlate with low sunspot numbers, and reduced solar activity. Figure 4 shows that the auroral activity was low in the vicinity of 1050 AD, 1320 AD and 1420-1500 AD (called the Oort, Wolf, and Spoerer grand minima, respectively), and that the ^{10}Be data attained high values similar to those during the Maunder and Dalton minima. In summary, Figure 4 shows that the cosmic-ray intensity has been high when solar activity was low, and that the intensity was ~40% lower during the intervening periods of higher sunspot numbers, and higher solar activity.

The geomagnetic field decreased[17] by ~25% in the period shown in Figure 4, and as a consequence, the ^{10}Be production rate increased by ~8.5 % throughout this period[15]. The dashed line in Figure 4 gives the observed concentration of ^{10}Be, while the solid line gives the ^{10}Be data after correction for the effects of these changes in the geomagnetic field. The solid line shows that there was no statistically significant change in the maximum cosmic-ray intensity at Earth between the maxima in 1050 AD and 1700 AD. That is, within the accuracy of these data, there is no evidence to suggest that the interstellar cosmic-ray intensity (i.e. outside the influence of the Sun's magnetic field) has changed over this interval[15].

The right-hand scales of Figure 4, and the dotted lines, indicate the manner in which the "modulation potential", Φ, has varied over time. The 22-year average values of Φ achieved low values (~100 MV) during the Oort, Spoerer, and the last part of the Maunder "grand solar minima", while the residual modulation during the Wolf and Dalton grand minima was substantially higher, in the vicinity of 200 MV. These results indicate that the cosmic-ray intensity at Earth during the Oort, Spoerer and last part of the Maunder minima approximated that in interstellar space, and that the interplanetary fields had little modulating effect upon the cosmic radiation incident on the heliosphere.

Since 850 AD the 22-year average cosmic-ray intensity (as measured by the ^{10}Be concentration) has returned repeatedly to low values that are similar to those of the present epoch (i.e. since 1950). Thus the ^{10}Be concentration at the South Pole

in Figure 4 exhibits minima within ±2% of 3.00×10^4 atoms/g for the 22-year averages centred on 940, 1132, 1220, 1360, 1740, and 1958 AD. This remarkable result indicates that the modulation process, and by inference, the properties of the heliospheric magnetic field, were similar during many of the periods of high solar activity between 850 and 1958. This may indicate that the interplanetary magnetic field near Earth is presently near an asymptotic value that it has approached on five previous occasions in the past 1150 years[26].

Figure 4 shows that the neutron-monitor era (1951-date), and the satellite era (1963-date) both represent one of the most extreme cosmic-ray modulation events in the past 1150 years[15,18]. Thus while the satellite and other cosmic-ray data provide us with a very detailed knowledge of the present-day three-dimensional heliosphere, this is not the typical condition of the heliosphere over the past millennium. This emphasizes the role of the cosmogenic nuclide data, in that they provide us with the means to explore the cosmic-ray modulation processes (and the level of solar activity) over time-scales of thousands of years, and during times when the heliosphere was significantly different from the present epoch[18].

The Cosmic-Ray Intensity during the Spoerer Grand Minimum

The cosmic-ray intensity maximum in the interval 1420 - 1540 (the Spoerer minimum) was the most prolonged in the past 1150 years (see Fig. 4). We study this interval in greater detail, to better understand the cosmic-ray modulation effects during a prolonged "grand minimum" of solar activity[15]. In effect, we use it as a "virtual laboratory" that allows us to investigate the quiet Sun. Figure 5 presents the annual ^{10}Be data from Greenland for the interval, 1420 - 1540. There was persistent and relatively large amplitude (~25%) modulation of the cosmic radiation throughout the whole interval. The persistent ~25 year repetition indicates the continuation of the 22-year periodicity in the solar magnetic field. Note also the 11 and ~5-year repetitions; the latter having been reported previously[19,20]. Annual measurements of tree ring ^{14}C exhibit an 11-year variation that confirms the presence of this modulation at this time[21].

The cosmic-ray intensity first attained a value close to the interstellar value (the line marked ^{10}Be(LIS)) at the beginning of the Spoerer Minimum, and then returned repeatedly to values that are statistically consistent with that value until 1530. These observations indicate that: (1) the heliosphere repeatedly returned to a condition of very low residual cosmic-ray modulation throughout the

Spoerer Minimum, 1420 - 1540, and (2) the cosmic radiation experienced solar control at both the 11- and 22-year periodicities. These cosmogenic cosmic-ray data provide a clear indication that solar activity continued in an episodic manner throughout this grand minimum, and provide a better insight into that period than is possible using the scanty historical sunspot and auroral records from that era. The continuation of cosmic-ray modulation through a portion of the Maunder minimum (1645 - 1715) has been established as well[15,18,22]. These results emphasise the important property of the cosmogenic cosmic-ray data to provide the means to investigate the 11- and 22- year solar variability far into the past.

The Cosmic-Ray Variability over the Past 10 000 Years

The ^{10}Be data in Figures 4 and 5 have allowed the cosmic-ray variations recorded in the cosmogenic record to be compared with those in the instrumental record. The availability of a detailed sunspot record since 1610; and sparse sunspot and auroral records prior to that, have allowed the relationship between solar activity and the cosmic radiation intensity to be established with some confidence. In general, (a) the highest cosmic-ray intensities, which indicate a 22-year average modulation potential of ~100 MV, correspond to extended periods

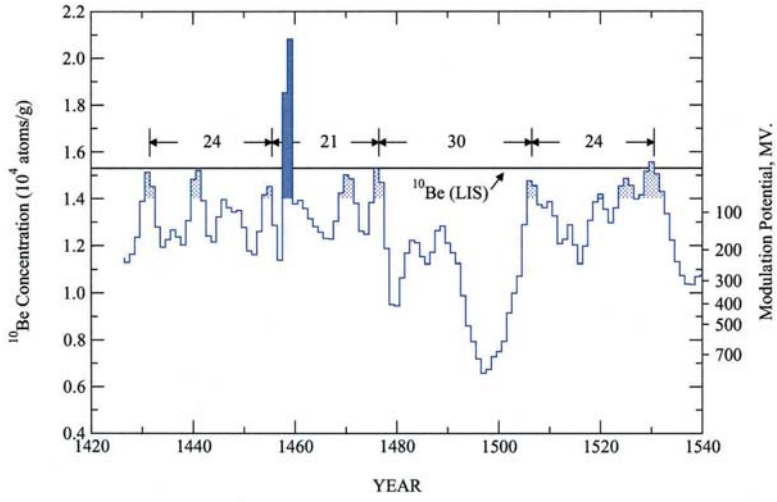


Figure 5. The ^{10}Be concentration at Dye 3, Greenland, corresponding to the Spoerer minimum in solar activity. The line labelled ^{10}Be(LIS) represents the estimated ^{10}Be concentration in the absence of any solar modulation[15]. The right-hand scale gives the modulation potential derived from these data, and it shows that the heliospheric field continued to modulate the cosmic radiation throughout this whole period. The two extreme points in the vicinity of 1460 are proposed to be due to the production of cosmic rays by the Sun, or by gamma rays from a nearby supernova (from Ref. 15).

(> 50 years) of low solar activity; while (b) the lowest cosmic-ray intensities, corresponding to a 22-year average modulation potential of ~700 MV, are associated with a very active Sun, as in the modern epoch 1950-2000. This provides a "calibration" that allows us to use the ^{10}Be and ^{14}C archives to study both the cosmic radiation intensity, and the degree of solar activity in the past, and to compare it to the past 1000 years summarised in Figure 4. Stimulated, in part, by the 2003 ISSI workshop, there is now active interest in using cosmogenic records to investigate the cosmic-ray (and thence, solar) effects in the past. The following is an outline of the progress made, and an indication of the progress that will be made in the near future.

The ^{10}Be signal found in ice cores reflects not only production changes, but also changes in the atmospheric transport and deposition processes. During periods when the climate is relatively stable, such as the past 12,000 years (the Holocene), these effects are generally much smaller than the production effects. This is not the case for glacial times when the accumulation rate of ice has been smaller by up to a factor of 2, yielding higher ^{10}Be concentrations. Further, the mixing effects in the oceans are believed to have changed during glacial times, thereby changing the "filtering" characteristics of the carbon cycle. For this reason, we will first consider the post-glacial period in this section, and then briefly consider part of the glacial period in the next section.

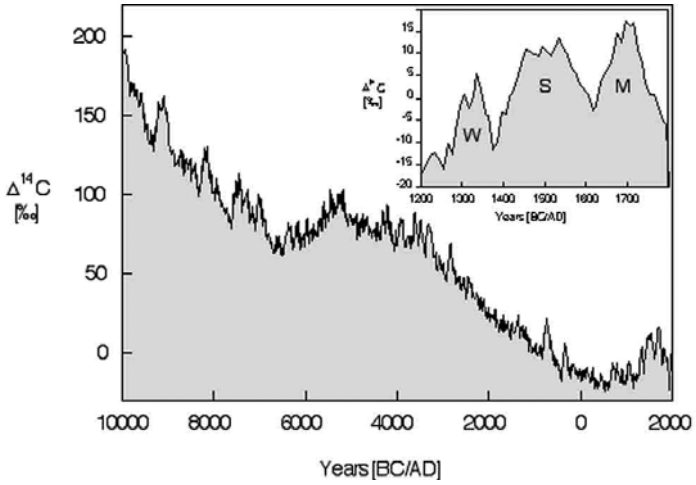


Figure 6. The cosmogenic ^{14}C data corresponding to the interval 10,000 BC to 1950 AD. The inset is for the interval 1200 - 1800 AD. Note that the century-scale variations in the inset are similar to those in Figure 4, corresponding to the Wolf (W), Spoerer (S) and Maunder (M) minima. These and the other century-scale variations are superimposed on a changing baseline due to the long-term changes in the strength of the geomagnetic dipole.

Figure 6 displays the ^{14}C data for the past 12,000 years[19]. This record consists of relatively short-term (~200 year) variations superimposed upon a long-term, slowly changing baseline. The latter is largely due to the long-term changes in the cosmic-ray geomagnetic cut-off, as a consequence of the ±30 - 40% changes in the Earth's dipole moment over the past 20,000 years.

The inset in Figure 6 displays 600 years of the ^{14}C record. It represents the same sequence of cosmic-ray fluctuations (from the Wolf to the Maunder events) that were evident in the ^{10}Be record in Figure 4, superposed upon a steady increase due to the "memory effect" discussed previously. There are other episodes of short-term (~200-year) enhancements in the 12,000 year ^{14}C record in Figure 6, of similar duration and amplitude to the Wolf-Maunder sequence in the inset. The "calibration" at the beginning of this section therefore suggests that the cosmic radiation has experienced a number of sequences of ~200-year modulation events over the past 12,000 years, similar to those in Figure 4, as a consequence of changes in the level of solar activity. Based upon this association, Figure 6 suggests that: (a) the Sun remained active for the 2000 year interval 2800-800 BC, without periods of low solar activity similar to those observed during the recent millennium (Fig. 4); (b) that by way of contrast, there were a number of "grand minima" in the 1200 year interval 4400-3200 BC; and (c) there were a number of substantially longer periods (~500 yr) of very low solar activity in the interval 9500-6600 BC.

In summary, the cosmogenic ^{10}Be and ^{14}C data now provide the means to make quantitative studies of the time variations in the cosmic radiation at Earth in the post-glacial period since 10,000 BC. The initial studies have demonstrated that there have been a number of modulation episodes similar to the Oort-Dalton sequence during that time. Several new ice cores have been drilled in recent years that include the post-glacial period, and these will allow more detailed analysis of the cosmic radiation intensities throughout this period.

The Cosmic-Ray Variations during the Past 60 000 Years

Astronomical observations indicate that the magnetic fields and gas density in interstellar space in the vicinity of Earth have varied substantially over the past 100,000 years[23]. This suggests that the cosmic-ray intensity near Earth may have varied as a consequence of the Earth's motion about the centre of our Galaxy, and if so it would provide important information regarding the structure of the Galaxy. In addition, there is the possibility that a supernova may have resulted in enhanced cosmic-ray intensity at Earth in the past. The cosmogenic measurements of the cosmic radiation allow both possibilities to be investigated.

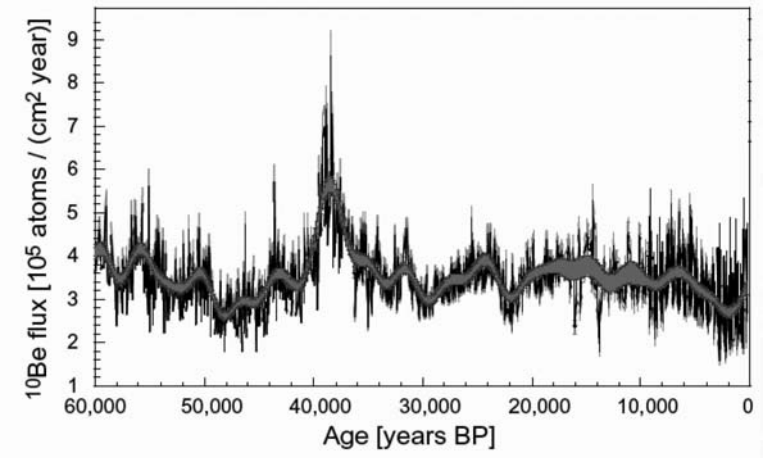


Figure 7. The ^{10}Be flux estimated from an ice core from Greenland, corresponding to the past 60,000 years BP (means years before 1950). The individual data points are 100-year averages, and the central line is the 3000-year running average of those data (from Ref. 24).

Figure 7 displays the ^{10}Be precipitation rate over the past 60,000 years. The period prior to 12,000 BP ("before present") corresponds to the last glacial epoch, when the accumulation rate of ice was smaller, resulting in higher ^{10}Be concentrations, which might be misinterpreted as higher cosmic-ray intensities. A number of the concurrent measurements in the ice cores yield estimates of the snow precipitation rate, allowing the ^{10}Be flux (i.e. the annual ^{10}Be precipitation rate) to be calculated, and this provides a measurement of the cosmic-ray intensity unaffected by the severe climate changes between the glacial and post glacial periods. Figure 7 shows that there were substantial changes in the ^{10}Be flux throughout the past 60,000 years, and we now briefly discus two of the more prominent variations: (1) the broad minimum in the vicinity of 5000 - 1000 BP; and (2) the high values in the interval 40,000 - 37,000 years before the present.

As discussed previously, Figure 3 implies that variations in the strength of the geomagnetic field result in out-of-phase variations in the cosmic-ray intensity at Earth. Paleomagnetic studies show that the geomagnetic dipole moment was low ~6,000 years ago (75% of the present value), that it attained a maximum 3,000 years ago that was 40% above the present value, and that it has declined rapidly over the past 1000 years. The relationship between the cosmic-ray intensity and the geomagnetic dipole strength is well-known (see, for example, Fig. 3), and fully explains the reduction in ^{10}Be flux in the interval 5000-1000 BP in Figure 7.

Measurements of the remnant magnetism in sea sediments have shown that the geomagnetic field reached an intensity minimum of 10-20% of its present value

about 40,000 BP, and that it was close to reversing its polarity for a period of ~5000 years[24] (called the "Laschamp magnetic event"). Figure 3 indicates that the ^{10}Be flux to Earth would therefore increase by a factor of 2 - 2.5, which is consistent with the observations in Figure 7. Both of the variations in the ^{10}Be flux considered above are therefore consistent with independent paleo-magnetic measurements, and other mechanisms are not required to explain them. Nevertheless, data such as in Figure 7 will be important to set limits on the effects due to changes in the Earth's interstellar environment, supernova explosions, interstellar shock waves, and extended periods of low solar activity in the past.

Inferred Properties of the Interplanetary Magnetic Field and Solar Activity in the Past

The majority of the cosmic radiation observed at Earth originates in the Galaxy, and reaches Earth after propagating through magnetic fields of solar origin that extend to the limits of the heliosphere, some 100 - 150 Astronomical Units from the Sun (1 AU = the Sun–Earth distance). The "cosmic-ray transport equation" describes the cosmic-ray propagation processes in terms of the properties of the heliomagnetic field, and the speed of the solar wind, and this shows that the intensity of the galactic cosmic radiation at Earth is determined by these several properties[25]. That is, the cosmic-ray intensity at Earth can be regarded as a measurement of the integrated properties of the heliospheric magnetic fields. Since the cosmogenic ^{10}Be data constitutes a measurement of the cosmic-ray intensity, they can be used to investigate the properties of the heliospheric magnetic field in the past. The initial studies of this field were based on extrapolations from the present, using statistical regressions between the sunspot number and ^{10}Be data to guide the extrapolation[12].

By inverting Parker's cosmic-ray transport equation[25], the cosmogenic ^{10}Be data given in Figure 4 have been used recently[26] to investigate the strength of the heliospheric field at Earth since 850 AD. These studies show that the 22-year average magnetic field was lowest (2 - 3.75 nanotesla, nT) during the Oort (1050 AD), Spoerer (1420 - 1540 AD) and the latter part of the Maunder minima. During each of the periods of high solar activity since 850 AD (Fig. 4), the 22-year average field was similar to the present-day value (~6 nT). Satellite measurements show that the 3-month average field has varied[29] over the range 5-10 nT since the 1960s. Together, these results suggest that the heliospheric field near Earth may vary by a factor of 3 - 5 between a grand minimum and the periods of enhanced solar activity.

A number of extrapolations of the sunspot number into the past have been made using empirical models of the correlation between sunspot number and the cosmic-ray intensity[27, 28]. Considerable progress in such studies, and companion extrapolations of the solar irradiance into the past, is to be expected using the cosmic-ray production rate based on ^{14}C, and ^{10}Be from several recently acquired ice cores.

Conclusions

Stimulated by the two ISSI workshops[4] in 1999 and 2003, there has been considerable progress towards using the cosmogenic nuclides to study the manner in which the galactic cosmic radiation at Earth has varied over historic time. These studies have determined that they represent a measurement that is broadly similar to that of a neutron monitor, but corresponding to somewhat lower cosmic-ray energy. Using ^{10}Be data, the cosmic-ray modulation potential has been estimated for the 1150-year period since 850 AD. This shows that the cosmic-ray intensity is highest during the "grand minima", approaching the intergalactic intensity that exists outside the heliosphere. The modulation effects of the heliospheric magnetic fields are considerably higher during periods of substantial solar activity. Using the cosmic-ray transport equation, the ^{10}Be data for the period 850-1950 have been inverted to yield estimates of the temporal variation of the strength of the heliospheric magnetic field at Earth. They show that the heliospheric field near Earth was in the range 2-3.75 nT during prolonged periods of low solar activity such as the Spoerer minimum, while satellite measurements show that it has been in the range 5-10 nT since 1960.

With the availability of a number of new ice cores, and further development of the methodology to interpret the cosmogenic data as cosmic-ray intensities, it is anticipated that there will be considerable advance in the study of "paleo-cosmic rays" in the near future. Further, the ^{10}Be and ^{14}C data themselves, and the inferred cosmic-ray parameters (such as the modulation potential), will be used increasingly to examine the properties of the interplanetary field, and the temporal variation of solar activity over the past millennia[28].

References

1. S.E. Forbush, *J.Geophys. Res.*, **63**, 651,1958.
2. J.A. Simpson, in Ref. 4, p. 11.
3. B. Peters, Proc. 5th Int. Conf. Cosmic Rays, Mexico, 1955, also D. Lal & B. Peters, *Progr. Elem. Cosmic Ray Phys.*, **6,** 1, 1962.

4. J.W. Bieber *et al*., "Cosmic Rays at Earth", Space Science Series of ISSI Vol 10, Kluwer Academic Publ., Dordrecht, and *Space Sci. Rev.,* **93,** No. 1-2, 2000.
5. J. Beer, in Ref. 4, p. 107.
6. J. Masarik & J. Beer, *J. Geophys. Res*., **104**, 12009, 1999.
7. K.G. McCracken, *J. Geophys. Res*., **109**, A04101, doi:10.1029/2003JA010060, 2004.
8. W.R. Webber & P.R. Higbie, *J. Geophys. Res*., **108**, 1355, doi:10.1029/2003JA009863, 2003.
9. G.C. Castagnoli & D. Lal, *Radiocarbon,* **22**, 133, 1980.
10. I.G. Usoskin *et al*., *J. Geophys. Res*., **107**, 1374, 2002.
11. G.M. Raisbeck *et al*., *Phil. Trans. Roy. Soc. London*, Series A **330**, 463, 1990.
12. M.A. Lockwood, *J. Geophys. Res*., **106**, 16021, 2001.
13. L.J. Gleeson & W.I. Axford, *Astrophys. J.,* **149**, 115, 1967.
14. H. Oeschger *et al*., *Tellus,* **27**, 168, 1975.
15. K.G. McCracken *et al*., *J. Geophys. Res*., **109**, A12103, doi:10.1029/2004JA010865, 2004.
16. M. Stuiver & P.D. Quay, *Science,* **207**, 11, 1980.
17. M.W. McElhinney & P.L. McFadden, Paleomagnetism: Continents and Oceans, Academic Press, San Diego, Calif., 2000.
18. I.G. Usoskin *et al*., *J. Geophys. Res*., **106**, 16039, 2001.
19. M. Stuiver & T.F. Braziunas, *The Holocene,* **3**, 289, 1993.
20. K.G. McCracken *et al*., *Geophys. Res. Lett*. **29**, 2161,doi:10.1029/2002GL015786, 2002.
21. H. Miyahara *et al*., *Proc. Int. Conf. Cosmic Rays*, 28th, 4139, 2003.
22. J. Beer, S. Tobias & N. Weiss, *Solar Phys*., **181**, 237, 1998.
23. P.C. Frisch, *J. Geophys. Res*., **105**, 10279, 2000.
24. R. Muscheler *et al*., *J. Quart.Sci*. in press.
25. E.N. Parker, *Planet. Space Sci*., **13**, 9, 1965.
26. R.A. Caballero-Lopez *et al*., *J. Geophys. Res*. **109**, doi: 10.1029/, 2004
27. I.G. Usoskin *et al*., *Astron.& Astrophys*., **413**, 745, 2004.
28. S. Solanki *et al*., *Nature,* **431**, 1084, 2004.
29. NASA "omnitape", http://nssdc.gsfc.nasa.gov/cohoweb/cw.html

The Sun, from Core to Corona and Solar Wind

R. von Steiger[a] and C. Fröhlich[b]

[a]*International Space Science Institute, Bern, Switzerland*
[b]*Physikalisch-Meteorologisches Observatorium and World Radiation Center, Davos, Switzerland*

Introduction

The Sun is an unspectacular star of spectral type G2 somewhere in the outskirts of our Milky Way galaxy, where there are millions or even billions of others. Yet it is the only star that can be studied in full detail, and it is the dominant body of our Solar System and the heliosphere in which it is embedded. The Sun concentrates almost all of the mass (nearly 99.9%), and it is the only relevant source of light and energy in the Solar System. Its magnetic field and tenuous outermost atmosphere, the corona and solar wind, permeate all of interplanetary space out to a distance of about 100 astronomical units, far beyond the orbits of the planets. It is therefore worth exploring this medium, not only because astronauts might want to travel there, but also because it affects all bodies in space, man-made or natural, most notably the Earth itself. This influence would be a relatively simple matter were it not for the fact that all solar output, be it radiation or particles, is variable on all time scales from less than minutes to billions of years.

Several ISSI workshops have addressed themes in this context[1-4], two of which are highlighted here. We first address the questions of why and how solar composition is being measured and what can be learned from it. Then we address the solar radiative output and its variability. We conclude by pointing out how these two seemingly independent aspects of solar physics are not only related, but literally intertwined.

Solar Composition

The Sun condensed 4.6 billion years ago from the protosolar cloud, which was probably well-mixed and had a homogeneous composition. Yet when we look around in the Solar System, each planet, moon, planetoid, comet and meteorite is composed differently: rocky inner planets, gaseous outer planets, icy comets, and so on. In order to assess these differences and how they evolved, it is essential to

know from where they started, i.e. the composition of the protosolar material. Fortunately this can be measured without travelling back in time: because the Sun represents by far the largest reservoir, it is hard to imagine how it may have been contaminated. Indeed the standard composition as assembled from both solar and meteoritic data and given by, for example, Grevesse & Sauval[5], or more recently, Lodders[6], represents the agreed baseline (proto-)solar composition.

The history of the formation and evolution of the Solar System is encoded in the composition (particularly of the isotopes) of its bodies relative to this baseline[7]. But how can the solar composition be measured? After all it is not possible to go and scoop up a solar sample and analyse it in the laboratory. There are two main methods: remote sensing and in-situ observation.

Most of what we know about solar composition (and all we know about the compositions of other stars) comes from optical remote sensing of absorption lines in the solar spectrum. Each absorption line can be attributed to a particular element, and the strength of the line is proportional to the abundance of that element. Simple as this sounds, in reality it is more complex. The ratio of the observed emission measures of two lines ideally can be taken as the abundance ratio of the emitting elements. But this is only true when the lines originate in the same volume of the solar atmosphere and when the contribution functions of the two elements to the emission measure have the same temperature dependence[8,9]. Thus all element abundances as given in standard tables are ratios relative to a normalizing element, in most cases hydrogen. Even so, the standard tables must not be considered as cast in stone. Owing to previously unrecognized contributions from other elements the oxygen abundance was corrected downwards by 20% in Grevesse & Sauval[5] from the previous standard (and since then by yet another 20%).

More seriously, some elements are not amenable to the remote-sensing method. Noble gases have spectral line energies that are too high to be excited at solar surface temperatures. Their abundances need to be determined by other methods. One novel method has become possible thanks to the long-duration, uninterrupted observations of the Sun from spacecraft such as ESA's SOHO mission, i.e. helioseismology. Like a three-dimensional drum, or bell, the Sun oscillates, and the oscillations of its surface can be observed from space; the main modes have oscillation periods of the order of 300 seconds, which is why they are called 5-minute oscillations. The frequencies of the different oscillation modes form a pattern that has been determined with unprecedented precision and accuracy using the SOHO helioseismic instruments (MDI, GOLF and VIRGO). Each mode essentially represents a sound wave, and different modes penetrate to different depths in the solar body. Therefore, helioseismology is a tool for literally

looking into the Sun's interior. This mode-frequency pattern can be compared to a synthetic pattern calculated numerically from a standard solar model[10]. The best such model pattern agrees with the observed pattern to a fantastic accuracy of a fraction of a percent[11]. Now, since one critical input to the standard solar model is the solar ratio of the two most abundant elements, helium to hydrogen, the agreement between model and observed pattern can be interpreted as a measurement of that ratio. The result is consistent with the primordial He abundance and theoretical calculations of its increment by galactic chemical evolution[12]. Unfortunately, elements heavier than helium are too rare to leave a discernible mark in the helioseismological mode-frequency pattern.

The other important method for obtaining solar abundances of many elements is by in-situ observation. Taking advantage of the fact that the Sun continually emits some of its matter in the form of a solar wind, it can be collected and/or observed anywhere in interplanetary space (anywhere outside planetary magnetospheres, that is). This can be done either by the foil collection technique, pioneered by J. Geiss on the Apollo missions to the Moon, with subsequent laboratory analysis of the exposed foils, or by flying a miniaturized mass-spectrometer on a spacecraft such as SWICS on Ulysses, or CELIAS on SOHO.

The Ulysses mission has now completed two full revolutions on its unique orbit that is almost at right angles to the ecliptic plane. The first orbit, in 1992-1998, occurred during low solar activity, whereas during the second orbit, in 1998-2004, the Sun was much more active. It was well known from solar-eclipse photographs, taken long before Ulysses, that the solar corona changes shape with solar activity, as illustrated in Figure 1. The two images from the LASCO coronagraph on SOHO show a solar minimum corona (left) and a solar maximum corona. It is obvious that the bright streamers, which emanate from behind the occulting disk looking like candle flames, are confined to the solar equator at solar minimum, but occur at all latitudes at solar maximum. The Ulysses mission expanded that picture and showed that this structure extends throughout the entire heliosphere. This is also illustrated in Figure 1 by polar plots of two parameters measured with the SWICS instrument, solar-wind speed (yellow) and the so-called freezing-in temperature derived from two charge states of oxygen (green); the latter parameter essentially indicates the temperature of the corona from where the wind originates. The striking feature in the solar minimum graph (left) is how well the heliosphere appears to be ordered: poleward of about 30 degrees it is filled with a very uniform, fast type of solar wind that originates from a relatively cool source, while equatorward of about 20 degrees a much more variable, but generally slower solar-wind type from a hotter source dominates. At in-between latitudes, the two solar-wind types alternate regularly due to solar rotation, with a remarkably sharp boundary between them[13]. The two

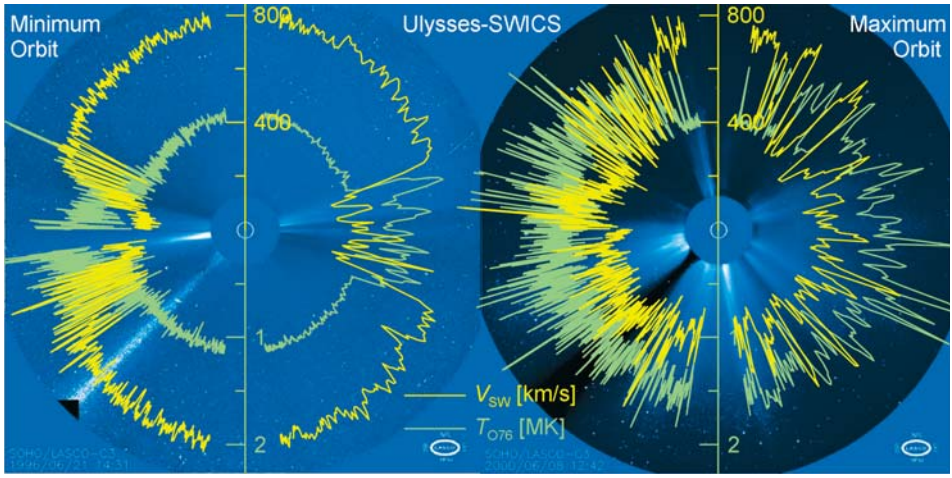


Figure 1. SOHO-LASCO images of the white-light corona taken with the C3 coronagraph, which has an occulting disk three times the apparent diameter of the Sun. Overlaid are Ulysses-SWICS polar plots of two solar-wind parameters, speed (yellow) and freezing-in temperature (green). The images were taken at solar minimum (left) and at solar maximum (right). They show that the structure of the corona, dipolar at solar minimum but "chaotic" at solar maximum, extends to the entire heliosphere. (Images courtesy of SOHO-LASCO consortium.)

types of solar wind, fast streams that originate from the cool coronal holes and slow, variable wind that originates from above the hot streamer belt, have long been known[14]. But only Ulysses in its high-inclination orbit and with its capability for continuous composition observations could reveal how remarkably well they are ordered in the heliosphere at solar minimum. Moreover, the fast solar-wind type could be shown to match solar surface composition rather closely, albeit not perfectly, and thus can be used to infer the solar composition[15].

The orderly picture at solar minimum is in stark contrast to the latitude distribution of the same two parameters during the second, solar maximum orbit of Ulysses (right). Fast and slow streams from cool and hot sources, respectively, appear to occur at all latitudes. In addition, transient events such as coronal mass ejections occur frequently at solar maximum and further complicate the picture. These events can be fast just like streams from coronal holes, but composition data can be used to unambiguously tell the difference. In fact, complicated as the mixture of parameters may look at solar maximum, composition data reveal that it's the same two types of solar wind as at solar minimum, but distributed over all latitudes and peppered with coronal mass ejections[16].

As already mentioned, the fast solar wind does not quite represent the solar surface composition, but the two are fractionated relative to each other: elements

with a low first ionisation potential (FIP) are enriched in the solar wind by a factor of 1.5 to 2. Only if this FIP fractionation is sufficiently well understood can the solar composition can be derived from the solar wind. No universal agreement exists about the nature of the fractionation process[17], but it is quite clear that it must be related to the separation of atoms and ions in the chromosphere, the only place in the solar atmosphere where neutral atoms of some elements can exist. In turn, modelling of the FIP fractionation mechanism together with observations of solar-wind element abundances can be used to learn about the conditions in the chromosphere.

Likewise, abundances of different charge states of an element can be used to infer the conditions, in particular the electron temperature, in the low and middle corona. At the base of the corona the plasma is in thermal equilibrium with the hot electrons at and above a million degrees. Then, as the solar wind begins to flow increasingly faster, the corona gets less and less dense, until finally the material "freezes in", i.e. drops out of thermal equilibrium because collisions with hot electrons become too infrequent. From that point onwards, the plasma retains the charge-state information out to large distances, where we can observe it and decode the coronal temperature from the ion charge-state ratios. This powerful tool can be used to classify the solar-wind types as explained above, and even to derive a rough temperature profile in the corona[13]. Some charge states can even be observed both by optical remote sensing as well as in-situ, as illustrated in Figure 2. The image was taken by the SUMER spectrograph on SOHO in light of wavelength 1242 Å emitted by Fe XII, or 11 times charged iron ions. Superposed is a graph taken in-situ with Ulysses-SWICS in the large polar fast streams at solar minimum. In the SUMER image the coronal hole (at lower left) looks pitch dark, while the SWICS graph shows that Fe^{11+} accounts for about 25% of all iron. This obvious contradiction needs further work, which may finally lead to a better understanding of the conditions and processes in coronal holes[18].

Solar Irradiance Variability

Conclusive evidence for solar-irradiance variability was achieved only after electrically calibrated radiometers (ECRs) were launched on space platforms to monitor the Sun more or less continuously, that is with the launch of the Earth Radiation Budget experiment with the radiometer HF on Nimbus 7 in November 1978[19]. In early 1980, the Solar Maximum Mission satellite (SMM) followed with the ACRIM-I radiometer[20], then the Earth Radiation Budget Satellite (ERBS)[21], the Upper Atmosphere Research Satellite (UARS) with ACRIM-II[22], the European Retrievable Carrier (Eureca) with SOVA[23], the Solar and

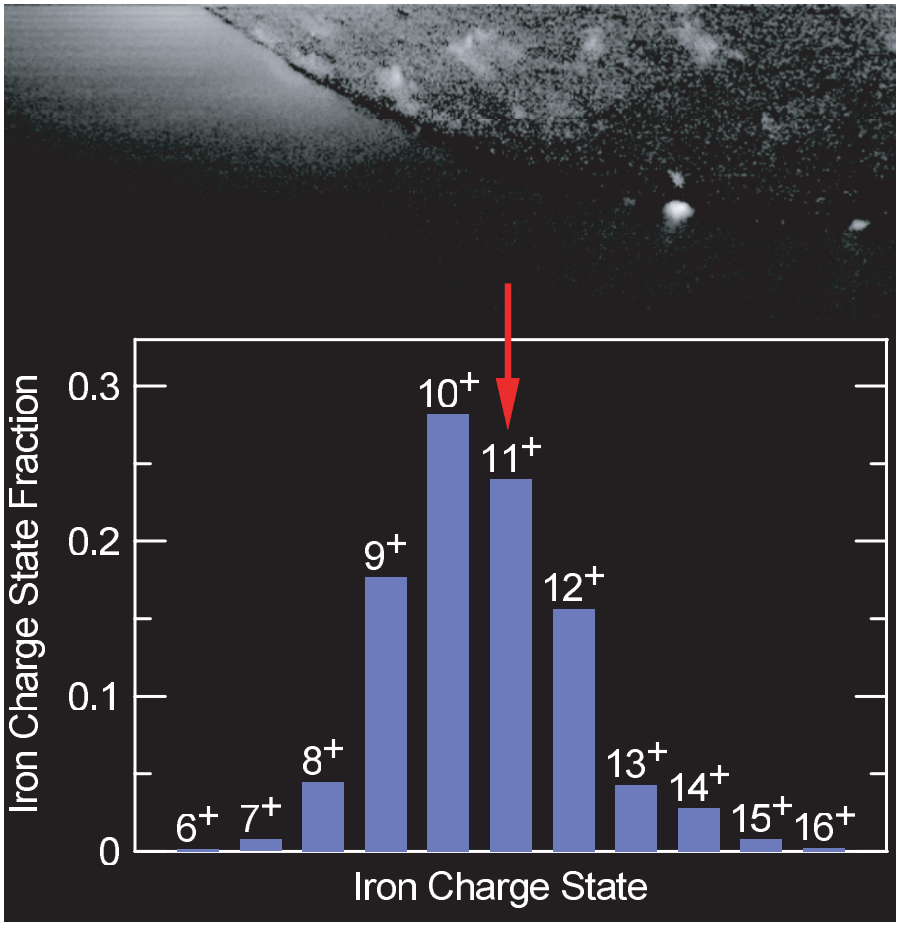


Figure 2. SOHO-SUMER Fe XII 1242 Å image taken at the boundary of the south polar coronal hole, overlaid with Ulysses-SWICS iron charge-state distribution taken at several AU. The picture illustrates how the same ion, indicated by the red arrow, can be observed both remotely and in-situ. Figure adapted from Ref. 18, image courtesy of SOHO-SUMER consortium.

Heliospheric Observatory (SOHO) with VIRGO[24,25], ACRIMSAT with ACRIM-III[26], and most recently the Solar Radiation and Climate Experiment (SORCE)[27]. Figure 3a compares the various irradiance data sets acquired from the corresponding missions. Offsets among the data sets reflect the different radiometric scales of the individual measurements. Since late 1978, at least two independent solar monitors have operated simultaneously in space.

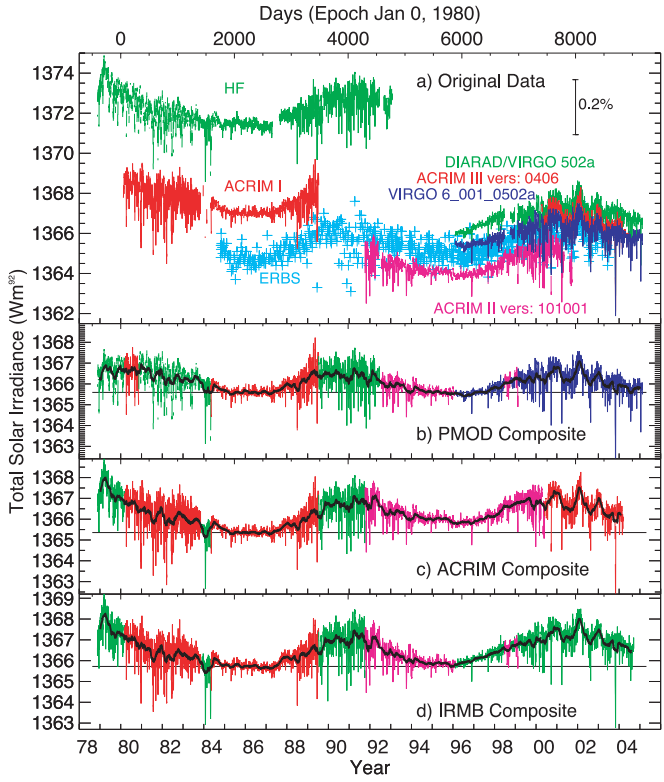


Figure 3. Compared in the top panel are daily averaged values of the Sun's total irradiance from different space platforms since November 1978, with the names of the radiometers for identification. The data are plotted as published by the corresponding instrument teams. Shown in the three bottom panels are the PMOD[24,28,29], the ACRIM[30,31] and the IRMB[32] composite irradiance time series compiled from the individual data sets together with a 81-day running average. The difference between the composites around the maximum of cycle 21 is due to the fact that only the PMOD composite corrects the HF data for degradation.

Construction of a total solar irradiance composite

Since the first presentation of a composite of total solar irradiance (TSI) by Fröhlich & Lean[33] at the IAU Symposium 185 in August 1997 and the extensive discussions during the ISSI workshop[3] the composite has not only been updated to cover now almost three solar cycles, but it has also been improved. In order to distinguish the different composites existing today, the Fröhlich & Lean[33] composite and its updates is called PMOD, the one constructed by Willson[30] and updated by Willson & Mordvinov[31]ACRIM, and the one compiled more recently by Dewitte *et al*.[32] IRMB. All three are shown in the bottom panels of Figure 3.

The description of the procedures and their updates used to construct the PMOD composite can be found in References 24, 28, 29 and 34. As only the results from the ACRIM and VIRGO radiometers have the possibility of in-flight assessment of degradation, all composites are radiometrically based on ACRIM-I. The PMOD and IRMB composites are adjusted to the Space Absolute Radiometer Reference (SARR) introduced by Crommelynck *et al*.[35], which allows comparison of the results from different space experiments. The main difference between the PMOD composite and the others are corrections applied to the HF instrument that compensate for degradation and early increase. In the most recent version, these corrections are determined for the whole period after the many slips have been removed[34]. Thus, there is no longer a need to treat the period of the gap between ACRIM-I and II separately, and the tracing of ACRIM-II to I can be performed with the independently corrected HF data. Another difference is the treatment of the degradation of ACRIM-I, which is based on the model for exposure-dependent changes, developed for the PMO6V radiometers within VIRGO. After the start of the VIRGO data, the ACRIM composite continues with data from ACRIM-II and III, whereas the PMOD and IRMB composite use VIRGO data. Note, however, that the former uses VIRGO TSI, which is determined from all available data, i.e. from both VIRGO radiometers (PMO6V and DIARAD), whereas the latter is only based on DIARAD, and thus called DIARAD TSI.

A comparison of the three composite irradiance records to ERBE is presented in Figure 4. The ERBE data cover the period from October 1984 to August 2003, so they do not show the first six years of the composite. For Figure 4 they have been corrected for the early increase with the coefficients determined for ACRIM-I and the dose received by ERBE during its <3 days total exposure time, and interestingly enough the long-term trend has been removed. Also the ACRIM and IRMB composite records use Nimbus 7 HF and ACRIM-I and II data prior to 1996 as published, which means uncorrected. The corrections of HF and ACRIM-I explain the difference with the PMOD composite during the maximum of cycle 21 and also over the ACRIM gap. The step of the ACRIM composite (Fig. 4) amounts almost exactly to the overall correction of HF applied for the PMOD composite during this period of time. The IRMB composite uses ERBE data for bridging the ACRIM gap, so a corresponding step is avoided, but the strong variation in the HF data during the gap support the fact that the correction of HF is needed. The most obvious difference for the IRMB composite is during the period of SOHO, which indicates a serious problem with the evaluation of the DIARAD data adopted by IRMB.

An estimate of the uncertainty of the long-term behaviour of the composite TSI can be deduced from the uncertainty of the slope relative to ERBE, which amounts for the PMOD composite to 1.1 ± 2.1 ppm/a. Although this uncertain-

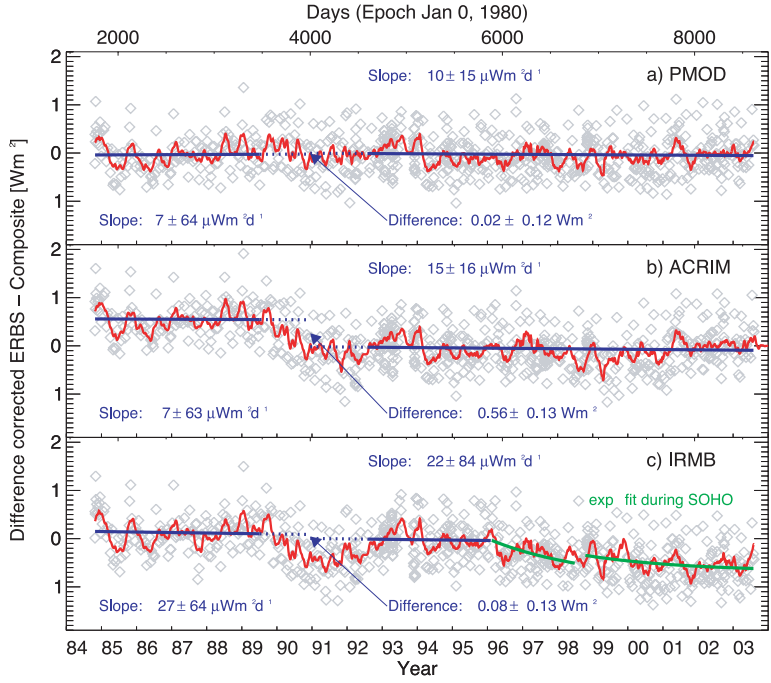


Figure 4. Shown here are the ratios of the three composite TSI records to the independent ERBS observations. The main difference is localized in the gap between the ACRIM I and ACRIM II missions. It is unlikely that ERBE instrumental effects can explain these differences as Willson & Mordvinov[31] suggest, since this would require an episodic sensitivity change confined only to this period. The IRMB composite uses ERBE data to bridge the ACRIM gap, which removes the step. The most obvious difference is during SOHO, which indicates a serious problem with the evaluation of the DIARAD data adopted by IRMB.

ty is partly determined by the sampling noise of ERBE, we can estimate the uncertainty of a possible trend to be <3 ppm/a. This implies a possible change of 50 – 80 ppm over the 25 years of the observations. If we add the uncertainties related to the tracing of ACRIM-II to I and of the HF correction (60 ppm), we get a total uncertainty of 92 ppm for the PMOD composite over the 25 years of observation. The difference between two successive minima, which could be interpreted as a secular change, amounts to -13 ppm, which is not significantly different from zero at the 6-σ level.

By just looking at the comparison of the three composites with ERBE in Figure 4, the PMOD composite is the most consistent. The Spearman's rank correlation coefficients of the three composites with ERBE of 0.751, 0.678 and 0.695, for PMOD, ACRIM and IRMB, respectively, supports this choice also from a statistical point of view. So, for climate and solar-physics studies, the use of the PMOD composite[36] is recommended (Fig. 5).

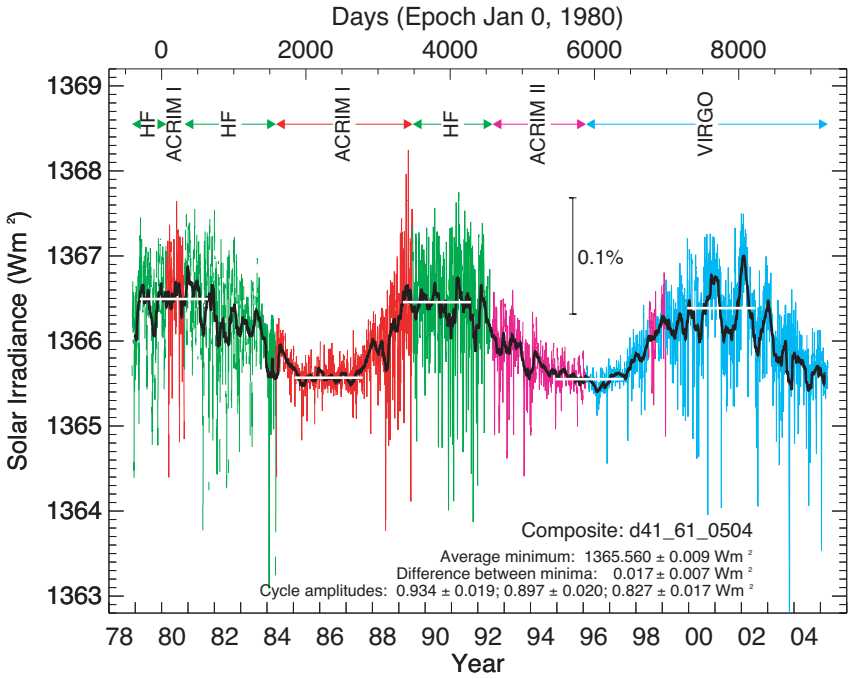


Figure 5. Shown is the PMOD composite[36] with the different colours indicating the origin of the time series used. Also indicated are the average values during the minima and maxima of the three solar cycles.

What can we learn from the TSI record of the last 25 years?

The zero difference between the last two solar minima indicates that there was no long-term change in the quiet Sun as defined by the irradiance at minimal activity. So, the mean value for the quiet Sun is 1365.56 Wm^{-2} with an uncertainty relative to SI of the order of 0.1 to 0.2%, the present estimate for room-temperature radiometry. With the low value of the Total Irradiance Monitor (TIM) on SORCE[37], however, this estimate may be too optimistic.

The relative uncertainty is much better with < 0.01% over the last 25 years and, for example, the differences in the amplitudes of the three cycles with 684.2 ± 13.6, 657.1 ± 15.1 and 605.9 ± 11.5 ppm are significant. The given uncertainties are 1-σ standard deviations of the average and are more formal, and do not represent any possible bias. Not only the amplitude change from cycle to cycle, but also the character of the activity is involved. The most recent cycle looks quite different mainly because there were many fewer sunspots associated with active regions than in the other two cycles, as shown by the large increases without many sunspots (Maunder-minimum type episodes). This points also to a problem

related to the reconstruction of TSI for the past, which uses the sunspot number for scaling the irradiance in the past[38-41]. The cycle amplitudes of the last three cycles of TSI and those of the sunspot group number[42] with 131.64 ± 0.24, 134.80 ± 0.25 and 95.81 ± 0.40 show little association. A more detailed analysis of the correlation between cycle-related long-term irradiance changes with the sunspot group number shows that it was quite reasonable during cycles 21 and 22, but definitively fails during cycle 23.

Synthesis and Conclusion

Particles, magnetic fields and radiation are varying with the 11-year solar activity cycle. Solar-wind composition varies by a factor of two between solar minimum and maximum, and by up to an order of magnitude in the short term. Total solar irradiance, on the other hand, varies by as little as 0.1% from minimum to maximum, and not at all over the past 25 years of direct observations, as manifested by no significant difference between the last two minima. On rotational time scales, changes of up to 0.5% can be observed if large sunspot groups happened to cross the visible disk, e.g. in October 2003.

Since the Sun's electromagnetic radiation is the dominant source of energy for the Earth, even small variations in irradiance have the potential to influence its climate and atmosphere, including the ozone layer[43,44]. Furthermore, the extinction of solar radiation by absorption and scattering in the Earth's atmosphere, and its reflection by land surfaces and oceans, are strongly wavelength-dependent, as are the processes through which climate responds to radiative input changes, involving atmospheric constituents such as water vapour and ozone, surface properties such as sea ice and snow cover, and most importantly clouds. The interest in solar-cycle variability is motivated by the fact that if a correlated variability is found in a terrestrial signal (such as a climate parameter), this may indicate a physical connection. Many such correlations have been claimed, but with a few exceptions[45,46] they have failed the "Geiss test", which states that a correlation holds only if it is improved with every solar cycle that is added to the database. Another correlation that has received considerable attention over the last decade is based on the correlation between cosmic-ray intensity and cloud cover[47], which, however, fails with the addition of more recent data.

The fact that TSI has not changed between the last two solar minima indicates that it is unlikely that the Sun was causing the observed global warming in recent decades. However, this must not be taken as a complete acquittal. Spectral solar irradiance, especially in the ultraviolet, is much more variable (e.g. a factor of approximately 2 for Ly-α at 121 nm), and through photochemical effects in the

middle atmosphere and nonlinear coupling, e.g. by gravity waves, to the troposphere a significant effect may result[46,48], which may be much larger than the simple changing of the incoming energy by a change in TSI.

Absorption and emission processes of gases in the Sun's atmosphere in various states of ionization produce spectral features with widths of typically a few tens of pm. Many spectral features are attributable to hydrogen, the most common component of the Sun's atmosphere, including prominent emission and absorption lines (e.g. Ly-α and H-α at 656.3 nm). Likewise, the second most common solar-atmosphere constituent, He, produces strong line emission (e.g. at 30.4 and 58.4 nm) and absorption (e.g. at 1083 nm). The variability of the solar ultraviolet irradiance, which is important for climate change, i.e. from the visible down to the wavelength of Ly-α, is mainly determined by the temperature distribution of the solar atmosphere and its composition. Knowledge of it allows quite accurate calculation of the solar spectrum[49,50], but it is based on observations and a self-consistent model for the temperature distribution from the photosphere to the lower corona is still missing. Moreover, the variabilities of the coronal and solar wind composition are related to the irradiance in EUV (at wavelengths below approximately 100 nm), which is mainly determined by the emission lines of highly ionized atoms originating in the chromosphere and corona. We have seen that only under the most favourable conditions measures of EUV line emission ratios can be taken as abundance ratios of the radiating elements. The lines must have been formed at the same location in the solar atmosphere, and the contribution function to the emission measure must have the same temperature dependence. A good model of the abundances and their variations in the solar atmosphere is therefore needed. Such a model must inherently be dynamic so as to include the FIP fractionation effect and the freezing-in of charge states. Discrepancies between EUV radiances and in-situ abundance observations (Fig. 2) indicate that our current understanding still needs improvement. The prospect of an improved understanding of solar variability and its influence on our climate should motivate us to work towards that.

References

1. C. Fröhlich, M.C.E. Huber, S. Solanki & R. von Steiger (Eds.), "Solar Composition and Its Evolution - From Core to Corona", Space Sciences Series of ISSI, Vol. 5, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **85**, Nos.1-2, 1998.
2. A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach & H. Kunow (Eds.), "Corotating Interaction Regions", Space Sciences Series of ISSI, Vol. 7, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **89**, Nos. 1-2, 1999.

3. E. Friis-Christensen, C. Fröhlich, J. Haigh, M. Schüssler & R. von Steiger (Eds.), "Solar Variability and Climate", Space Sciences Series of ISSI, Vol. 11, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **94**, Nos. 1-2, 2000.
4. H. Kunow, N. Crooker, J. Linker, R. Schwenn & R. von Steiger (Eds.), "Coronal Mass Ejections", Space Sciences Series of ISSI, Vol. 22, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, in preparation, 2005.
5. N. Grevesse & A.J. Sauval, in Ref. 1, p. 161.
6. K. Lodders, Solar system abundances and condensation temperatures of the elements, *Astrophys. J.*, **591**, 1220, 2003.
7. R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T. Owen & F. Robert (Eds.), "Solar System History from Isotopic Signatures of Volatile Elements", Space Sciences Series of ISSI, Vol. 16, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **106**, Nos.1-4, 2003.
8. P.R. Young & H.E. Mason, Atomic physics for composition measurements, in Ref. 1, p. 315.
9. K. Wilhelm, Spectroradiometry of spatially-resolved solar plasma structures, in "The Radiometric Calibration of SOHO", A. Pauluhn, M.C.E. Huber & R. von Steiger (Eds.), ISSI Scientific Report Series, Vol. 2, p. 37, ESA Publications Division, Noordwijk, 2002.
10. J. Christensen-Dalsgaard, in Ref. 1, p. 19.
11. S. Turck-Chièze, in Ref. 1, p. 125.
12. J. Geiss & G. Gloeckler, this volume, 2005.
13. J. Geiss *et al.*, *Science*, **268**, 1033,1995.
14. S.J. Bame, J.R. Asbridge, W.C. Feldman & J.T. Gosling, *J. Geophys. Res.*, **82**, 1487, 1977.
15. R. von Steiger *et al.*, *J. Geophys. Res.*, **105**, 27'217, 2000.
16. T.H. Zurbuchen, L.A. Fisk, G. Gloeckler & R. von Steiger, *Geophys. Res. Lett.*, **29**, doi:10.1029/2001GL013946, 2002.
17. R. von Steiger, in Ref. 1, p. 407.
18. R. von Steiger *et al.*, in "Solar and Galactic Composition", R.F. Wimmer-Schweingruber (Ed.), AIP Conference Proceedings, Vol. 598, p. 13, 2001.
19. D.V. Hoyt, H.L. Kyle, J.R. Hickey & R.H. Maschhoff, *J. Geophys. Res.*, **97**, 51, 1992.
20. R.C. Willson, *Space Sci. Rev.*, **38**, 203, 1984.
21. R.B. Lee III, B.R. Barkstrom & R.D. Cess, *Appl. Opt.,* **26**, 3090, 1987.
22. R.C. Willson, "The Sun as a Variable Star, Solar and Stellar Irradiance Variations", J. Pap, C. Fröhlich, H.S. Hudson & S. Solanki (Eds.), p. 54, Cambridge University Press, Cambridge UK, 1994.
23. D. Crommelynck *et al.*, *Metrologia*, **30**, 375, 1993.
24. C. Fröhlich, *Metrologia*, **40**, 60, 2003.
25. S. Dewitte, D. Crommelynck & A. Joukoff, *J. Geophys. Res.*, **109**, A02102, \doi {10.1029/2002JA009694}, 2004.
26. R.C. Willson, http://www.acrim.com/Data%20Products.htm, 2001.
27. G. Kopp *et al.*, in American Geophysical Union, Spring Meeting 2001, Abstract #SH52A-08, p. 52, 2001.
28. C. Fröhlich & J. Lean, *Geophys. Res. Lett.*, **25**, 4377, 1998.

29. C. Fröhlich, Observations of irradiance variations. in Ref. 3, p. 15.
30. R.C. Willson, *Science*, **277**, 1963, 1997.
31. R.C. Willson & A.V. Mordvinov, *Geophys. Res. Lett.*, 30, 1199, doi:10.1029/2002GL016038, 2003.
32. S. Dewitte, D. Crommelinck, S. Mekaoui & A. Joukoff, *Sol. Phys.*, in press, 2005.
33. C. Fröhlich & J. Lean, in IAU Symposium 185: "New Eyes to See Inside the Sun and Stars", F.L. Deubner, J. Christensen-Dalsgaard & D. Kurtz (Eds.), p. 89, Kluwer Academic Publ., Dordrecht, 1998.
34. C. Fröhlich, AGU Fall Meeting Abstracts, p. A301, 2004. The poster is available at ftp://ftp.pmodwrc.ch/pub/Claus/AGU Fall2004/AGU_poster_Fall2004.pdf.
35. D. Crommelynck, A. Fichot, R.B. Lee III & J. Romero, *Adv. Space Res.,* **16**, (8)17, 1995.
36. C. Fröhlich, Total solar irradiance: The PMOD-composite, 2005. A description of the construction and the newest version of the composite can be found at http://www.pmod-wrc.ch/pmod.php?topic=tsi/composite/SolarConstant.
37. G.A. Kopp, G. Lawrence & G. Rottman, AGU Fall Meeting Abstracts, #SH31C-C7, 2003.
38. J. Lean, in Ref. 3, p. 39.
39. J. Lean, *Geophys. Res. Lett.*, **28**, 4119, doi:10.1029/2001GL013969, 2001.
40. S.S. Foster, PhD Thesis, University of Southampton, 2004.
41. M. Lockwood, in "Saas-Fee Advanced Courses, Number 34: The Sun, Solar Analogs and the Climate", I. Rüedi, M. Güdel & W. Schmutz (Eds.), p. 109, Springer-Verlag, Heidelberg, 2004.
42. D.V. Hoyt & K.H. Schatten, *Sol. Phys.*, **181**, 491, 1998.
43. U. Cubasch & R. Voss, in Ref. 3, p. 185.
44. D. Rind, *Science,* **296**, 673, 2002.
45. H. van Loon & K.Labitzke, in Ref. 3, p. 259.
46. J.D. Haigh, *Phil. Trans. Roy. Soc. A*, **361**, 95, 2003.
47. N. Marsh & H. Svensmark, in Ref. 3, p. 215.
48. D. Rind *et al.*, *Journal of Climate*, **17**, 906, 2004.
49. J.M. Fontenla, O.R. White, P.A. Fox, E.H. Avrett & R.L. Kurucz, *Ap. J.*, **518**, 480, 1999.
50. J.M. Fontenla *et al.*, *App. J.*, **605**, L85, 2004.

Space Plasma Physics

B. Hultqvist[a], G. Paschmann[b,c], D. Sibeck[d], T. Terasawa[e], R.A. Treumann[b] and L. Zelenyi[f]

[a]*Swedish Institute of Space Physics, Kiruna, Sweden*
[b]*Max-Planck-Institut für extraterrestrische Physik, Garching, Germany*
[c]*International Space Science Institute, Bern, Switzerland*
[d]*NASA/Goddard Space Flight Center, Greenbelt, USA*
[e]*Dept. Earth Planet. Science, Unversity of Tokyo, Tokyo, Japan*
[f]*Space Research Institute (IKI), Russian Academy of Sciences, Moscow*

Introduction

Almost all the matter in near-Earth space is highly if not fully ionized and thus dominated by electromagnetic forces. Such matter is in the fourth state of matter, the plasma state, which on Earth is rarely accessible in comparable purity. Its investigation requires the use of either rockets or spacecraft.

Space plasma physics dominated research in the space sciences during the first two decades after Sputnik and still remains one of the largest research fields in terms of the number of scientists involved. As such, it has played a major role in the programme at ISSI during its first ten years. Today near-Earth space has become a "laboratory" in which to study plasma physics.

The space-plasma-physics field has reached a certain degree of maturity during the space era, but is still a young research field in the sense that unexpected observations from all new space missions continue to surprise. The physics of space plasma has been found to be complex. It is impossible to derive more than very limited conclusions from basic principles alone. Theory needs strong guidance from experiments in order to stay on track. On the other hand, the limitations of the observational possibilities in space, with its enormous spatial dimensions and temporal variations, make theoretical and numerical models essential for the interpretation of the observations. The fact that important new results come as surprises means that many models are still in an early stage.

Among the space sciences, space plasma physics is nevertheless at an advanced stage. Detailed experiments aimed at clarifying specific physical problems are

needed and are possible to perform in the "laboratory" of near space by launching small, specialised satellites, with high-resolution instruments onboard, into the right orbits. The high temporal and spatial resolutions of some recent plasma instruments have led to the opening of new parameter spaces for measurements. During the first ten years of ISSI, for instance, the small satellites Freja (Sweden/Germany) and FAST (USA) are examples of projects that have provided experimental data with much more temporal and spatial structures than have been available before. New dimensions have been introduced by ESA's four Cluster satellites, which for the first time allow for the direct determination of vector quantities in space.

ISSI's role in developing space plasma physics has been to bring together groups of scientists to discuss and summarise the current understanding of specific physical problems or sub-fields and, most importantly, to integrate, generally for the first time, major bodies of experimental and theoretical results, involving the World's leading scientists who have worked in the field. Some important new results have come out of these integrating efforts.

At ISSI, teams, working groups and workshop projects have covered a wide spectrum of subjects in space plasma physics, ranging from the very specialised to broad research areas. It is, of course, not possible to report all of the results here, so we have selected below a number of topics to which major efforts have been devoted within the ISSI programme.

Source and Loss Processes of Magnetospheric Plasma

Before the first ion-composition measurements of magnetospheric plasma were made, there was a general belief among space physicists that the ionosphere is negligible as a plasma source for the magnetosphere and that all plasma in the magnetosphere is of solar-wind origin. That view was based on the fact that most of the processes for transporting ionospheric plasma into the magnetosphere were unknown at the time, and have only been discovered later. When the Lockheed group launched its first ion mass-spectrometer on a small US military low-orbiting satellite in the late 1960s, they found that the keV ions, which precipitated into the atmosphere from the magnetosphere, contained an appreciable fraction of O^+ ions, which can only originate in the ionosphere. These results were so surprising that they did not dare to publish them until they had launched another ion mass-spectrometer on another similar satellite and found the same results[1]. Another small military research satellite, S3-3, was launched in 1976 into an elliptic polar orbit with an 8000 km apogee at high latitudes, and it produced surprising results showing field-aligned ion beams, with a large fraction

of O^+ ions, coming out of the ionosphere[2]. When the first ion mass-spectrometer was sent into the central magnetosphere on ESA's GEOS-1 spacecraft by the Bern group in 1977, they could conclude that the ionosphere is a source of similar importance to the solar wind in the region of the magnetosphere reached by GEOS-1, i.e. within about 8 R_E geocentric distance[3]. On the basis of further satellite measurements it was even argued by some scientists[4] that the ionospheric source could provide all the plasma seen in the magnetosphere up until then (not including the tail). This was thus a 180° change of view compared with that generally held before the launch of the first ion mass-spectrometer. Such was the situation in 1996 when ISSI initiated a study project on "Source and Loss Processes of Magnetospheric Plasma" [5,6].

Plasma sources

By then, many years of direct measurements at the magnetosphere's inner boundary of the ionospheric ion outflow into the magnetosphere had shown a total ion outflow amounting to 2×10^{25} s^{-1} to 10^{26} s^{-1} for H^+ and to 0.5×10^{25} s^{-1} to 3×10^{26} s^{-1} for O^+ at low and high levels of solar and magnetic activity, respectively[6,7].

It is known from ion-composition measurements in the magnetosphere that the solar wind is an important source of magnetospheric plasma. For instance, fairly large amounts of He^{2+} ions, which cannot come from the ionosphere, are always present. This demonstrates definitively that the outer boundary of the magnetosphere, the magnetopause, is not an impenetrable boundary for solar-wind plasma. Direct measurements of solar-wind plasma transport across the magnetopause are, however, exceedingly difficult to make because the plasma flow and the magnetic field are essentially directed tangential to the magnetopause. Any transport of plasma across the magnetopause is only a small perturbation on the tangential transport. This is contrary to the ionospheric-outflow case, where the outflow is aligned with the more or less undisturbed magnetic field and is usually the dominant component. Any measurements apply only locally and it is clear that the direct determination of the total transport of plasma into the magnetosphere, in a way analogous to what has been done for the ionosphere, i.e. by direct *in-situ* measurements distributed over the entire magnetopause, is not possible. Instead indirect methods have to be used.

Of the physical processes that may contribute to the transport of plasma across the magnetopause, magnetic reconnection is the one most studied and best supported by experimental results. It also provides the largest number of testable predictions among all the mechanisms that have been discussed. In the magnetic reconnection model, the "frozen-in" condition of plasma and magnetic field, in which low-energy plasma (but not energetic particles) and magnetic field lines

can be considered as moving together (the field lines "stick" to the plasma elements), is violated only in a localised region, called the "diffusion region", where fields may diffuse and become interconnected. Outside of the diffusion region the frozen-in condition is a good approximation everywhere in the magnetosphere and the solar wind, except in regions with a magnetic-field-aligned electric field, such as other diffusion regions in the magnetotail and in auroral acceleration regions near Earth. Reconnected field lines stay connected when they are pulled along the magnetopause into the tail by the solar wind. As it is much easier for charged particles to move along magnetic field lines than perpendicular to them, particles can fairly easily move from the solar-wind side to the magnetosphere side of the magnetopause.

Estimates of the total inflow rate across the dayside magnetopause give the order of magnitude 10^{26} s^{-1}, which corresponds to a mass influx of approximately 1 kg s^{-1}. This inflow rate has the same order of magnitude as the total outflow from the ionosphere. For the tail magnetopause no direct observational results have been reported, so in order to arrive at an estimate of the total solar-wind plasma entrance rate through the magnetopause of the magnetotail we have to use other observational results from the tail.

Only data from the passages of ISEE-3, Pioneer-7, and Geotail through the distant tail have been reported in the literature. Although the number of spacecraft passes through the distant tail is limited, the published data provide a consistent picture of the ion flow through the tail. From the ISEE-3 observations, the following fluxes of anti-sunward-moving ions for different down-tail distance ranges have been derived: 2×10^{26} s^{-1} in the distance range 0-60 R_E, 7×10^{26} s^{-1} at 60-120 R_E from Earth, 2×10^{27} s^{-1} for 120-180 R_E and somewhere between 3×10^{27} s^{-1} and 3×10^{28} s^{-1} beyond 180 R_E. Recently reported Geotail observations have given results entirely consistent with these numbers. About 1000 R_E down-tail, magnetotail densities and velocities, observed during a single passage of Pioneer-7, correspond to an anti-sunward flux of about 6×10^{28} s^{-1}.

From the above data, one clear conclusion can be drawn: namely that the supply rate of plasma through the dayside magnetopause and from the ionosphere (both of order 10^{26} s^{-1}) are orders of magnitude too small to account for the anti-sunward plasma flow through the deep tail. Therefore, plasma has to enter the magnetosphere elsewhere at high rates; along the magnetotail is the only remaining alternative. The reconnection model provides the required free access of plasma along a part of the tail magnetopause. Other mechanisms for transferring plasma across the magnetopause, which may contribute to the transport, are not likely to provide the required amounts of plasma, simply because they only work locally and the known ones do not provide the kind of free access that the reconnection does.

A consequence of the situation sketched above is that most of the solar-wind plasma that enters the magnetosphere streams along the tail and out through the distant throat back into the interplanetary space, without affecting the closed-field-line region around Earth (see Fig. 1).

Plasma sinks

The high-latitude ionosphere is not only a source of magnetospheric plasma, but also a sink. Magnetospheric ions and electrons precipitate into the upper atmosphere primarily in the auroral regions, but also in the polar caps. In order to reach the atmosphere, ions and electrons must have velocities nearly parallel to the geomagnetic field, since they will otherwise be reflected above the atmosphere by the magnetic mirror force. The directions relative to the magnetic field for which ions can precipitate into the atmosphere (in practice, they reach about 100 km altitude) define the loss cone. At auroral latitudes, the loss cone as seen near the equatorial plane is generally quite small (cone half angle 1-4°).

Statistical investigations have demonstrated that the total average flux into both hemispheres is 2×10^{24} to 1×10^{25} s^{-1} for ions (with energies between 30 eV and 30 keV) and 1×10^{26} s^{-1} to 6×10^{26} s^{-1} (for 50 eV to 20 keV energies) for electrons, with the higher numbers corresponding to higher magnetospheric disturbance levels[8]. Comparison of these numbers with the total outflow figures given in the previous section shows that the total ion outflow rate is an order of magnitude larger than the total ion precipitation rate. The magnetosphere thus returns to the upper atmosphere much less plasma than it extracts from it. This is not surprising considering that generally all ionospheric ions that achieve sufficient speed to escape from the gravitational field have free access to the magnetosphere, whereas ions must have a velocity direction within the small loss cone in order to precipitate.

As for the transport of plasma into the magnetosphere across the magnetopause, it is not possible to evaluate the total loss rate of plasma from the magnetosphere through the magnetopause from direct *in-situ* measurements. Observations of both energetic ions and electrons outside the magnetopause have been reported and it seems likely that they come from the magnetosphere, but no quantitative flow numbers are available. Assuming that the reconnection model discussed earlier is applicable, a simple estimate can be made. For a southward-directed interplanetary magnetic field, the uniform electric field on the two sides of the magnetopause that is implied by the reconnection model pushes the plasma on both sides towards the magnetopause at equal speeds if the magnetic field strengths on the two sides are equal. In this case, the ratio of inflow and outflow scales as the ratio of the number densities on the two sides. As the number density of the magnetospheric plasma is typically one order of magnitude less than

that of the solar-wind plasma outside the magnetopause, the outflow of magnetospheric plasma due to reconnection is expected to be an order of magnitude smaller than the inflow. For an inflow through the dayside magnetopause on order 10^{26} s^{-1} (see previous section), the total ion outflow through the dayside magnetopause will be of the order of 10^{25} s^{-1}. The dayside magnetic-field intensity on the inner side of the magnetopause is generally stronger than that on the outside, and the outflow rate is correspondingly reduced according to this rough way of estimating. We thus see that, as in the case of the ionosphere, more plasma is provided to the magnetosphere across the magnetopause than is lost through it. This imbalance is removed by the high rate of plasma loss through the deep tail exhaust already described in a previous section.

Conclusions from the balance of sources and losses

As mentioned earlier, the anti-sunward ion flow in the magnetotail has been found to increase with distance from Earth, from 2×10^{26} s^{-1} in the near-Earth tail to $(3 - 30) \times 10^{27}$ s^{-1} beyond 180 R_E and even 6×10^{28} s^{-1} at 1000 R_E. Contrary to what was argued by some research groups in the 1980s (see above), none of the measured or quantitatively estimated sources near Earth can provide such large ion fluxes, so we concluded that solar-wind plasma must enter all along the tail magnetopause. We thus have a situation where the strength of a plasma source that has not been directly observed is defined by the requirements of an observed plasma sink in the deep magnetotail. Regions with reconnected field lines along the entire tail on both sides of it may match the down-tail loss. Using large-scale simulations, the total inflow of solar-wind plasma along reconnected field lines has recently been calculated and the result is consistent with the measured outflow along the tail[9].

A summary of source and loss rates for different parts of the inner and outer boundaries of the magnetosphere is given in Table 1[6]. The underlined numbers are based on direct measurements. The others are estimated as indicated earlier. Figure 1 contains a schematic summary of some of the information presented above. The balance between source and loss processes is a dynamic one, with significant variations in the efficiency of the various processes at the various boundaries, depending on the flow speed and density of the solar wind, the direction of the interplanetary magnetic field, the height distributions of plasma density and temperature in the upper ionosphere, etc. Table 1 and Figure 1 show average numbers.

The observed total flow of solar-wind ions through the magnetosphere, amounting to the order of 10^{28} s^{-1}, is not more than an order of magnitude less than the undisturbed solar-wind flow through an area the size of the magnetospheric cross-section. Only a small fraction of the solar-wind ions affect the closed-

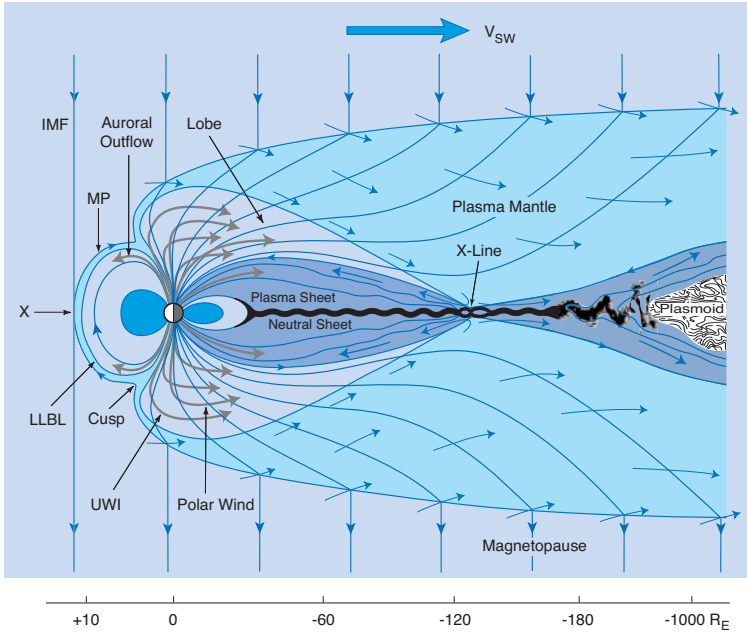


Figure 1. Two-dimensional view (not to scale) of the magnetosphere, showing the magnetic field and plasma configurations and the three main plasma sources/losses. These processes are magnetopause and tail reconnection and plasma flow from and to the ionosphere. The plasma flow is indicated by arrows, blue for solar-wind plasma and violet for ionospheric plasma. The blue lines with arrows show the magnetic field lines (after Figure 7.1 in Ref. 6).

field-line region in the magnetosphere, i.e. the plasma sheet earthward of the reconnection region in the tail, and the dayside magnetosphere. The source strength for that region is two orders of magnitude less, i.e. 10^{26} s^{-1}. The ionosphere and the solar wind both contribute significantly to the plasma content there, but the proportions vary with magnetospheric disturbance level and solar-cycle phase. The dominant sink for this region is the same as that for the solar-wind-dominated plasma outside the closed-field-line region, namely the downtail flow out of the magnetosphere. The ionosphere is the dominating plasma source in a region of the magnetosphere near Earth.

	Source rates	Loss rates
High-latitude ionosphere	$\underline{10^{26}}$	$\underline{10^{25}}$
Plasmasphere	$<< 10^{26}$	$< 10^{25}$
Magnetopause, dayside	10^{26}	10^{25}
Magnetopause, along magnetotail	$10^{28} - 10^{29}$	?
Deep tail "exhaust pipe"	?	$\underline{10^{28}} - \underline{10^{29}}$

Table 1. Summary of Source and Loss Rates (per second, orders of magnitude). Underlined numbers are based on direct measurements.

The results reported here came out of the ISSI study project[6], which for the first time brought experts on and data from the various boundary regions together and produced an integrated view of the problem based on direct observations. The conclusion – that the main source region for the plasma in the magnetosphere as a whole is the tail magnetopause – is consistent with the reconnection model. It is an example of ISSI, in its integrating role, being able to conclude a long-lasting discussion.

The Aurora

The aurora is one of the most spectacular visible natural phenomena, and at the same time one of the richest and most fascinating subjects in all of space plasma physics. While the aurora is one of the directly observable manifestations of the Sun-Earth connection chain, the underlying plasma processes are expected to be ubiquitous in the plasma universe.

Auroral light is produced when electrons and protons impinging on Earth's upper atmosphere excite the atmospheric atoms and molecules. In the night-side auroral oval, this usually occurs in the form of auroral arcs, narrow structures oriented tangential to magnetic latitude circles, a few hundred metres to several tens of kilometres wide in latitude, while extending several thousand kilometres in longitude. The arc often displays a relatively sharp lower boundary at around 100 km altitude.

Key questions regarding the origin of the aurora concern the relationship between auroral displays and the large-scale magnetospheric current systems and boundaries, and, most importantly, the nature of the process that accelerates the primary auroral particles. Recent high-resolution *in-situ* observations from satellites (notably Viking, Freja, and FAST) and sounding rockets, together with remote-sensing of the optical auroral displays, have led to large steps forward in answering these questions. The time therefore seemed ripe to gather a team of experts at ISSI and ask them to write a comprehensive and well-integrated book on the subject. The work was organized through three workshops, starting in March 1999, and the book was published in 2002 as Volume 15 (Auroral Plasma Physics) in the Space Science Series of ISSI[10]. The textbook nature of the volume is reflected in the fact that the book is not a collection of articles by different authors, but a monograph with 32 authors.

Because it is created by charged particles hitting the upper atmosphere, the aurora represents a major loss mechanism for the plasma in the magnetosphere. At the same time, particles are also injected back into the magnetosphere through

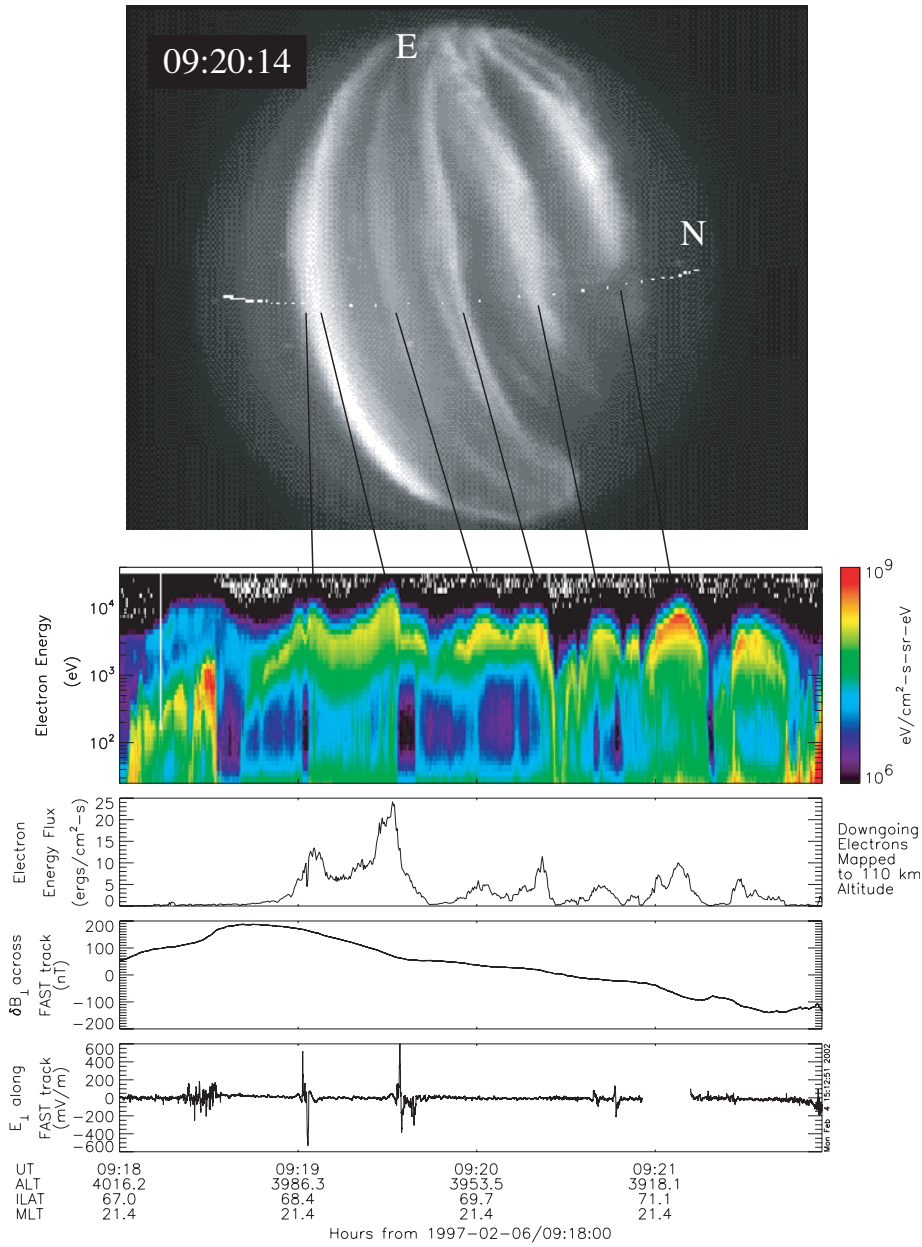


Figure 2. Relation between the optical aurora and particle measurements. The *top panel* shows multiple auroral arcs seen in the all-sky image, taken by a low-light level TV-camera from an aircraft, and the FAST satellite orbit mapped to 110 km altitude. The *second panel* shows the electron energy-time spectrogram measured by the FAST satellite, where particle intensity is colour-coded as shown by the bar on the right. The *third panel* is the precipitated electron energy flux. In the *fourth panel*, the perpendicular magnetic-field perturbation is shown. In the *fifth panel*, the electric field along the FAST track is displayed (from Ref. 12).

the effects of reflection, electric fields, and heating. The aurora is therefore also a source of plasma. For this reason, the aurora already played a major role in an earlier workshop[6] that dealt with magnetospheric plasma sources and losses (see previous section).

Alfvén waves play an important role in the coupling between the auroral acceleration region and the magnetosphere and in the acceleration of auroral electrons. One of the earlier international teams at ISSI focussed on exactly this topic and produced a comprehensive review[11]. In the following sections, we describe some of the key results that have been presented in Reference 10.

Remote-sensing and in-situ observations

The complementary role of *in-situ* and remote-sensing observations in the study of the aurora is illustrated in Figure 2, which shows an example of a satellite crossing above multiple parallel arcs and simultaneous optical coverage of these arcs from an aircraft. While optical observations reveal the overall morphology of the auroral display, spacecraft provide detailed particles and field observations, albeit only along the spacecraft trajectory. The arcs can be identified by peaks in the energy spectra and flux of the downward-going electron component (called "inverted-V events" because of their resemblance to an inverted-V in the spectrograms).

Relationships with global current systems

As a synthesis of the recent observations, notably those from the Freja and FAST missions, Figure 3 outlines the electric-current structures that delineate the auroral zones, i.e. upward-current regions, downward-current regions, and time-varying-current regions, and summarizes the associated observational characteristics. This figure served as the guide to the presentations in Reference 10.

Figure 4 provides an example of *in-situ* satellite data illustrating the regions and processes listed in Figure 3: upward-current regions designated with blue markers in the top panel; downward-current regions in green, and Alfvénic regions in red. The polar cap is to the right.

In the upward-current region, precipitating electrons are narrow in energy and broad in pitch angle; in the downward-current region, the (up-going) electrons are broad in energy and narrow in pitch angle. In the Alfvénic region, the electrons are variable and counter-streaming. Each region has characteristic ion outflow, with differing energy and pitch-angle structure and outflow magnitude. The bottom two panels illustrate the great variety of wave activity associated with the aurora.

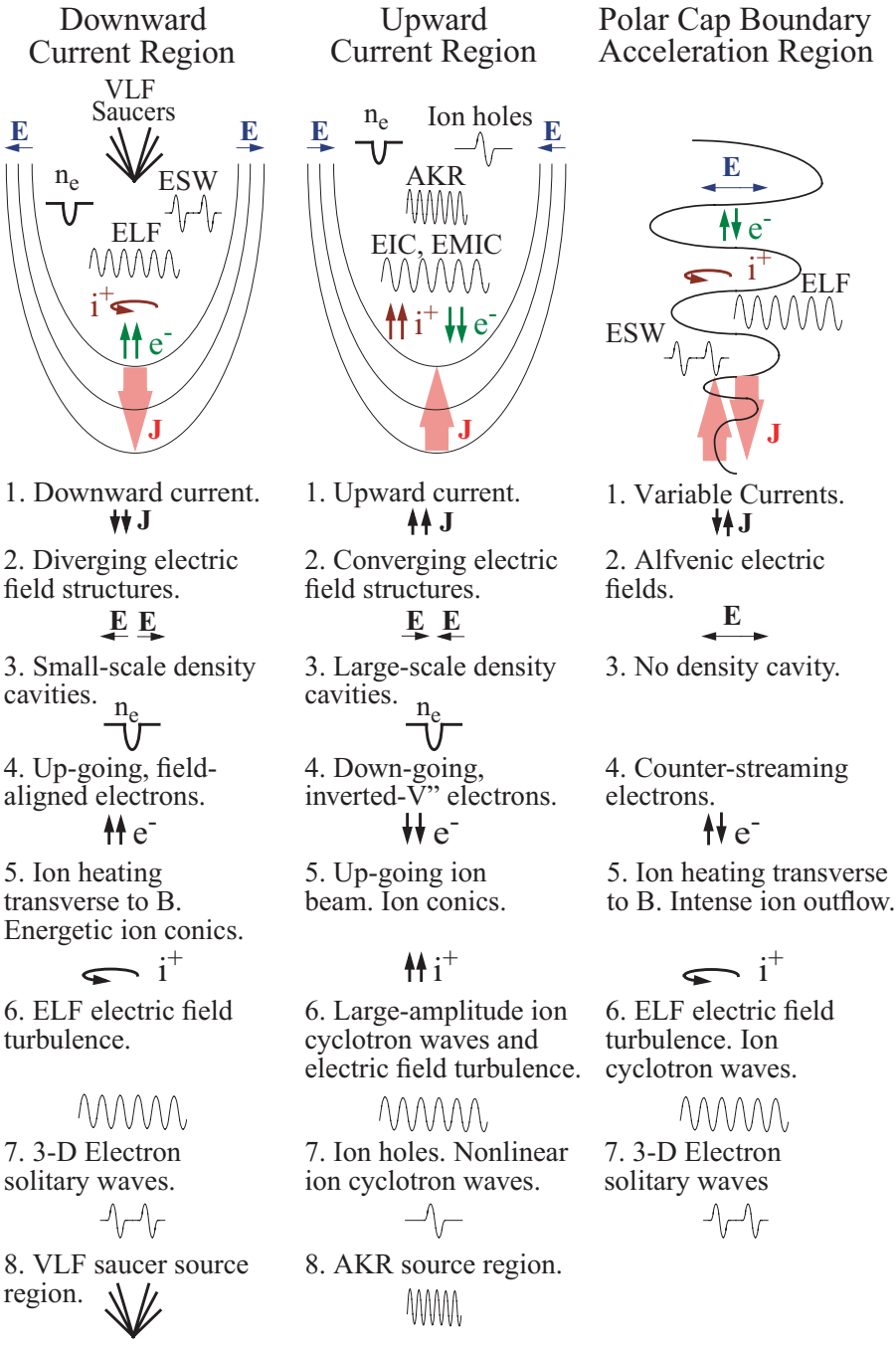


Figure 3. An illustration of the field-aligned current and electric potential systems in the ionosphere and magnetosphere. The numbered items refer to key characteristics of these regions (from Ref. 13).

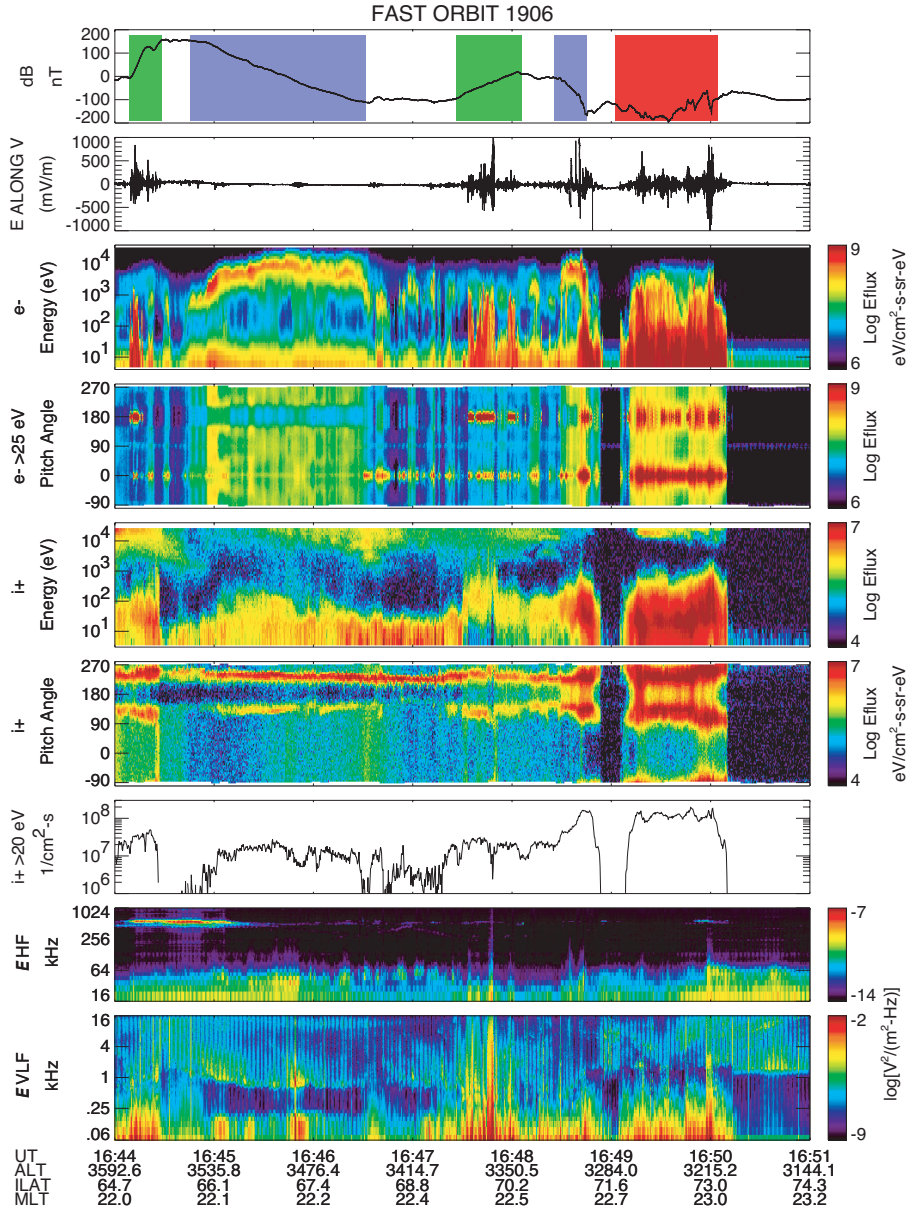


Figure 4. An auroral pass as seen by the FAST spacecraft. The *top panel* shows the magnetic-field perturbation, with the inferred field-aligned currents indicated in green (downward) and blue (upward), and the Alfvénic currents in red. The DC electric-field fluctuations in the *second panel* show the electrostatic shock structures associated with the auroral acceleration region. The *next four panels* show ion and electron spectrograms, i.e. intensities versus energy and pitch-angle, colour-coded, using the colour bars on the right. The *third panel from bottom* shows the integrated ion outflow. The *bottom two panels* show wave activity from near-DC to MHz frequencies (from Ref. 13).

Parallel potential drops

One of the major breakthroughs in our understanding of the aurora has come from the observational proof that it is electric fields directed along the geomagnetic field lines that are responsible for the acceleration of the auroral particles (as originally proposed by Alfvén).

While direct measurement of these parallel electric fields is difficult, much information about the parallel potential profile – and therefore, the parallel field – can be derived from particle data. The precipitating-electron data show the magnitude of the potential drop above the observation point. The ion-beam data and the electron-loss-cone data show its magnitude below the observation point.

According to the concept illustrated in Figure 3, one can compare the perpendicular potential drop that the spacecraft sees as it moves across the top of the potential structure with the ion-beam and electron-loss-cone measurements of the parallel potential drop below the spacecraft. Observationally, there is indeed good agreement between the ion-beam energy and the potential along the spacecraft trajectory as measured by the integrated electric field component perpendicular to the magnetic field lines. This agreement provides clear evidence for the potential-drop model of auroral acceleration.

As auroral arcs are produced by thin sheets of precipitating electrons, they also carry a considerable upward-directed field-aligned current. In simple terms, the relation between auroral particle fluxes and the field-aligned currents stems from the need to maintain current continuity (and charge neutrality) in the presence of a low number density of current carrying electrons and increasing current concentration just above the topside ionosphere resulting from the convergence of the magnetic field lines. The current continuity is achieved by downward acceleration of the electrons in an upward electric potential drop. Figure 5 is a sketch that illustrates the relationship between the auroral arc, the upward current and the electric field.

As illustrated in the sketch in Figure 3, the return (downward-) current region has a divergent shock structure which mirrors the convergent structures of the inverted-V in the upward-current region. Observations of isolated divergent electric-field structures by Freja and FAST have shown excellent agreement between observed suprathermal electron fluxes and the total current determined from magnetic-field measurements. Good agreement has also been found between the characteristic electron energy and the measured electric potential drop. All this supports the conclusion that the upward electron beams are accelerated by quasi-static potential structures.

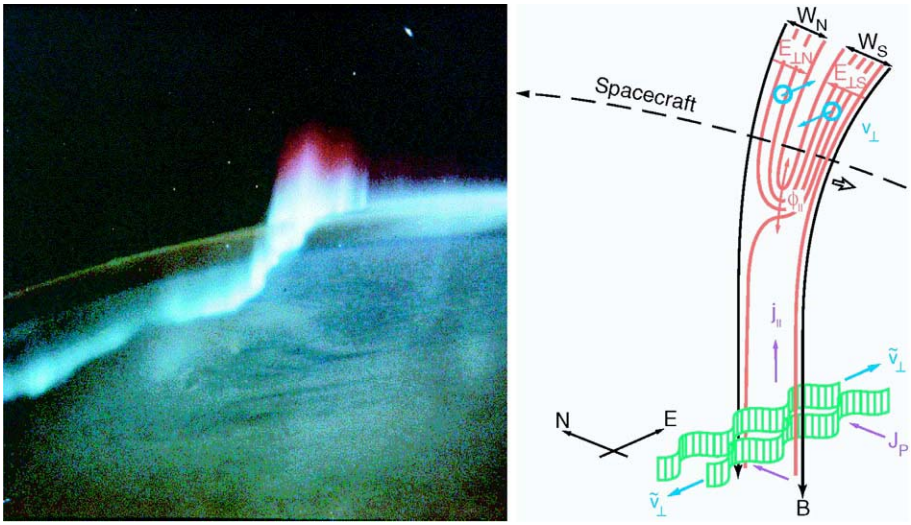


Figure 5. A combination of an optical auroral arc system on Earth (*left*) as seen from the Space Station (Spacelab) and the corresponding schematic model (*right*) showing a view of parallel currents, $j_{||}$, plasma motions, $v_{\perp}$, electric fields, $E_{\perp}$, and potential contours (red lines) above the auroral arc, which is sketched in green (from Ref. 12).

It is important to note that the formation of the electric-potential patterns and associated currents implies the existence of a generator region somewhere above the auroral acceleration region. The nature of this generator is still far from clear. Also not understood at present among the main characteristics of the aurora is the cause for the often very narrow width of auroral arcs.

Plasma Boundaries

Thin boundaries are ubiquitous in space plasmas. Examples are current sheets that separate plasmas with different magnetisation (such as the magnetopauses or tail current sheets in planetary magnetospheres, the heliospheric current sheet), shocks standing in a plasma ahead of obstacles or propagating through a plasma (such as planetary bow shocks, interplanetary shocks, or the heliospheric termination shock). *In-situ* observations from spacecraft have provided a wealth of information on the formation and characteristics of such boundaries, notably those occurring when the solar wind impinges on Earth's magnetosphere.

The terrestrial bow shock is formed when the supersonic plasma emitted from the Sun (the solar wind) encounters Earth's magnetic field. The dipole magnetic field of Earth acts as an (almost) impenetrable obstacle to the solar wind, which

therefore has to slow down and flow around the obstacle. In this process, the magnetopause is formed, separating the magnetic field inside from the solar wind that flows around it. Ahead of the magnetopause, the bow shock forms a surface across which the solar-wind plasma is heated and slowed down from supersonic to subsonic speeds.

Earth's bow shock is the best-known example of a collisionless plasma shock and has been the subject of extensive observational and theoretical investigations since the start of the space age. Collisionless shocks abound throughout the astrophysical world and are believed to play critical roles in flow dynamics and heating, as well as being a prime acceleration environment for charged particles with energies up to those of cosmic rays.

The magnetopause is a thin sheet of electric current that, as mentioned above, to first order separates the solar wind from Earth's magnetosphere. Across the magnetopause, the plasma environment changes from the dense, cold, weakly magnetised solar-wind plasma to the dilute, hot, and strongly magnetised plasma inside the magnetosphere. On closer examination, the magnetopause is not impermeable, but allows a fraction of the solar wind's mass, momentum and energy to be transferred to the magnetosphere, primarily through a process called "magnetic reconnection". Reconnection requires some process to unfreeze electrons and ions from the magnetic field, but this process needs only to be present in the immediate vicinity of the reconnection site. Elsewhere on the magnetopause, it is the normal magnetic-field component accompanying reconnection that creates a direct magnetic coupling across the magnetopause and allows solar-wind plasma to flow into the magnetosphere.

Plasma boundaries have received much attention in ISSI's programme. The first workshop series in the field of solar-terrestrial physics that started in 1995 was dedicated to the identification of the source and loss processes of magnetospheric plasma. As any mass transfer across the magnetopause constitutes, by definition, a source or loss of magnetospheric plasma, the magnetopause played an important role in that series of workshops, which resulted in Volumes 2 and 6 of the SSSI series[5,6].

Recognizing the need to develop and test methods and tools for the analysis of ESA's forthcoming Cluster mission, ISSI hosted an international team of experts to write a book on multi-spacecraft analysis methods. The result was published as the first volume in the ISSI Scientific Report series[14], which has become the reference for much of the Cluster-related data analysis, and from which much of the text of this section has been taken.

Plasma boundaries are constantly moving, evolving and changing their orientation. They are thus a prime target for the study by the closely spaced fleet of Cluster spacecraft. It was natural then that the bow shock and magnetopause should become the focus of an ISSI workshop series (entitled Dayside Solar Wind-Magnetosphere Interaction: Cluster Results) that began in 2003 and has led to the publication of another ISSI Volume[15] in 2005.

Plasma boundaries have also been studied by international teams that have met at ISSI. One dealt with numerical simulations of selected bow-shock problems, and led to a comprehensive review[16]. Another team looked at plasma measurements from the closely spaced AMPTE-IRM and-UKS spacecraft, in preparation for the Cluster multi-spacecraft mission. Several publications resulted from this effort. In the following, we will present some of the highlights from the work at ISSI on plasma boundaries.

The bow shock

It has been shown by an ISSI team[16] that "quasi-perpendicular" shocks should become non-stationary when the fraction of reflected ions exceeds some critical value. Non-stationarity of the shock largely determines the electron reflection process. Full particle simulations of high-Mach-number shocks were used in order to study electron behaviour during shock crossings in detail. Instabilities between the incoming electrons and the reflected ions lead to fluctuations on electron scales, which trap and accelerate the electrons.

Energetic particles are known to exist upstream of "quasi-parallel" shocks. The investigation of cross-field diffusion processes of the energetic particles upstream and downstream of the shock involves three-dimensional simulations. A still-outstanding question concerns the so-called injection process, i.e. how a certain part of the thermal upstream ions are injected at the shock into a diffusive shock-acceleration mechanism. Hybrid simulations show that pickup ions, i.e. neutral interstellar particles, which are ionized in the inner heliosphere and then picked up by the solar wind, are easily reflected and injected into an acceleration mechanism at quasi-parallel shocks.

The Cluster mission has made important contributions to the understanding of the physics of the bow shock. Firstly, by making the first detailed, three-dimensional studies of individual shock crossings, the phenomenology and physical processes within and in the vicinity of the bow shock, under specific conditions, could be clarified. Secondly, through the ability to make unambiguous determinations of the vector quantities associated with the shock it has been possible to underpin and re-examine the statistical studies of shock motion, and local and overall shock orientation. Cluster has explored spatial scales from 100 km to

5000 km and this range will be extended to 10 000 km and beyond before the end of the mission. Key findings include[17]:

1. The definitive determination of absolute shock scales.
2. The temporal/spatial variability: motion and internal dynamics.
3. The proof that ion beams manage to emerge from particles reflected at quasi-perpendicular shocks.
4. The surprisingly small-scale sub-structure of large structures at quasi-parallel shocks.

Shock scales. Cluster studies have measured the width of the ramp at the quasi-perpendicular bow shock over a range of upstream parameters (Mach number, etc.). The width is a critical indicator of the internal shock processes, which in turn govern the partition of energy among the incident particle populations[18].

Figure 6 shows the thickness of the shock ramp for almost 100 quasi-perpendicular shock crossings as a function of Mach number. Statistically, the measured ramp scale size was proportional to the gyro-radius of trapped ions, over a large range of Mach numbers.

Shock variability. Cluster determination of the speed of the bow shock has shown that variations in the upstream parameters have an immediate and direct impact on the location and gross motion of the shock. However, Cluster electric- and magnetic-field observations have also highlighted considerable variability in the shock structure and profile, even over relatively small scales.

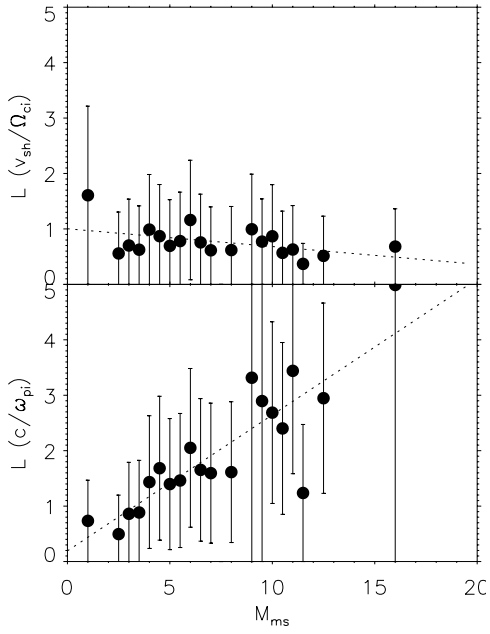


Figure 6. Relationship between the thickness of the (quasi-perpendicular) bow shock and the upstream magnetosonic Mach number. In the *top panel*, the thickness has been scaled to the downstream ion gyro radius, while in the *lower panel* it has been scaled to the ion inertial length. The figure shows that the gyro-radius scaling renders the thickness approximately constant over a large range of Mach numbers, while the ion inertial scaling increases with Mach number (from Ref. 18).

The near-simultaneous measurement of the shock profile by four spacecraft allows the study of spatial and temporal variability in ways that have not previously been possible. By considering the magnetic-field profile through a nearly perpendicular supercritical shock, it was possible to identify structures that had a stable phase with respect to the main shock ramp (and those that did not). Magnetic-field magnitude structures were found not to vary significantly over the spacecraft separation (around 600 km) or the time differences between the shock passages of the different spacecraft (up to 30 s), which is an interesting new observational result.

By contrast, the downstream large-amplitude waves (with a polarisation and frequency consistent with ion-cyclotron waves generated by non-gyrotropic ion distributions) varied significantly between spacecraft, confirming that these waves were not stationary with respect to the shock. In addition, it was not possible to identify the same waves at different spacecraft, implying that the scale sizes of these waves along the shock front were no larger than the spacecraft separations of around 600 km – their wavelength along the shock front was around 100 km.

Beam origin. Simultaneous Cluster ion observations at several locations have provided unambiguous evidence that field-aligned beams found upstream of the quasi-perpendicular bow shock emerge out of the reflected and partially scattered population at the shock itself rather than originating deeper in the magnetosheath[19]. Observations show that the upstream beam occupies a portion of the phase space that is empty downstream. These observations indicate that the field-aligned beams most likely result from effective scattering in pitch-angle during reflection in the shock ramp.

Substructures. Cluster measurements of large-amplitude structures, which are believed to be the building blocks of collisionless shocks under quasi-parallel conditions, have revealed the surprising result that they appear quite different even at scales 10% of their overall size. Moreover, these differences are not the same in the electric and magnetic components. Thus the previously-believed monoliths are in fact quite filamentary and ethereal[20]. Only at 100 km separation do the four spacecraft records become essentially identical.

The magnetopause

Volume 2 in the SSSI series on "Transport Across the Boundaries of the Magnetopsphere" [5] contained a number of articles[21-24] that reviewed our knowledge of the complex plasma-transfer processes at the magnetopause (MP) at the start of the workshop series on "Source and Loss Processes of Magnetospheric Plasma". The outcome of the workshop deliberations regarding the solar wind as a plasma source was summarized in an article[25] in SSSI Volume 6[6]. One of the

conclusions was that about 10^{28} solar-wind particles cross the magnetopause and enter the magnetosphere every second when the IMF is strongly southward and magnetic reconnection proceeds at its maximum rate (see also the section on "Source and Loss Processes of Magnetospheric Plasma" above).

As in the case of the bow shock, the Cluster mission has provided new insights into the physics of the magnetopause. Among the new results are[26]:

1. The accurate determination of MP orientation, speed and thickness.
2. The measurement of MP currents based on the curlometer method.
3. The determination of the 2D structure of the MP based on integration of the Grad-Shafranov equation.
4. The identification of small-scale structures, in between the ion and electron scales.
5. The identification of cases where reconnection was essentially stationary.
6. The clarification of the causes for flux-transfer events (FTEs).
7. The identification of overturning of Kelvin-Helmholtz waves.
8. The identification of detached solar-wind plasma blobs inside the MP.

Items 6 and 8 are dealt with in the section on "Transient phenomena" above and item 7 in the section on "Waves and Turbulence". In the following we will focus on the remaining items.

MP orientation, speed and thickness. As in the case of the bow shock, one can use the timing of the crossings recorded by the four spacecraft to directly determine the orientation and velocity of the magnetopause and from this calculate its thickness. From an analysis of 24 magnetopause encounters by the four Cluster spacecraft during a single pass along the dawn-side magnetopause, the thickness was found to vary over a wide range, from less than 200 km to thousands of km[27]. In simple models, the thickness of the magnetopause current layer should be determined by the ion gyro-radius, because it determines how deeply the incident solar-wind ions should penetrate Earth's magnetic field. However, the gyro-radius was measured to be only about 50 km for the cases studied. Thus the magnetopause is usually very much thicker than simple theory would predict, and the gyro-radius is not a good scale for the thickness.

Current structure. The internal structure of the MP current sheet has been determined[27] using Cluster's ability to infer the electric currents from a direct application of Ampère's law, which relates the currents to the spatial derivatives of the magnetic field. Cluster is the first mission where this technique, referred to as the "curlometer", could be applied[14].

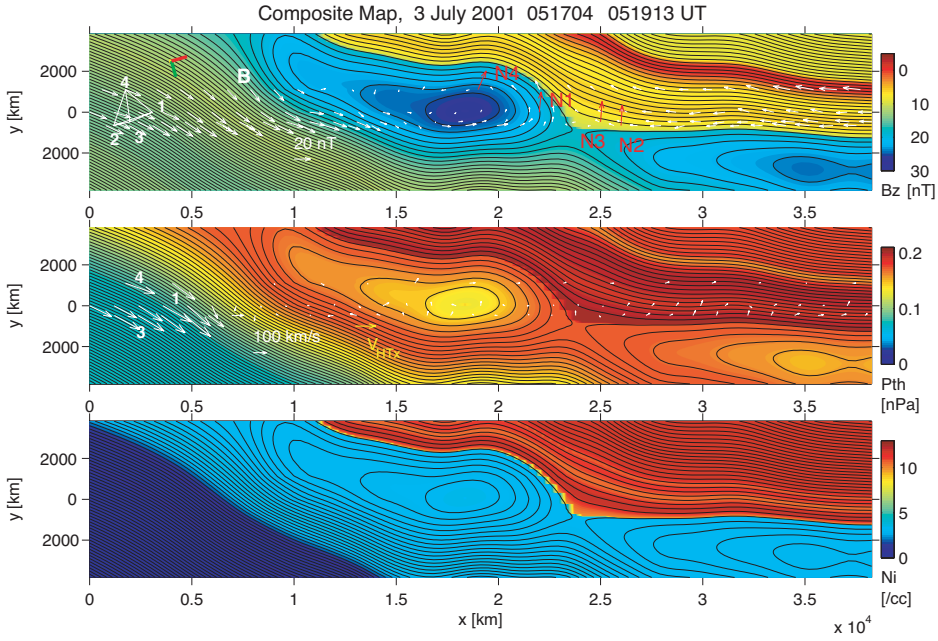


Figure 7. Composite magnetic field maps for the 3 July 2001 event. In all panels the black contour lines represent magnetic field lines projected into the reconstruction plane, as inferred from the integration of the Grad-Shafranov equation, using measurements of the magnetic field and plasma pressure by the four Cluster spacecraft. In the *top panel*, the white arrows show the measured field vectors, anchored at points along the spacecraft trajectory (which run from left to right), and the red arrows show the boundary normal determined from minimum-variance analysis; the colours represent the axial magnetic-field component, which changes across the magnetopause. In the *middle panel*, the white arrows show the measured ion bulk velocity vectors from three spacecraft, transformed into the deHoffmann-Teller frame, and the colours now represent the plasma pressure. In the *bottom panel* the colour shows the ion density (from Ref. 28).

2D structure. Assuming that the magnetopause is in magnetostatic equilibrium and stationary, one can construct a two-dimensional map of the magnetopause from the magnetic-field and plasma measurements. While this technique, based on integration of the Grad-Shafranov equation, was originally developed for single-spacecraft measurements, Cluster has provided the unique opportunity to check these results by comparing the map constructed from measurements from one of the spacecraft, with what the other three are actually observing when they cross the region covered by the map. With the validity of the method thus proven, it was then improved to become a genuine multi-spacecraft technique that produces a single field map by ingesting data from all four Cluster spacecraft. Figure 7 shows an example of such a composite map[28] obtained from a

dawn-side magnetopause crossing. The black lines in all three panels represent the magnetic field lines in the reconstruction plane. The maps not only indicate that the MP is curved at this time, but that magnetic flux ropes are embedded in the current layer, presumably formed by magnetic reconnection that occurred further upstream.

Ion-diffusion region. At the closest spacecraft separations (~100 km), Cluster high-resolution magnetic- and electric-field measurements have allowed small-scale structures within the magnetopause related to magnetic reconnection[29] to be resolved. The left side of Figure 8 shows predictions from a 2D two-fluid MHD numerical simulation of the diffusion region near the reconnection site, using plasma parameters similar to those in the observations. The inset shows the Cluster configuration and their schematic trajectory through the diffusion region. Panel 'a' shows the measured reconnecting magnetic field components, B_L. From the time delay between the current-sheet crossings by the four spacecraft, one can determine the spatial scale of the current layer, as shown at the bottom of the figure. The current-sheet thickness is about 300 km. All four spacecraft observe a very similar structure of the current sheet (panel a), implying a planar structure on the scale of the spacecraft separation.

The out-of-plane magnetic-field component B_M (panel b) shows the bipolar variation expected for the Hall fields. The fact that all four spacecraft observed Hall fields of this large amplitude indicates that this is a stable spatial feature of the diffusion region, rather than some brief temporal variation. The presence of a non-zero normal component of the magnetic field $B_N \approx$ -3 nT (panel c) also suggests ongoing reconnection. As shown in panel 'd', there is good agreement between the measured electric-field component normal to the magnetopause, E_n, and the Hall term ($\mathbf{j} \times \mathbf{B}/ne$) within the narrow region of strong E_n, indicating that E_n can be balanced by the Hall term in the Generalised Ohm's law. This confirms the major role of the Hall term in the formation of the structure of the diffusion region.

Evidence for quasi-continuous reconnection. An important issue is whether magnetic reconnection is necessarily transient in nature or can be operating continuously. When single spacecraft periodically encountered the layer of accelerated flows that are a signature of reconnection, one could never be certain that reconnection did not actually stop in between those isolated encounters. With Cluster, one has now seen cases where the closely spaced encounters of these flows by the four spacecraft start filling in the gaps left by each individual spacecraft alone. Furthermore, in a fortuitous conjunction between Cluster and the IMAGE spacecraft, it has been possible to infer that reconnection was truly continuous over many hours[30].

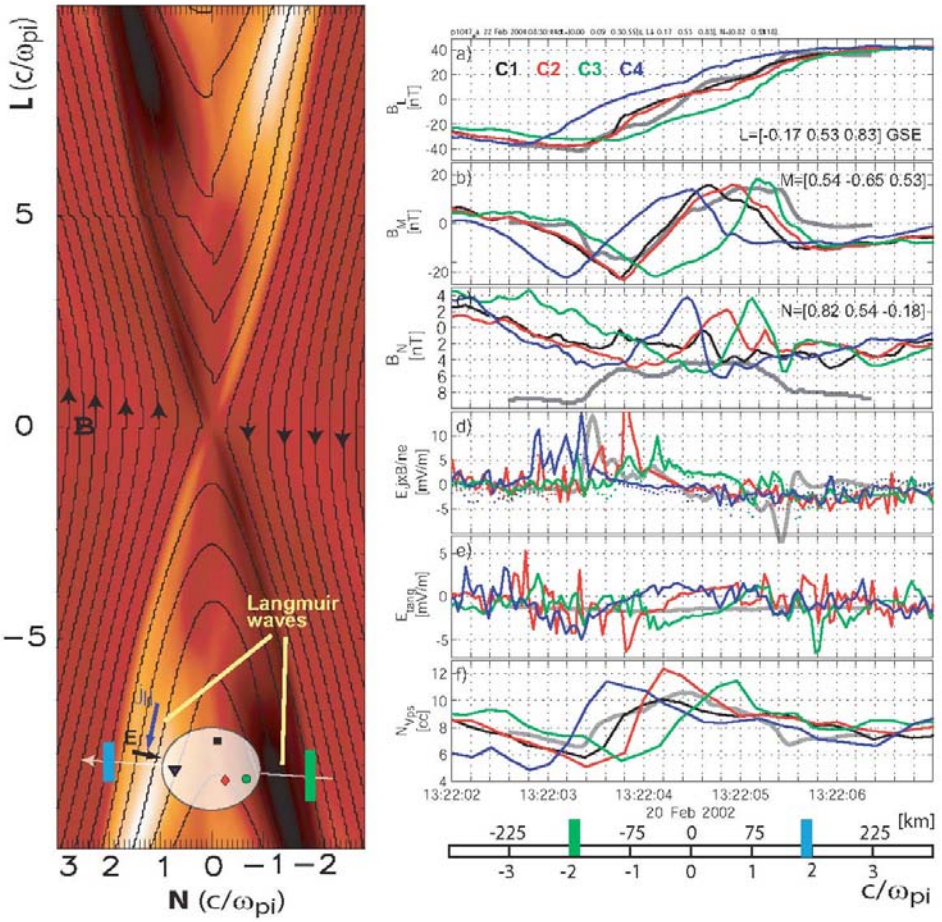


Figure 8. *Left*: Structure of the diffusion region from a numerical Hall fluid simulation of magnetic reconnection. The magnetic field lines, projected into the plane of the figure, are shown as black lines with arrows; their out-of-plane component is shown by the colour: light when pointing into the plane, dark when pointing out of the plane. Also shown is the configuration of the properly colour-coded Cluster spacecraft, and their location relative to the diffusion region. *Right*: Cluster observations: (a) reconnecting magnetic-field component, showing the field reversal across the diffusion region, as illustrated by the black arrows in the picture on the left; (b) out-of-plane magnetic-field component, showing same sign change as the simulation on the left; (c) normal magnetic-field component, which is the component in the horizontal direction in the picture on the left; (d) electric field normal to the structure, E_n (solid lines), and $\mathbf{j} \times \mathbf{B}/ne$ (dotted lines); (e): tangential electric field with average value about 1 mV/m; (f): plasma density from satellite potential. At the bottom, the spatial scale computed from the four-spacecraft magnetopause velocity estimate is given. The grey lines in all panels are "data" extracted from the simulation on the left along a Cluster-like trajectory (from Ref. 29).

Transient Phenomena

Stationary or quasi-stationary phenomena have, for natural reasons, dominated space plasma physics research throughout most of the space era. Only in the last decade or two have the multipoint observations and *in-situ* measurement techniques needed to study dynamic phenomena become available.

Transient events are common in the vicinity of the magnetopause and in the high-latitude ionosphere at the footprints of magnetic field lines that map to the outer magnetosphere. They provide evidence for one or more unsteady solar-wind – magnetosphere interaction mechanisms. Several models, including plasma instabilities, impulsive penetration of solar-wind plasma into the magnetosphere, reconnection highly structured in space and time ("patchy" and "bursty" reconnection), and pressure-pulse-driven boundary waves on the magnetopause, purport to account for the unsteady interaction. Associating events at the magnetopause with those in the ionosphere, and both with the proper driving mechanism, is important. If sufficiently numerous, large and widespread, the transient events might dominate the solar-wind – magnetosphere interaction.

Investigations of transient phenomena have played an important role in several ISSI workshops and the resulting books[5,6,15]. Predicted characteristics of events driven by the Kelvin-Helmholtz plasma instability have been summarized. Just as winds generate waves on the surface of lakes, the shocked solar wind plasma drives quasi-periodic anti-sunward-moving waves on Earth's magnetopause as it flows past the magnetosphere. During periods of enhanced solar-wind velocity, the instability criteria are more likely to be satisfied and the waves should pass an observer more rapidly. The corresponding signatures in the high-latitude ionosphere should be periodic swirls (or vortex flows) that accelerate as they move anti-sunward. Multipoint Cluster observations are presently being used[26] to better determine the characteristics of boundary waves on the magnetopause.

Similarly, expectations for events produced by impulsive penetration have been outlined[23,25]. These events should be associated in a one-to-one manner with "blobs" of solar-wind plasma, whose momenta suffice to traverse the magnetosheath and magnetopause. Once inside the magnetosphere, the events should decelerate as they move anti-sunward. The corresponding ionospheric signatures should be isolated convection vortices that decelerate as they move anti-sunward and equatorward. Doubts have been raised about the existence of such plasma blobs, whether they can retain an excess momentum, and whether they can penetrate the magnetopause. Nevertheless, some recent Cluster observations indicate[27] that blobs of solar-wind plasma do penetrate the magnetopause and become detached.

Bursty merging on the dayside magnetopause generates bubbles of intermixed magnetosheath and magnetospheric plasmas on interconnected magnetic field lines that bulge outward into both regions. During periods of southward interplanetary magnetic field (IMF) orientation, the bubbles form along and move away from a line passing through the sub-solar point whose tilt depends upon the IMF orientation. Magnetic curvature forces pull bubbles connected to the northern ionosphere dawnward during periods of duskward IMF orientation and duskward during periods of dawnward IMF orientation. Because the bubbles bulge outward into both the magnetosheath and the magnetosphere, their passage generates characteristic bipolar magnetic-field signatures in the direction normal to the nominal magnetopause. These signatures are known as flux-transfer events, or FTEs.

Recent Cluster observations confirm[30] that FTEs occur on the equatorial magnetopause for southward IMF orientations, they recur each ~8 min, they exhibit velocities of ~100 km s^{-1}, and they have ~1 R_E dimensions along and perpendicular to the magnetopause. The observations also confirm that at least some events bulge outwards into both the magnetosheath and magnetosphere. By means of special theoretical methods (Grad-Shafranov techniques) applied to the Cluster observations, it has been possible to reconstruct the magnetic-field structure within and outside one such FTE[31]. The internal structure of some FTEs may be best explained in terms of a coaxial current pattern. The dimensions of at least two FTEs along the magnetopause were greater further within the magnetosphere than near the magnetopause, contradicting expectations from existing FTE models. Some transient events observed on the edge of the interior magnetic cusp exhibit the plasma signatures expected for FTEs, but not the bipolar signatures normal to the nominal magnetopause.

The ionospheric signatures originally predicted for FTEs were azimuthally- and poleward-moving convection vortices. More recently, FTEs have been associated with latitudinally limited, but longitudinally extensive, poleward-moving transient azimuthal flux bursts[32]. These events tend to occur for southward IMF orientations, recur every 8 min, and exhibit dawnward or duskward flows in accordance with merging model predictions and FTE observations. Cluster measurements have been used to confirm the simultaneous existence of recurring FTEs at the magnetopause and recurring poleward-moving flow bursts and auroral forms in the ionosphere.

Although the intrinsic pressure variations associated with interplanetary shocks can be large, kinetic processes within Earth's foreshock almost constantly introduce transient density and dynamic pressure variations with even greater amplitudes into the incoming solar wind. The pressure variations batter the magnetopause, drive large-amplitude boundary waves, and launch fast mode compressional waves into

the magnetosphere. The boundary waves move dawnward across local noon during typical periods of spiral IMF orientation, and duskward across local noon during periods of ortho-spiral IMF orientation. Although the north/south orientation of the IMF should have no bearing upon the occurrence of the wave patterns, they should be more common during periods of radial IMF orientation and high solar-wind velocities, because under these conditions the foreshock lies upstream of the dayside magnetopause and more energy is available to drive the kinetic processes.

ISSI teams have considered the magnetospheric and ionospheric response to intrinsic solar-wind pressure variations and those generated within the foreshock. The first evidence indicating that kinetic processes at the bow shock were capable of driving large-amplitude magnetopause motion was reported by one of the teams[33]. As illustrated in Figure 9, kinetic processes associated with the interaction of an otherwise undistinguished solar-wind tangential discontinuity with Earth's bow shock resulted in a dramatic, but local, decrease in the solar-wind dynamic pressure applied to the magnetosphere. The drop in the pressure permitted the magnetopause to bulge outwards and created a gross deformation that propagated anti-sunward with the solar wind flow. Other authors subsequently associated this event with perturbations in the geosynchronous magnetic-field strength, transient auroral brightenings, and travelling convection vortices (TCVs) in the ionospheric flow.

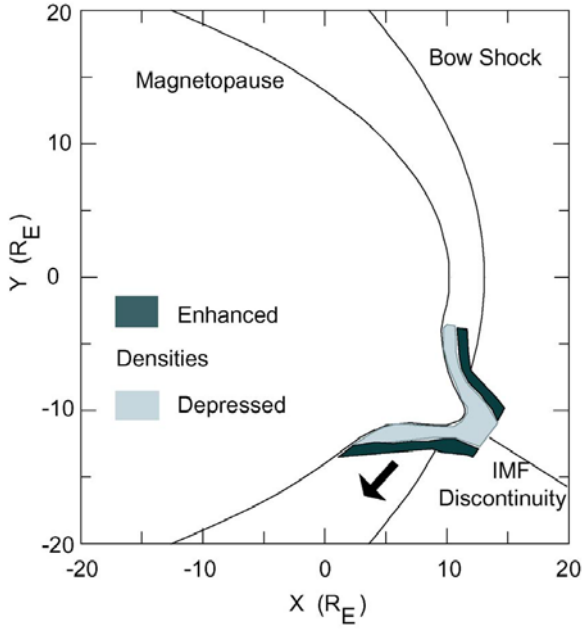


Figure 9. Kinetic processes, associated with the interaction of an interplanetary-magnetic-field (IMF) discontinuity with Earth's bow shock, generated a density cavity, which allowed the magnetopause to expand outward, resulting in a large-amplitude anti-sunward propagating wave on the magnetopause boundary (from Ref. 34).

Because spacecraft are seldom located in the foreshock, they rarely observe the pressure variations striking the magnetosphere. Another ISSI study[34] explored the possibility of employing equatorial ground magnetograms to track the pressure variations striking the magnetosphere. It has long been known that each and every abrupt change in the solar wind's dynamic pressure generates a fast mode wave that propagates into the magnetosphere and that the signature of this fast mode wave is enhanced in equatorial dayside ground magnetograms. However, there are many signatures in the equatorial magnetograms and it is not clear if each of them has a corresponding solar-wind signature. The team reported that about half the abrupt signatures they observed in equatorial ground magnetograms could be associated with variations in the solar-wind pressure, but that half could be associated with substorm onsets.

By contrast, all compressional signatures at the dayside geosynchronous orbit are associated with variations in the solar-wind pressure applied to the magnetosphere. Sequences of transient events in the high-latitude ionosphere were associated with corresponding compressional signatures in the magnetic-field and energetic-particle signatures observed by geosynchronous spacecraft. These signatures at geosynchronous orbit move dawnward or duskward in the direction predicted by the pressure-pulse model for the given spiral/ortho-spiral IMF orientation[35].

An ISSI team has also investigated the characteristics of travelling convection vortices (TCVs). A detailed case study of a single TCV indicated that it was produced by a pair of field-aligned currents flowing alternatively out of and then into both hemispheres[36,37]. In contrast to model predictions, the current maxima occurred on the perimeter of the vortices.

Another study[38] reported that the Geotail spacecraft had observed the flow pattern expected for a large-amplitude boundary wave in the vicinity of the equatorial magnetopause at the longitude and time of an ionospheric TCV. Using high-time-resolution observations to track the motion of the pressure front associated with a TCV through both the equatorial magnetosphere and ionosphere, it was shown[39] that this motion was consistent with foreshock-generated pressure pulses striking the magnetosphere for the observed IMF orientation. A model for the field-aligned currents associated with boundary motion, which did not make simplifying assumptions and therefore settled an ongoing controversy, was presented.

Statistical studies[40,41] have demonstrated that there is a statistically significant relationship between TCVs and higher frequency Pc3 (22 - 100 mHz) pulsations, both of which propagate away from local noon with similar speeds. Because the foreshock is a well-known source of Pc3 pulsations, these results indicate a close

connection between TCVs and pressure variations generated in the foreshock. A statistical survey of TCV occurrence patterns as a function of solar-wind conditions has given a number of conflicting results. Consistent with the predictions of both the Kelvin-Helmholtz and pressure-pulse models, solar-wind velocity has been found to control the occurrence and strength of TCVs. IMF Bz determines the occurrence of events near local noon, as the bursty reconnection model predicts. The standard deviations of various IMF parameters also control event occurrence, as expected from the pressure-pulse model. Much work remains to be done to explain the complex relations between the solar-wind properties and their effects in the ionosphere. In particular, it appears that events produced by different mechanisms can exhibit similar signatures.

Concluding remarks. ISSI workshops have provided comprehensive reviews of our present understanding of transient events at the magnetopause, while the international teams have contributed case studies and a plan of attack describing how the characteristics of transient events in the ionosphere can be related to those at the magnetopause, and how both can be associated with the driving mechanism. Nevertheless, many questions remain unanswered. Fortunately, Cluster observations and forthcoming NASA THEMIS observations should prove decisive in this regard. Cluster observations can be used to discriminate between boundary waves and FTEs bulging into both the magnetosphere and the magnetosheath. They can also be used to determine the direction and speed of event motion as a function of solar-wind conditions, an important discriminator between the proposed models. THEMIS is a mission consisting of five spacecraft: three at the magnetopause, one in the foreshock, and one in the solar wind. Its observations will be used to determine whether fluctuations in the solar-wind dynamic pressure or IMF orientation trigger the occurrence of FTEs, and to define the response of the magnetosphere to pressure pulses generated within the foreshock. Its observations will also be used to determine when and where boundary waves result from the Kelvin-Helmholtz instability or from magnetopause motion directly driven by variations in the solar-wind dynamic pressure. Questions like these are left for future workshops.

Waves and Turbulence

Collisionless plasmas are not in strict thermal equilibrium. They are thus subject to the excitation of large-amplitude fluctuations annihilating and dispersing the excess energy. Such unstable fluctuations appear in the form of waves, wave packets, turbulence, and radiation. Radiation escapes from the plasma carrying away energy. Waves confined to the plasma, in contrast, disperse and redistribute energy throughout the plasma, making it available at remote places and for

processes like plasma heating, dissipation and generation of turbulence. They trigger violent transitions, like "magnetic reconnection", formation of "collisionless shocks", and "diffusive transport" of particles and magnetic field.

Wave excitation. Plasma waves are excited by spatial inhomogeneities and by deviations of the particle velocity distributions from the equilibrium (Maxwellian) distribution. In either case the free energy available is converted into fluctuations resonant with the plasma particles, a process described as "wave-particle interaction". This means that electrically charged particles not in equilibrium, the resonant particles, start oscillating, thereby exciting a fluctuating electromagnetic wave field. Strong wave excitation proceeds preferentially in active plasma regions such as plasma boundaries and current sheets, like the heliospheric current sheet, shock waves and their environments, magnetopauses, ionopauses, in planetary magnetotails, auroral zones and radiation belts.

Turbulence. Whenever many waves are present in a plasma, the fluctuation spectrum is not resolvable into single wave modes. In this case the plasma is turbulent. In contrast to wave-particle interaction, turbulence evolves from interaction among waves. Turbulence does not exhibit full disorder; rather it contains "order in chaos". It is structured into eddies, vortices, and plasma clumps of different sizes. Turbulence is "multi-scale". Magnetic turbulence is of low frequency and macroscopic scale, related to current filaments and current sheets. It is observed all over in space plasmas, from the solar corona through the solar wind, interplanetary shocks, planetary and cometary bow-shock foreshock regions, magnetosheaths, magnetopause transitions, and plasma sheets in magnetospheric tails, wherever the kinetic energy density in the plasma exceeds the magnetic.

Radiation. While serving as a sensitive tracer of microscopic plasma processes, radiation is usually ineffective for space plasmas (contrary to in astrophysics) in releasing free energy. Very small amounts of energy only are converted into radiation. In the collisionless plasmas in near-Earth space, radiation is in the domain of radio-wavelengths. Measurably high power in radiation is found near shock waves, where it is generated at harmonics of the local electron plasma frequency $f_{pe} \propto \sqrt{n}$ by head-on collisions between plasma waves excited by electron beams. It provides a measure of the local plasma density n. In magnetospheres in relation to aurorae, intense radiation, the celebrated auroral kilometric radiation (or AKR), is provided by energetic electron velocity distributions, which form in the presence of a magnetic field-aligned electric-potential drop (at Earth at altitudes of 2000 - 8000 km during substorm activity). Distributions of this kind resemble "excited states" known to release their excess energy like masers and lasers instantaneously in concert and offering high radiation yields. Such radiation is at the electron cyclotron frequency f_{ce} , which is proportional to the magnetic field strength, B, providing a measure of the latter.

Earth's wave environment

Wave processes have been a focus of scientific activity at ISSI, encompassing the debate about the acceleration and transport constituting the "Astrophysics of Galactic Cosmic Rays" (GCR)[42], where fluctuations in the interstellar medium have been identified[43] as scatterers of GCR. The "heliospheric laboratory" served as a paradigm for models of GCR acceleration[44, 45]. Turbulent magnetic fields are deemed responsible for CR scattering deep into and across the heliosphere[46]. The main focus at ISSI was on the many plasma-wave phenomena in the solar wind, magnetosheath and magnetosphere. These include shock-wave formation and the acceleration of charged particles at collisionless shocks[16], the generation of magnetosheath turbulence behind Earth's bow shock[5,6,15], magnetic reconnection at the magnetopause and in the magnetotail plasma sheet, and the role of waves in the physics of the aurora[10]. In addition to the ISSI Workshops, 19 out of 77, i.e. ~25% of the ISSI teams dealt with some aspect of plasma waves or turbulence.

Waves and turbulence provided the guidelines in the discussion of the generation of the polar wind and the acceleration of ions into ion conics[7], investigation of particle precipitation[5,47], analysis of reconnection[22] and diffusive entry[24] at the magnetopause, discussions which have been continued, extended and deepened in the follow-up workshop on "Magnetospheric Sources and Losses" [6] and on "Auroral Plasma Physics" [10], which draws extensively from the importance of wave phenomena in the physics of the aurora. Finally, the recent ISSI workshop "Magnetospheric Boundaries and Turbulence: Cluster Results" [15], which dealt with the dayside solar-wind/bow-shock/magneto-sheath/cusp/magnetopause transition, emphasizes the generation of waves and turbulence in the interaction between the solar wind and the magnetosphere. Additional effort has been made by smaller teams, ranging from the investigation of the microscopic dynamics of shocked plasmas[16], and the generation of so-far-unexplained power-law distributions in otherwise collisionless space plasmas[48], up to the detection of extra-solar planets by observation of their radio emissions[49].

Waves at plasma boundaries

Magnetopause surface waves. Figure 10 shows what dominant mode eroding the magnetopause is to be expected when the close-to-sonic magnetosheath plasma flow passes over the magnetopause and exerts friction and stress on the magnetospheric magnetic field. The friction forces the magnetopause to develop ripples and vortex-like circulations of plasma and magnetic field. The passing magnetosheath plasma and field experience adhesion and retardation on these ripples. Waves of this kind are very-large-scale, in the order of several Earth radii in wavelength along the magnetopause[50]. When decaying into smaller structures, the turbulent eddies and vortices reach down to length scales of only hundreds

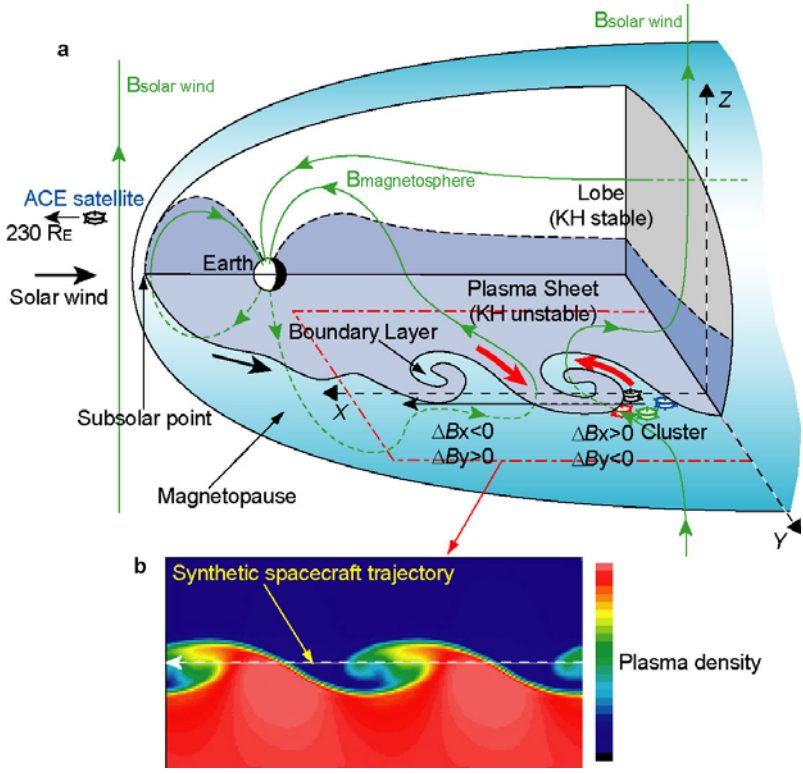


Figure 10. *Upper Part*: A three-dimensional cut-away view of Earth's magnetosphere with the signatures of Kelvin-Helmholtz ripples on the dusk-side flank of the magnetopause. The Cluster spacecraft were located near the dusk equatorial plane, and were separated by 2000 km from each other. The red arrows show the inferred circulation of plasma in the Kelvin-Helmholtz vortices in the boundary layer. Indicated are the signs of the observed variations in the components of the magnetic field **B**. *Lower Part*: Formation of vortices and mixing, resulting from a 3D numerical simulation of the magnetohydrodynamic KHI under a magnetosphere-like geometry. Density, velocity and magnetic-field data extracted from the simulation along the trajectory shown by the horizontal line agree quite well with the actual Cluster observations (from Ref. 51).

of kilometres. They cause mixing of plasmas and fields on both sides of the magnetopause (Fig. 10) at the downstream flanks[51]. In this way, they provide transport of plasma across the magnetopause and the formation of a boundary layer.

Magnetosheath and boundary-layer turbulence. Figure 11 (left part) shows a typical magnetosheath wave power spectrum close to the dayside magnetopause. Most of the wave energy is stored in low-frequency magnetic fluctuations, $\delta\mathbf{B}$. At higher frequency, electric-field fluctuations, $\delta\mathbf{E}$, take over. Cluster multi-spacecraft measurements have shown[52] this turbulence to consist of a superposition of low-frequency magnetic oscillations like Alfvén and mirror waves.

Alfvén waves are string-like oscillations of magnetic field lines, while mirror waves cause magnetic depressions in the plasma. Alfvén waves are excited by reflected ion beams in the foreshock, from where they enter the magnetosheath[52]. Mirror waves occur when the plasma has higher pressure perpendicular to than parallel to the magnetic field. This is the normal situation in the magnetosheath.

ISSI provided the forum for the development of the sophisticated analysis techniques[14] for multi-spacecraft missions (the so-called "k-filtering" methods[53]) by which the dispersion of plasma waves could be determined and the wave modes identified. The agreement between the inferred mirror wave iso-contours and the mirror dispersion relation is visible from one particular reconstruction in Figure 11 (right part). The investigation of the linear and nonlinear theory of the mirror mode has formed part of the theoretical studies of three ISSI teams, leading to new insight into mirror turbulence.

At polar-cusp latitudes, the low-frequency boundary-layer spectral shapes differ from the upstream magnetosheath and solar wind. They are the result of the development of "Self-Organized Criticality" and turbulence, investigated by visiting scientists at ISSI and a number of ISSI teams.

Electric waves, contributing to reconnection and diffusion. Electric waves at the magnetopause are excited by the magnetopause density gradient and the diamagnetic magnetopause current. They scatter the plasma particles out of their orbits and thus contribute to diffusive plasma entry across the magnetopause, formation of the magnetospheric boundary layer, and magnetic reconnection. Earlier estimates based on other spacecraft observations seemed to deny this possibility[6]. New Cluster measurements reported in an ISSI workshop[54] have detected very strong electric-wave activity on field lines connecting to the magnetopause reconnection region, confirming numerical simulation studies[55] of their importance. Waves of this kind are also the signature of electron beams accelerated in reconnection and reaching the spacecraft along the magnetic field.

Auroral plasma waves

Wave activity in the magnetosphere is most intense at nightside auroral latitudes between 1000 and 8000 km altitude. Figure 3 schematically summarized the relevant wave phenomena related to the aurora in the upward and downward auroral-current regions. The two dominant families of waves in the auroral magnetosphere are VLF waves and a sporadic high-frequency electromagnetic radiation at km-wavelengths, called auroral kilometric radiation (AKR).

VLF emissions. The lowest frequency waves in the aurora are Alfvén waves. Their properties have been summarized in a review paper[11] written as a team

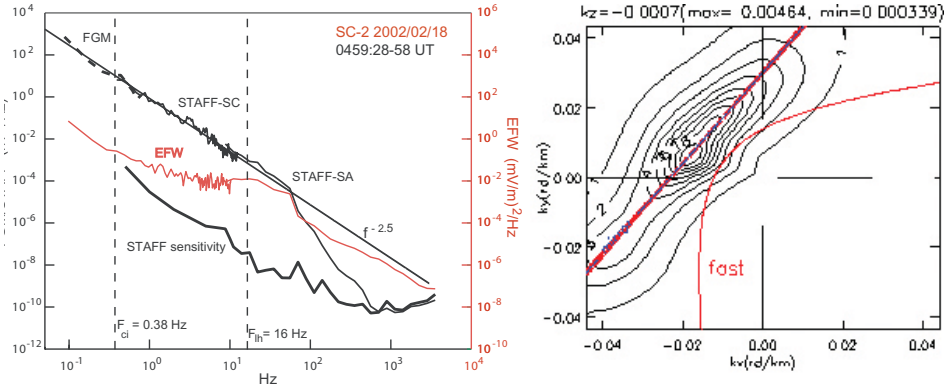


Figure 11. *Left*: Cluster measurements of low-frequency magnetic (blue) and electric (red) wave power spectra in the magnetosheath close to the magnetopause, exhibiting the power-law dependence of the wave power (vertical axis) on frequency (horizontal axis), which is typical for well-developed turbulence. The magnetic power cuts off at the electron plasma frequency, while the electric power exhibits a maximum at the lower-hybrid frequency. *Right*: Reconstruction of the wave-mode constituents in the magnetosheath based on the k-filtering technique, showing that in this particular case the main spectral contribution at very low frequencies is the mirror mode. The figure shows, for a fixed frequency of 0.6 Hz and a selected wave vector component k_Z, parallel to the magnetic field, the iso-contours of wave power in the plane perpendicular to the magnetic field. Superimposed (in red) are the dispersion curves for the mirror mode and the fast mode (from Ref. 52).

effort at ISSI. Auroral VLF waves have been discussed in an ISSI-related review paper[56]. A large variety of different modes in the whistler frequency band between the ion and electron cyclotron frequencies contribute to VLF. These waves are partly oscillations of the electrostatic potential in resonance with the electrons. In their presence, the auroral electron beam and ring distribution is grossly deformed[57]. It is one of the achievements of the "Auroral Plasma Physics" workshop activity at ISSI that this point has ultimately been clarified. Previously, this kind of deformation was constantly attributed solely to the action of the AKR.

One particular type of VLF radiation is the "VLF-saucer" emission (which takes its name from the saucer-like form of its dynamic spectrum), a spectacular example of which is shown in Figure 12. Saucers are electromagnetic waves belonging to the whistler family, known from disturbances of broadcasting signals by lightning discharges. Saucers are excited at a very narrow spatial location in the presence of high fluxes of field-aligned ~100 eV electrons in the aurora (described in the auroral section). Saucers thus indicate the presence of small-scale structures (so-called "phase space holes") in the auroral downward-current region[10].

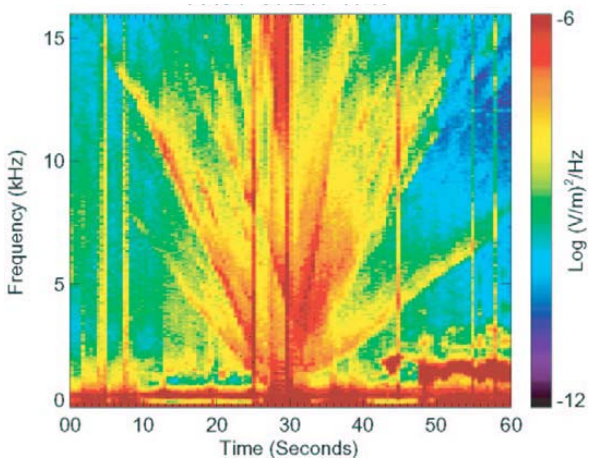


Figure 12. Frequency-time spectrogram of the electric wave power (colour-coded as shown by the bar on the right) observed by the FAST spacecraft in the auroral magnetosphere at an altitude of ~4000 km, showing a spectacular sequence of so-called saucer emissions. All wave traces start from a single narrow spatial location (from Ref. 57).

Auroral kilometric radiation. In the upward field-aligned auroral-current regions, the auroral electron beam impinges on the ionosphere and a parallel electric field evacuates the ionospheric plasma. Here AKR is generated in a process similar to maser emission[57]. The tenuous upper auroral plasma is unable to dissipate the auroral electron-beam energy other than by generating radiation. AKR radiates away into free space several percent of the total energy of a substorm. AKR is generated perpendicular to the magnetic field by the slightly relativistic auroral-beam electrons. The emission frequency in this case is below the local electron cyclotron frequency.

Closer inspection of the spectrum showed that AKR consists of very narrow emission lines, "elementary radiators". These elementary radiators are directly related to the electron holes. This fascinating radiation mechanism has tremendous importance in prospective applications to astrophysical conditions, though its microscopic relation to holes has not yet been fully clarified. It is one of the achievements of the ISSI auroral-wave activities that this problem was explicitly formulated.

Wave-particle interaction and particle precipitation

Plasma-wave turbulence in the radiation belts and plasma sheet[58] causes precipitation of energetic particles into the ionosphere. Historically, this effect was the

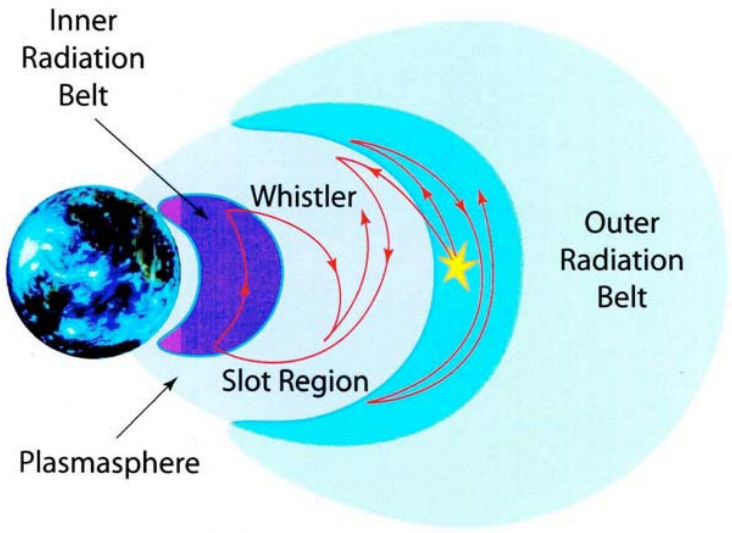


Figure 13. The Van Allen radiation belts and the formation of the slot region (shown in dark blue) between the inner and outer belts (shown in light blue), by generation of whistler waves (red lines) that can cause particle loss into the atmosphere in a collision shown as the yellow star-symbol (from Ref. 62).

first realization of the importance of plasma turbulence. Precipitation[59] represents loss of particles from the magnetosphere to the ionosphere; it loads the ionosphere with plasma, being of higher energy than the solar UV-generated ionospheric populations and causing an increase in the ionospheric conductivity. It closes the ionospheric current system via field-aligned currents, and it is believed to contribute to, if not to cause, the diffuse aurorae.

Pitch-angle diffusion. Precipitation results from scattering of magnetospheric particles by plasma waves[60]. Collisions with the waves rotate the particle velocity vectors towards becoming more parallel to the magnetic field. This process is called "pitch-angle diffusion" and was at the heart of the work of two visiting senior scientists (Drs. Rycroft and Trakhtengerts), which will result in a separate monograph.

Radiation-belt lifetimes. Energetic, > 40 keV, Van Allen-radiation-belt electrons excite a broad spectrum of whistler waves. Pitch-angle scattering at these whistlers limits the lifetime of the particles in the radiation belts. In the inner belt, scattering is rather inefficient and lifeimes are long, explaining the stability of the inner Van Allen belt over years. Wave excitation and scattering is strongest at 4-5 R_E equatorial distance from Earth. Here lifetimes are short, and a gap is gouged into the radiation belts (cf. Fig. 13), forming the "radiation belt

slot" that separates the inner belt from the less energetic outer belt. Ions undergo similar interactions with electromagnetic ion-cyclotron waves and precipitate predominantly at lower latitudes, where they cause the mid-latitude "auroral red arcs" during magnetospheric storms.

Plasma-sheet precipitation. The plasma sheet does not generate substantial whistler activity. Waves of presumably electrostatic nature are responsible for precipitation of electrons and ions from the plasma sheet. This remains an unresolved problem that was extensively discussed at the ISSI workshops[61] in relation to diffuse aurorae and substorm onset. Continuous field-aligned inflow of plasma-sheet electrons may be due to a broad spectrum of waves excited in the plasma sheet, and contributes to diffuse aurora. The more violent processes of substorms are related to the reconnection problem. Today's understanding of the wave processes in the plasma sheet is still at a premature stage. Its resolution requires the complete understanding of the collisionless reconnection problem via high-resolution multi-spacecraft observation and three-dimensional numerical simulations.

Concluding Remarks

Space plasma physics exploits for research the only directly accessible reservoir of collisionless plasma in the entire Universe[62]. As such, space plasma physics is not only of interest in itself as the experimental science that enables investigation of the fourth state of matter *in-situ*. Since most of the baryonic matter in the Universe is known to be in this state, it provides the paradigm for the processes taking place also in very remote cosmic regions. These regions reach out from the magnetospheres and ionospheres of the planets in our planetary system, through the magnetospheres and magnetized winds of stars, the solar and stellar atmospheres and coronae, stellar spheres comparable to our own heliosphere, to the magnetospheres of neutron stars, and to the formation of magnetic boundaries, reconnection, shock-wave formation, shock structure, particle acceleration, diffusion processes, and generation of radiation in collisionless plasma. Aurorae have been detected on the large magnetized planets of our Solar System. Similar phenomena take place in the solar corona and are expected to be found in the atmospheres of other stars and on exo-planets. Magnetospheres are known to exist around neutron stars and white dwarfs. Shock waves can be found everywhere in the galaxies accelerating particles to cosmic-ray energies.

The short review of the ISSI activities in the field of space plasma physics that has been given here is intended to provide an impression of the achievements over the past decade. It may also give a few hints regarding extrapolation to

remote systems. Some of these extrapolations have already been made early in the history of space plasma physics, when the concept of stellar winds was drawn up from the solar wind, and planetary as well as neutron-star magnetospheres were modelled along the lines of Earth's magnetosphere. However, the more subtle knowledge of microscopic space plasma physics accumulated in recent years leads us to expect that some of these pictures will change at least as much as space plasma physics itself has been modified, leading to deeper understanding of internal processes.

References

1 E.G. Shelley *et al*., *J. Geophys. Res*., **77**, 6104, 1972.
2. E.G. Shelley *et al*., *Geophys. Res. Lett*., **3**, 654, 1976.
3. J. Geiss *et al*., *Space Sci. Rev*., **22**, 537,1978.
4. C.R. Chappell *et al*., *J. Geophys. Res*., **92**, 5896, 1987.
5. B. Hultqvist & M. Øieroset (Eds.), Transport Across the Boundaries of the Magnetosphere, SSSI Vol. 2, Kluwer Academic Publ., Dordrecht, 1997, and *Space Sci. Rev.,* **80**, Nos. 1-2, 1997.
6. B. Hultqvist, M. Øieroset, G. Paschmann & R. Treumann (Eds.), Magnetospheric Plasma Sources and Losses, SSSI Vol. 6, Kluwer Academic Publ., Dordrecht, 1999, and *Space Sci. Rev.,* **88**, Nos. 1-2, 1999.
7. A.W. Yau & M. André, in Ref. 5, p.1.
8. D.A. Hardy *et al*., *J. Geophys. Res.,* **90**, 4229, 1985 and dto. **94**, 370, 1989.
9. R.J. Walker *et al*., *Physics of Space Plasmas*, p. 561, AGU, 1996.
10. G. Paschmann, S. Haaland & R. Treumann (Eds.), Auroral Plasma Physics, SSSI Vol. 15, Kluwer Academic Publ., Dordrecht, 2002, and *Space Sci. Rev.,* **103**, Nos. 1-4, 2002.
11. K. Stasiewitz *et al*., *Space Sci. Rev.,* **92**, 423, 2000.
12. Ref. 10, Chapter 2.
13. Ref. 10, Chapter 4.
14. G. Paschmann & P. Daly (Eds), Analysis Methods for Multi-Spacecraft Data, ISSI-SR 1, ESA, 2000.
15. G. Paschmann, S. Schwartz, P. Escoubet & S. Haaland (Eds.), Outer Magnetospheric Boundaries: Cluster Results, SSSI Vol. 20, Springer Verlag, Dordrecht, 2005, and *Space Sci. Rev.* (in press) 2005.
16. B. Lembège *et al*., *Space Sci. Rev*., **110**, 161, 2004.
17. Ref. 15, Part II (The Bow Shock).
18. Ref. 15, Part II, Chapter 2 (Quasiperpendicular Shock Structure).
19. Ref. 15, Part II, Chapter 2 ; K. Kucharek *et al., Ann. Geophys.***, 22**, 2301, 2004.
20. Ref. 15, Part II, Chapter 3 (Quasiparallel Shock Structure).
21. G. Paschmann, in Ref. 5, p. 217.
22. J.D. Scudder, in Ref. 5, p. 235.

23. R. Lundin, in Ref. 5, p. 269.
24. M. Scholer & R.A. Treumann, in Ref. 5, p. 341.
25. D.G. Sibeck *et al*., in Ref. 6, p. 207.
26. Ref. 15, Part III (Magnetopause and Cusp).
27. Ref. 15, Part III, Chapter 1 (Magnetopause and Boundary Layer).
28. Ref. 15, Part III, Chapter 1; H. Hasegawa *et al*., *Ann Geophys*., **22**, 1251, 2004.
29. Ref. 15, Part III, Chapter 3 (Magnetopause Processes); A. Vaivads *et al*., *Phys. Rev. Lett*., **93**, 105001, 2004.
30. Ref. 15, Part III, Chapter 3; T.D. Phan *et al*., *Ann. Geophys*., **22**, 2355, 2004.
31. Ref. 15, Part III, Chapter 1; B.U.Ö. Sonnerup *et al*., *Geophys. Res. Lett*., **31**, 10.1029/2004GL020134, 2004.
32. e.g. M. Lockwood *et al*., *J. Geophys. Res.,* **95**, 17117, 1990.
33. D.G. Sibeck *et al*., *Geophys. Res. Lett*., **25**, 453, 1998.
34. D.G. Sibeck *et al*., *J. Geophys. Res*., **103**, 6763, 1998.
35. G.I. Korotov *et al*., *J. Geophys. Res*., **107**, 10.1029/2002JA009477, 2002.
36. D.L. Murr *et al*., *J. Geophys. Res*., **107**, 10.1029/2002JA009456, 2002.
37. O. Amm *et al*., *J. Geophys. Res*., **107**, 10.1029/2002JA009472, 2002.
38. T.M. Moretto *et al*., *J. Geophys. Res*., **107**, 10.1029/2001JA000049, 2002.
39. D.G. Sibeck *et al*., *J. Geophys. Res*., **108**, 10.1029/2002JA009675, 2003.
40. D.W. Shields *et al*., *J. Geophys. Res*., **108**, 10.1029/2002JA009397, 2003.
41. C.R. Clauer & V.G. Petrov, *J. Geophys. Res*., **107**, 10-1029/2001JA000228, 2002.
42. R. Diehl, E. Parizot, R. Kallenbach & R. von Steiger (Eds.), The Astrophysics of Galactic Cosmic Rays, SSSI Vol. 13, Kluwer Academic Publ., Dordrecht, 2001, and *Space Sci. Rev.,* **99**, Nos. 1-4, 2001.
43. G.M. Mason, in Ref. 42, p.119; R.A. Treumann & T. Terasawa, in Ref. 36, p.135.
44. S.R. Spangler, in Ref. 42, p.261; E. Bereshko, p. 295; D.C. Ellison, p. 305; A.M. Bykov, p. 317.
45. J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger & R. Kallenbach (Eds.), Cosmic Rays and Earth, SSSI Vol.10, Kluwer Academic Publ., Dordrecht, 2000, and *Space Sci. Rev.,* **93**, Nos. 1-2, 2000.
46. W. Droege, in Ref. 45, p. 121.
47. T.G. Onsager & M. Lockwood, in Ref. 5, p. 77; L.R. Lyons, p. 109; H. Koskinen, p. 122.
48. cf. Ref. 6, Section 5.4; R.A. Treumann, *Phys. Scripta,* **59**, 19 and 253, 1999.
49. P. Zarka *et al*., *Astrophys. Space Sci*., **277**, 293, 2001.
50. Ref. 6, Section 5.5 on the Kelvin-Helmholtz instability.
51. Ref. 15, Part III, Chapter 3 (Magnetopause Processes); H. Hasegawa *et al*., *Nature,* **430**, 755, 2004.
52. Ref. 15, Part I, Chapter 3 (Magnetosheath).
53. J.-L. Pinçon & U. Motschmann, in Ref. 14, p. 65.
54. cf. Ref. 15, part III, Chapter 3 (Magnetopause Processes); A. Vaivads *et al*., *Geophys. Res. Lett*., **31**, 10.1029/2003GL018142, 2004.
55. B. Rogers *et al*., *Phys. Rev. Lett*., **87**, 195004, 2001; M. Scholer, *Phys. Plasmas,* **10**, 3521, 2003.

56. J. LaBelle & R.A. Treumann, *Space Sci. Rev*., **101**, 295, 2002.
57. cf. Ref. 10, Chapter 4 (In-situ Measurements in the Auroral Plasma).
58. L.R. Lyons, in Ref. 5, p. 109; H.E. Koskinen, in Ref. 5, p. 133.
59. cf. Ref. 10, Section 4.3 (Waves and Radiation).
60. dto., Section 3.5 (Wave-Particle Interactions).
61. cf. Ref. 6, Section 3.4 (Precipitation from the Plasma Sheet).
62. J.A.M. Bleeker, J. Geiss & M.C.E. Huber (Eds.), The Century of Space Science, Vol. 2, p. 1495, Kluwer Academic Publ., Dordrecht, 2003.

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

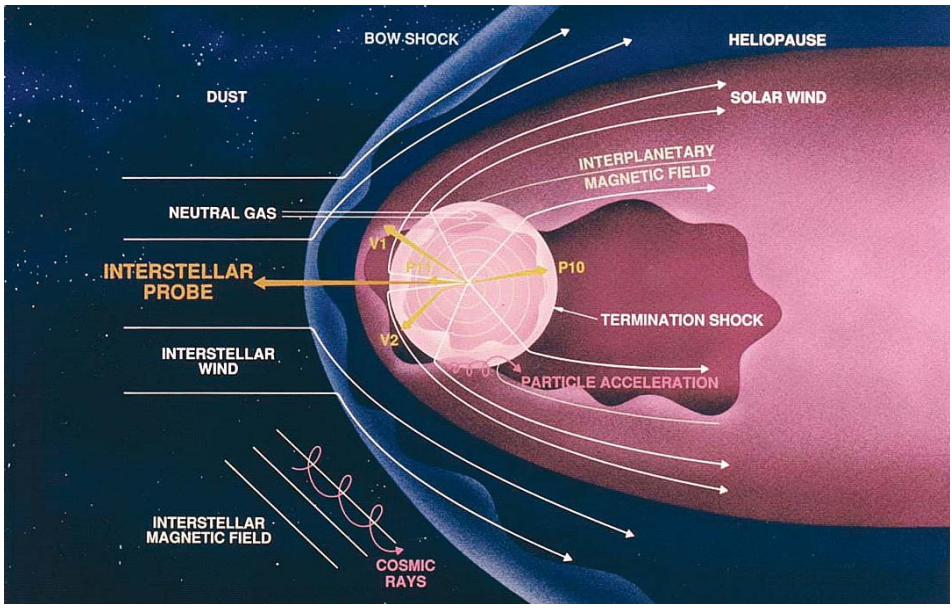


Figure 1. The heliosphere and the heliospheric interface. The heliopause is a contact discontinuity which separates the solar wind from interstellar plasma component. The termination and bow shocks are formed to decelerate the supersonic solar and interstellar winds, respectively, before these ionised 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,

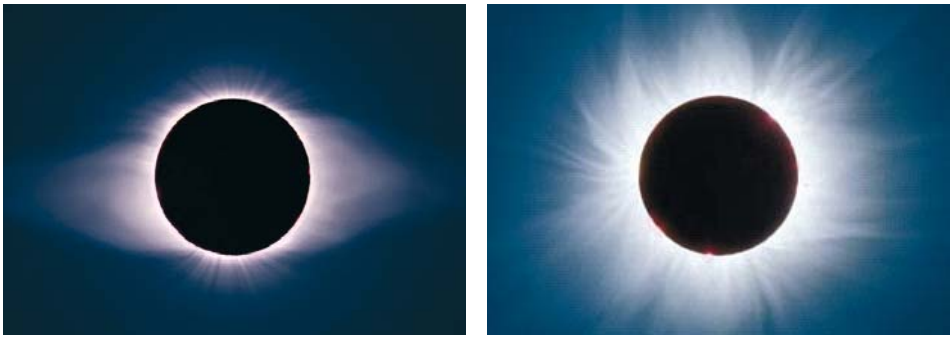


Figure 2. Solar-eclipse photographs of the solar corona. On the left, the corona at solar minimum on 24 October 1995, which shows the magnetically closed equatorial streamers and the large polar coronal holes with the characteristic polar plumes. On the right, the corona on 11 August 1999, close to solar 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.)

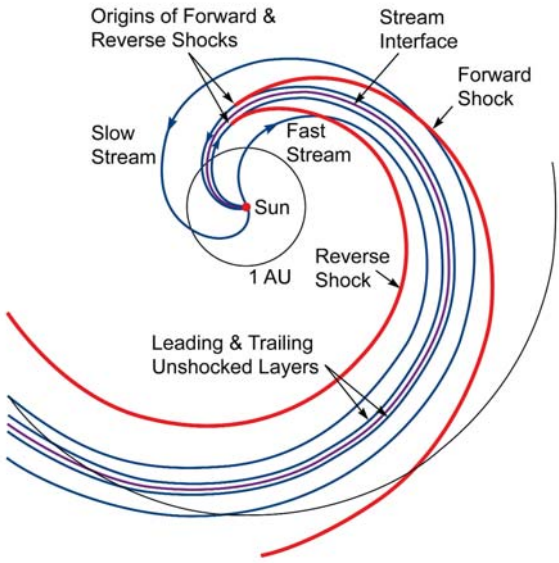


Figure 3. The equatorial projection of Corotating Interaction Regions, illustrating their main features (after Crooker *et al*., in Ref. 11). The shock fronts are in red, magnetic field lines in blue, and the contact surface between two plasmas in magenta.

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 $\pm 30^{o}$ 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 penetrat-

ed 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 ($\sim 10^{12}$ 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

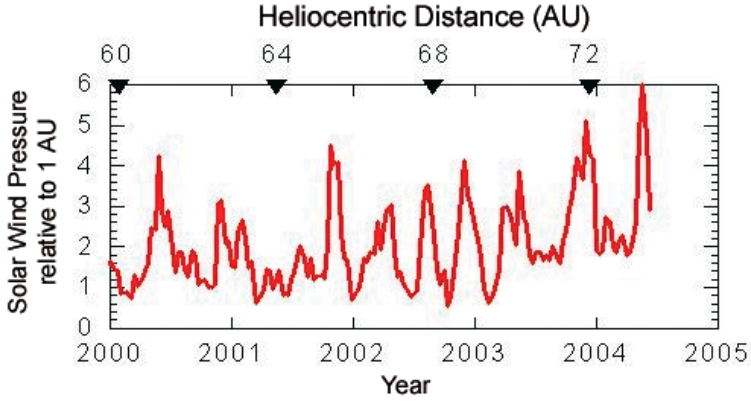


Figure 5. The variable pressure of the solar wind in the outer heliosphere, as measured by the Voyager 2 spacecraft in the years following the last solar maximum. (The pressure has been normalised to 1 AU.) The pressure pulses arise from the coalescing remnants of Coronal Mass Ejections and affect the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

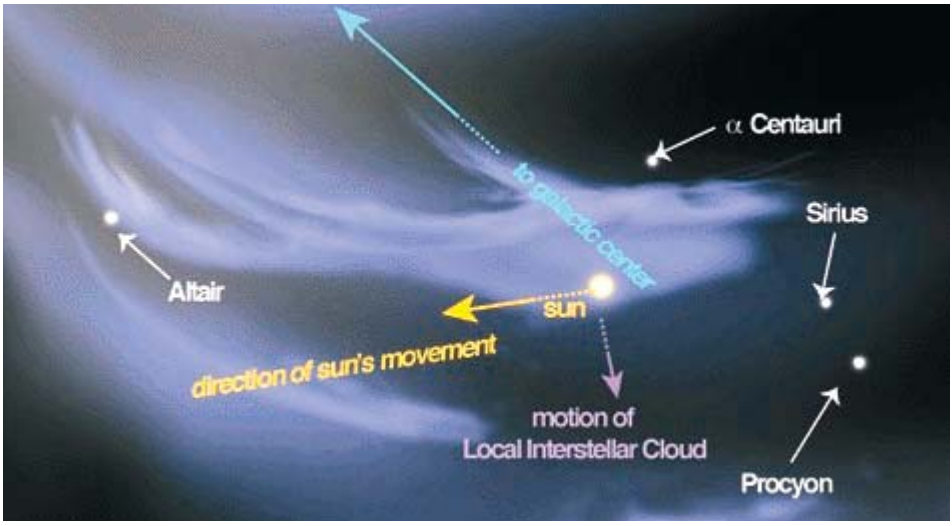


Figure 6. The structure of the Interstellar Medium in the vicinity of the Sun – within 10 light-years. The Sun is embedded into the partly ionized Local Interstellar Cloud (LIC). Directions of the Sun's motion, the LIC motion and towards the galactic centre are shown. (After P. Frisch, Ref. 16.)

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].

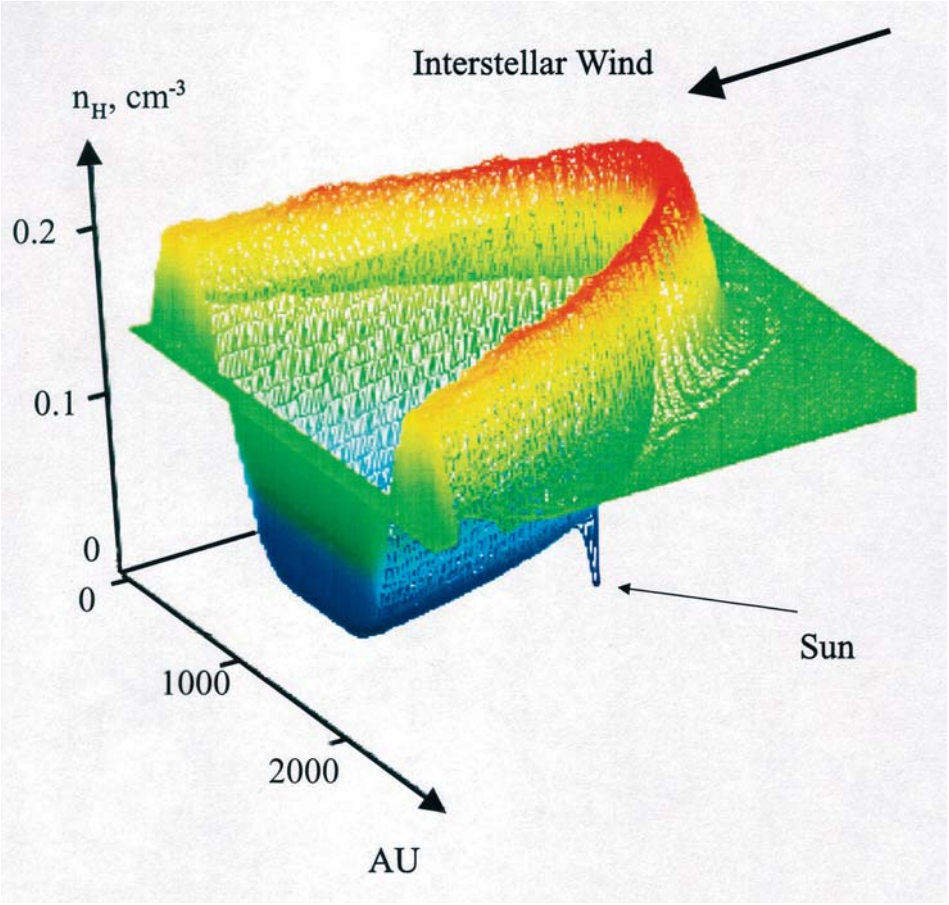


Figure 7. Hydrogen wall – an increase in the number density of interstellar H atoms around the heliopause of the Sun. The hydrogen wall is created due to charge exchange of primary H atoms with interstellar protons decelerated in the vicinity of the heliopause. (After Gruntman *et al*., Ref. 32.)

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 indi-

rect 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 $n_{p,LIC}$ = 0.11-0.12 cm^{-3} and $n_{H,LIC}$ ~ 0.22 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 ~0.1 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.

References

1. L.E. Davis, Jr., *Phys. Rev.*, **100**, 1440, 1955.
2. E.N. Parker, *Astrophys. J.*, **128**, 664, 1958.
3. The early results were reviewed by R. Lüst, in "Solar-Terrestrial Physics", J.W. King & W.S. Newman (Eds.), Academic Press, London, New York, 1, 1967; an overview has been given by M. Neugebauer & R. von Steiger, in "Century of Space Science", J. Bleeker, J. Geiss & M.C.E. Huber (Eds.), Kluwer Academic Publ., Dordrecht, p. 1115, 2001.
4. L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger & K-P. Wenzel, "Cosmic Rays in the Heliosphere", Space Science Series of ISSI Vol. 3, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **83**, 1998; J. W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger & R. Kallenbach (Eds.), "Cosmic Rays and Earth", Space Science Series of ISSI Vol. 10, Kluwer Academic Publ., Dordrecht, and *Space Sci. Rev.*, **93**, 2000.
5. R.G. Marsden (Ed.), "The Sun and the Heliosphere in Three Dimensions", D. Reidel, Dordrecht, 1986; J.R. Jokipii, , C.P. Sonett & S. Giampapa (Eds.), "Cosmic Winds and The Heliosphere", The University of Arizona Press, Tucson, 1997; A. Balogh & L.A. Fisk, in "Century of Space Science" (see in Ref. 5), p. 1141, 2001.
6. R. von Steiger, R. Lallement & M.A. Lee (Eds.), "The Heliosphere in the Local Interstellar Medium", Space Science Series of ISSI Vol. 1, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.*, **78,** 1996.
7. The results of the Helios 1 and 2 spacecraft that explored the innermost region of the heliosphere to study the connection between the solar wind and its origins on the Sun were reviewed by R. Schwenn and E. Marsch (Eds.), "Physics of the Inner Heliosphere", Vols. 1 and 2, Springer-Verlag, Berlin-Heidelberg, 1990, 1991; in particular by R. Schwenn, Vol. 1, p. 99, 1990.
8. J. Geiss, *Space Sci. Rev.*, **33**, 201, 1982; R. von Steiger, J. Geiss & G. Gloeckler, in "Cosmic Winds and the Heliosphere" (see Ref. 5), p. 581, 1997.
9. E.J. Smith, in "Cosmic Winds and the Heliosphere" (see Ref. 5), p. 425, 1997.
10. The results of the Ulysses spacecraft that has explored the heliosphere over the poles of the Sun have been reviewed in R.G. Marsden (Ed.), "The high latitude heliosphere", Kluwer Academic Publ., Dordrecht, 1995, and in A. Balogh, R.G. Marsden & E.J. Smith (Eds.), "The Heliosphere near Solar Minimum: The Ulysses Perspective", Springer-Praxis Series Astrophys. Astron., Springer, Berlin, 2000.
11. A. Balogh, J.T. Gosling, J. R. Jokipii, R. Kallenbach & H. Kunow (Eds.), "Corotating Interaction Regions", Space Science Series of ISSI Vol. 7, Kluwer Academic Publishers, Dordrecht, and *Space Sci. Rev.* **89**, 1999; in particular, see N.U. Cooker *et al.*, p. 179, 1999.
12. M.A. Lee, in "Cosmic Winds and the Heliosphere" (see Ref. 5), p. 857, 1997.
13. E.J. Smith, *J. Geophys. Res.,* **106**, 15819, 2001.
14. R.G. Marsden (Ed.), "The 3-D Heliosphere at Solar Maximum", Kluwer Academic Publ., Dordrecht, 2001.
15. F.B. McDonald & L.F. Burlaga, in "Cosmic Winds and the Heliosphere" (as in Ref. 5), p. 389,

1997; J. D. Richardson, K.I. Paularena, C. Wang & L.F. Burlaga, *J. Geophys. Res.*, **107**, 1041, 2002.

16. P. Frisch, *American Scientist*, **88**, 52, 2000.
17. M. Witte, M. Banaszkiewicz & H. Rosenbauer, in Ref. 6, **78**, 289, 1996; M. Witte, *Astron. Astrophys.*, **426**, 835, 2004; see, also, special section of *Astron. Astrophys.* , **426**.
18. R. Lallement, in Ref. 6, **78**, 361, 1996.
19. E.N. Parker, *Astrophys. J.,* **134**, 20, 1961; V.B. Baranov, K.V. Krasnobaev & A.G. Kulikovksy, *Sov. Phys. Dokl.*, **15**, 791, 1971.
20. E.C. Stone, *Science,* **293**, 55, 2001.
21. V.V. Izmodenov, in "The Sun and the Heliosphere as an Integrated System", G. Poletto & S. Suess (Eds.), Kluwer Academic Publ., Dordrecht, 2004; G. Zank, *Space Sci. Rev.*, **89**, 413, 1999.
22. V.B. Baranov & Y.G. Malama, *J. Geophys. Res.,* **98**, 15 157, 1993.
23. J. Linsky & B. Wood, *Astrophys. J.,* **463**, 254, 1996.
24. V.V. Izmodenov, R. Lallement & Y.G. Malama, *Astrophys.*, **342**, L13, 1999; B.E. Wood, H.-R. Müller, G. P. Zank, V.V. Izmodenov & J.L. Linsky, *Adv. Space Res.,* **34**, 66, 2004; B.E. Wood, H.-R. Müller & G.P. Zank, *Astrophys. J.,* **542**, 493, 2000; V.V. Izmodenov, B.E. Wood & R. Lallement, *J. Geophys. Res.,* **107**, SSH 13-1, CiteID 1308, DOI 10.1029/2002JA009394, 2002. For the most recent review, see A. Balogh, in "Living Reviews in Solar Physics" **1**, 2, URL: http://www.livingreviews.org/lrsp-2004-2, 2004.
25. G. Gloeckler & J. Geiss, *Adv. Space Res.,* **34**, 53, 2004; G. Gloeckler & J. Geiss, *Space Sci. Rev.,* **97**, 169, 2001.
26. V. Izmodenov, Y. Malama, G. Gloeckler & J. Geiss, *Astron. Astrophys.*, **414**, L29, 2004.
27. D.A. Gurnett, W. Kurth, S. Allendorf & R. Poynter, *Science,* **262**, 199, 1993; D.A. Gurnett & W. Kurth, in Ref. 6, p. 53, 1996.
28. V. Izmodenov, Y. Malama, G. Gloeckler & J. Geiss, *Astrophys. J.,* **594**, L59, 2003; V. Izmodenov, G. Gloeckler & Y. Malama, *Geophys. Res. Lett.,* **30**, 3-1, 2003.
29. S.M. Krimigis *et al.*, *Nature,* **426**, 45, 2003.
30. F.B. McDonald *et al.*, *Nature,* **426**, 48, 2003.
31. B. Wolf, D. Koester & R. Lallement, *Astron. Astrophys.*, **346**, 969, 1999.
32. M. Gruntman, E.C. Roelof & D.G. Mitchell, *J. Geophys. Res.,* **106**, 15767, 2001.
33. Our understanding of the heliosphere has been immensely increased by the interaction of scientists who met repeatedly within the creative environment of ISSI for discussions that led to five volumes serving as reference works on this topic. V.I. was supported in part by ISSI, INTAS grant 2001-0270, and RFBR grant 04-02-16559.

Acceleration in the Heliosphere

E. Möbius[a] and R. Kallenbach[b]

[a]*Space Science Center and Department of Physics, University of New Hampshire, Durham, USA*
[b]*International Space Science Institute, Bern, Switzerland*

Introduction

Since the discovery by Victor Hess on a balloon flight in 1912 that the intensity of energetic radiation increases with altitude, it has been known that cosmic rays are an important ingredient of the space environment. The space era has opened this important phenomenon to detailed study. The extension of optical astronomy into the radio regime in the 1940's, and later with space-based telescopes into the more energetic X-ray and γ-ray regimes, has provided ample evidence for humongous and violent objects in the Universe[1] that generate floods of extremely energetic particles, which ultimately form the cosmic-ray population in our galactic neighbourhood. Our living environment on Earth can be affected by this potentially dangerous radiation and its mutation-driving impacts, but it is very effectively protected by a three-layer shield system: (1) the heliosphere with the solar wind and its embedded magnetic field; (2) Earth's magnetic field; (3) Earth's atmosphere. Satellites and space probes have gathered information about the variation in the radiation intensity and spectra in response to changes in the interplanetary medium and solar activity[2,3]. They have also provided us with in-situ observations of sources, acceleration, and transport of energetic particles generated in our immediate neighbourhood, in and around Earth's magnetosphere, at the Sun, and throughout the heliosphere. Our ability to model and scale acceleration processes enables us to understand not only our own environment, but also that of distant particle accelerators, which cannot be studied in-situ. As shown in Figure 1, the average fluence spectra in interplanetary space[4] extend from the solar wind up to several 100 MeV/atomic mass unit [amu] with a power law for most of the range similar to cosmic rays. Shock waves, or structures where the solar wind flow is abruptly decelerated from super- to sub-sonic speeds, have been identified as powerful particle accelerators in the heliosphere. They serve as a model for supernova blast waves, which accelerate galactic cosmic rays. With ever-increasing sophistication in spacecraft instrumentation, we make good use of the laboratory on our front doorstep by simultaneously observing source and energetic-particle populations as well as magnetic and electric fields near shocks. Yet the first step from the bulk plasma into the accelerated distribution (marked with ? in Fig. 1) is still very much under debate.

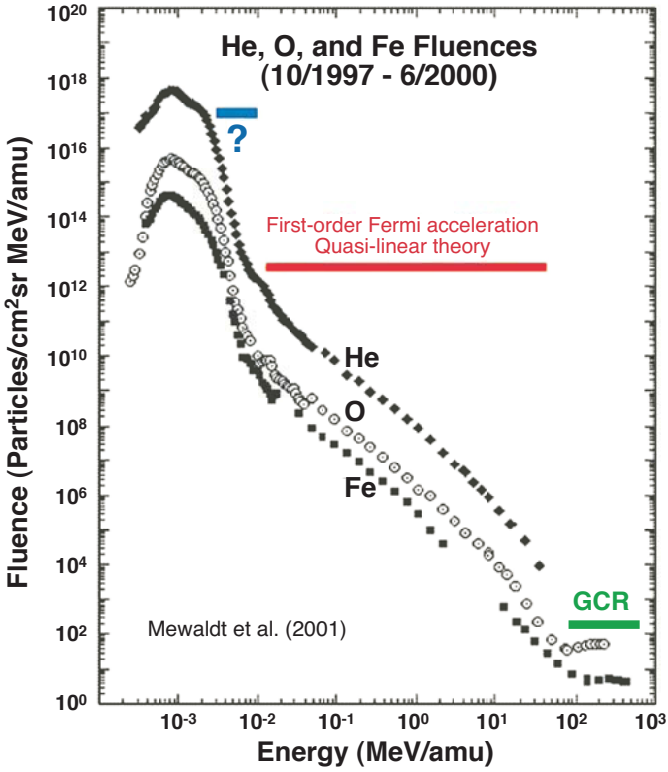


Figure 1. Total fluence spectra during the rise to the 2000 solar maximum. Solar and heliospheric energetic particles reach to galactic cosmic rays (adapted from Ref. 4).

As this brief introduction shows, particle acceleration is a genuinely interdisciplinary topic, which benefits greatly from collaborations across the dividing lines between disciplines. The International Space Science Institute (ISSI) is providing an excellent forum for such discussions with workshops and scientific teams composed of scientists with a variety of backgrounds. For cosmic rays, ISSI's impact has already been demonstrated through several cross-discipline books from such activities on the topic.

Taking a more local view, it also happens that ISSI in its first ten years of existence has hosted symposia and working teams that have compiled comprehensive summaries, provided a critical evaluation of our current understanding, and focused on the open questions for two of the key heliospheric phenomena, which are generators of energetic particles, co-rotating interaction regions (CIRs)[5] and coronal mass ejections (CMEs)[6]. CIRs are the dominant structures and generators of energetic particles in interplanetary space close to and during the solar

activity minimum. The combination of a rotating Sun and an orderly emission pattern of fast and slow solar wind leads to the periodic overtaking of slow by fast wind, with the formation of compression regions and shocks. CMEs constitute blast waves from eruptions in the solar corona and thus are dominant during solar maximum. Both types of shocks, those of CIRs and of CMEs, in fact also serve as model cases for the heliospheric termination shock, the structure at presumably about 100 astronomical units (AU) where the solar wind undergoes its final transition from supersonic to subsonic flow. The study of the heliospheric termination shock has become an ISSI research area specially funded by INTAS in Brussels, supporting the collaboration between theoreticians of the countries belonging to the former Soviet Union and scientists from the European Union or Switzerland.

The basic ISSI consensus on particle acceleration in the heliosphere formulated in the CIR book is that: (1) both solar wind and interstellar gas penetrating the heliosphere contribute as important sources to the energetic particles; (2) shocks themselves provide a means for efficient acceleration; and (3) particles are required to attain a minimum energy to be injected into shock acceleration, termed the "injection threshold", which is substantially higher than the thermal energy of solar wind particles, and may at times be higher than the maximum energy of pickup ions from neutral gas. We have organized this paper such that we devote a section to each of these three issues following this introduction, ending with the injection problem, i.e. the question of how particles make the transition from being a member of the bulk source to entering the energetic-particle population. This has been thoroughly debated in the CIR book, but basically left unsolved. We will close with a brief summary of where we stand right now, what the open issues are, and how we can apply the local results to more distant phenomena in the Universe.

Locations and Sources for Particle Acceleration

Let us start with the question of how we can identify where particles are actually accelerated. In the following we will repeatedly refer to Figure 2, which shows a simplified view of the heliosphere with rather schematic shapes for a CME and a CIR. The figure also contains the simplified spatial distribution of the most important source populations for particle acceleration, as they have been identified observationally.

Particle acceleration, where?

In the 1960's so-called "energetic storm particles" (ESP) were observed in interplanetary space with maximum intensity when the disturbance passed the space-

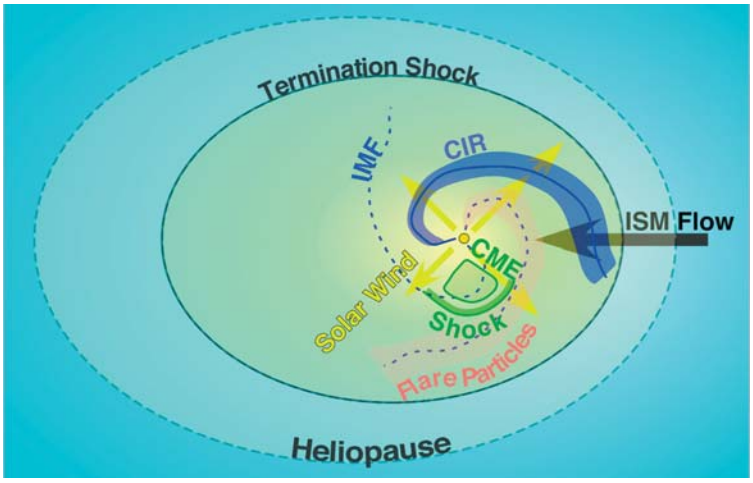


Figure 2. Schematic view of the heliosphere with interstellar medium (ISM) flow, solar wind and interplanetary magnetic field (IMF). Acceleration structures, such as a co-rotating interaction region (CIR) and a coronal mass ejection (CME), are also sketched.

craft after a major eruption on the Sun. The peak of the particle flux could be associated with the passage of interplanetary shocks that signal the arrival of a CME (shown in green in Fig. 2), and the co-location therefore suggested a causal connection with local acceleration.

Spacecraft also detected energetic particles that appeared to be connected with recurrent activity regions on the Sun. Pioneer 10 and 11 found that the fluxes increased with distance from the Sun, rather than decreasing, as is expected for a solar origin[7]. They peaked at 3-4 AU and then declined sharply. Evidence for local acceleration of these particles was the close association of the peak fluxes with the leading and trailing edge of the compression ahead of high-speed solar-wind streams[8], indicated in blue in Figure 2. CIRs are the dominant structures in the inner heliosphere during the decline and minimum of the 11-year solar cycle when the solar magnetic field is a fairly stable dipole, which is tilted relative to the Earth's orbit. Let us use a lawn sprinkler as an analogy: While the sprinkler ejects water, the rotating Sun emits solar wind. It emits fast wind from the strongly tilted northern and southern caps, and slow wind from the equator. Over the course of one rotation, a succession of slow and fast wind is seen in the ecliptic. As a consequence, the fast wind runs into slow wind that was emitted earlier, and compression regions form at the interface, the CIRs. Since even the speed difference is supersonic, shocks form on both sides of the compressions as the "sprinkler" spiral is wound up at larger distances.

Source populations for acceleration

For CMEs and CIRs, acceleration occurs in regions where plasmas with different speed converge. But what is the material that is accelerated? Composition measurements can identify the source populations. With the accelerating structures, Figure 2 also shows typical spatial distributions of major particle sources in the heliosphere. Streaming away from the Sun, the corona and solar wind are key plasma sources, with an elemental composition similar to the Sun and highly ionized because of the high temperature. Comprehensive surveys of solar energetic particles have established that the average composition of strong particle events, which are associated with CMEs, closely resembles that of the corona and solar wind[9]. These observations led to the paradigm that this energetic particle population is mostly accelerated out of the corona and/or the solar wind. The energetic ion composition in CIRs was also found to be similar to that of solar energetic particles and solar wind, but with some noticeable differences[8], in particular for He and C.

The advent of high resolution and collection power composition instruments for both bulk distributions and energetic-particle populations on spacecraft, such as ACE, SAMPEX, SOHO, Ulysses and Wind, has made it possible to establish the composition patterns, energy spectra, and spatial and temporal variations of sources and energetic particles in detail. Substantial deviations from the previously obtained averages and huge variations, which were mostly hidden in the past because of lack in resolution and counting statistics, have now called for additional energetic-particle sources. Substantial ^{3}He enhancements were found in CME events[10], which cannot be explained by selective acceleration out of the solar wind, as its ^{3}He abundance is very low[11]. However, ^{3}He and heavy ion enrichment have long been known as the hallmark of impulsive flares[12]. In fact, the active Sun peppers the heliosphere with impulsive flare material, which starts its journey from compact flare sites narrowly confined to magnetic flux tubes[13], as indicated in Figure 2. Over time and through frequency of occurrence, these particles presumably spread out into another important source distribution. A new comprehensive survey[14] provides clear evidence for the presence of such material in the background population far ahead of the shocks, where acceleration of ^{3}He and heavy ions is observed. This key observation suggests that preexisting energetic and suprathermal particle populations, such as a remnant mixture from impulsive flares and previous CMEs, may be a substantial feeder into particle acceleration, thus indicating an efficient recycling of energetic particles.

A second channel for composition measurements is the distinction of ionic charge states, which are the hallmark of the ionization environment of the source material. Interstellar gas, which streams into the heliosphere as a wind, is a

major player with increasing distance from the Sun (Fig. 2)[15]. These particles appear as pickup ions in the solar-wind plasma and as freshly ionized particles they are mostly singly charged. The, at first, puzzling observation of a substantial He^+ fraction of interplanetary energetic particles[16] in the multiply charged solar wind has found its natural explanation after the detection of interstellar pickup He^+.[17] The identification of pickup ions as the major contributor to energetic He in a CIR at 4.5 AU[18] has led to the suggestion that pickup ions are generally an important source for efficient acceleration at interplanetary shocks[19]. With an average of ≈25%, the He^+ abundance is much reduced at 1 AU compared with its dominance at 4 - 5 AU. But, as shown in Figure 3, the observed increase in He^+/He^{2+} with time elapsed from the start of each CIR reflects the increasing importance of interstellar He as source material with distance from the Sun, as the spacecraft is magnetically connected to the CIR at larger and larger distances[20].

Interstellar pickup ions are also an important contributor at interplanetary travelling shocks. The overwhelming majority of the energetic He^+ at these shocks cannot stem from cold prominence material, even for CMEs that show an overabundance of He^+ in the solar wind bulk flow, as the substantial enhancements in the energetic population occur outside the CME cloud[21,22]. A survey of He^+ and He^{2+} demonstrates that, with $He^+/He^{2+} \approx 0.06$, He^+ is in fact the third most abundant ion species of the energetic-particle population in the inner heliosphere[22].

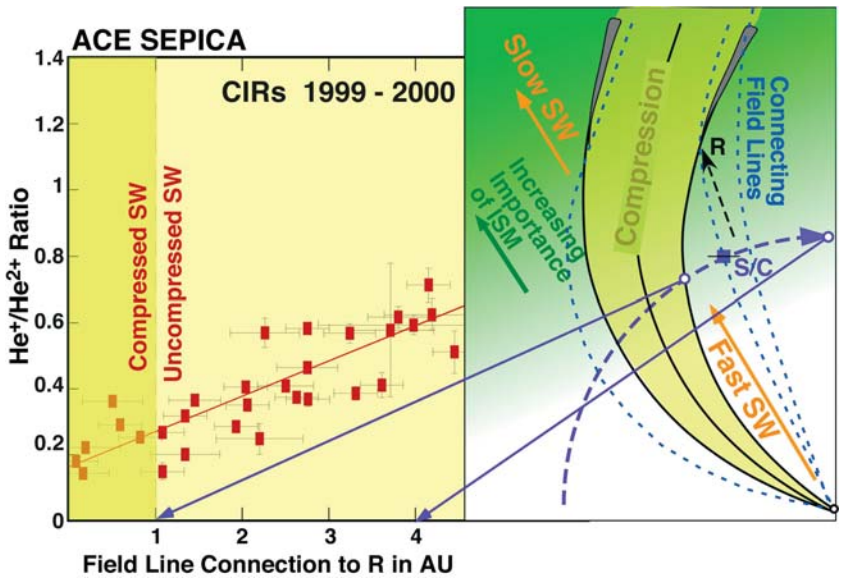


Figure 3. Temporal variation of the He^+/He^{2+} ratio shown as a dependence on the distance of a CIR from the Sun to which the field line at the S/C connects. The rise indicates the increasing importance of interstellar He^+ pickup ions as source.

The relative abundance of He^{+} over He^{2+} is substantially (by 50-200 times) enhanced in the energetic population over the source population, i.e. the abundance ratio of pickup ions over solar wind He^{2+}.

Based on the observed increased injection efficiency of pickup ions into acceleration, it was argued that the "inner source" of pickup ions may also contribute substantially to the energetic particle population in CIRs[19]. The inner source is thought to be solar wind that is implanted in interplanetary dust grains and then re-emitted as neutral atoms[23], or solar wind that is neutralized by penetrating very small dust particles[24], thus leading to a somewhat modified solar-wind composition. Although revealing He^{+} and Ne^{+}, detailed studies of the energetic CIR population at ≈1 MeV/amu did not find any evidence for the expected singly charged C, O, or Mg, and the observed charge state distributions clearly resemble those of the adjacent solar wind[25,26].

Acceleration Processes

In a nutshell, particle distributions that feature a high-energy extension from the bulk flow, such as remnant energetic particles and pickup ions, appear to be efficiently accelerated further. To discuss the acceleration processes, it is therefore justified to start with particles that already have a wide velocity distribution compared with the relatively cold bulk plasma. We will leave the question of how particles make the transition into this suprathermal distribution until the last chapter. Under this premise, the particles are rather mobile in the frame of the bulk plasma. Their motion can be altered by obstacles in the bulk flow, which we term "scattering centres", and they have the ability to cross discontinuities between plasmas of different velocity freely. It is these qualities that make the particles susceptible to acceleration.

First-order Fermi acceleration

First-order Fermi acceleration is an unavoidable consequence in the presence of any jump in plasma bulk velocity V_b, if particles cross such a jump multiple times and undergo multiple collisions[27]. A pictorial view of this process is shown in Figure 4 for a parallel shock, i.e. the magnetic field is parallel to the shock normal. Plasma with embedded magnetic field and fluctuations is approaching from the left with high speed (indicated by the blue arrows). It then slows down either abruptly at a shock (upper half) or gradually in a compression region (lower half). The transition in speed V_b is indicated in the two graphs above and below the main figure and by the shorter red arrows. Grey shading indicates the related compression in density. Particles that move already relative to the bulk flow may get turned around by magnetic fluctuations (indicated as scattering

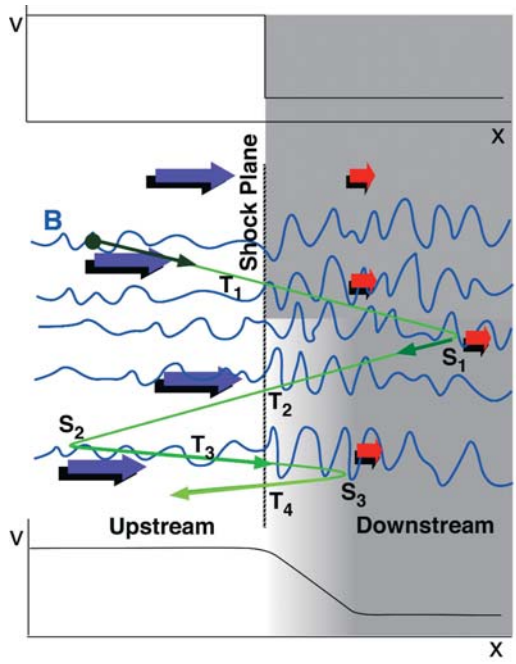


Figure 4. Schematic representation of first-order Fermi acceleration for plasma flow with embedded magnetic-field fluctuations. Ions with motion relative to the flow gain energy through a combination of scattering ($\mathbf{S_i}$) upstream and downstream and transitions ($\mathbf{T_i}$) of the speed change, independent of the smoothness of the change. Speed is shown as line plots and the related density change as grey shading. Top: Strong shock; Bottom: Gradual change.

events $\mathbf{S_i}$), which may be described like an elastic collision with a massive partner, as the magnetic field is strongly coupled to the bulk flow. After one complete cycle of scattering downstream ($\mathbf{S_1}$) and upstream ($\mathbf{S_2}$) with dual transition ($\mathbf{T_1}$, $\mathbf{T_2}$) of the velocity gradient, the original momentum mv_i has increased to $mv_f = mv_i + 2m(V_u - V_d)$, where V_u and V_d are the upstream and downstream values of the flow speed. Consecutive cycles add momentum in equal increments.

Figure 5 shows the process in velocity space in terms of the classical Fermi acceleration "picture"[28]. The velocity space trajectories, shown here in the reference frame of the shock, describe the velocity evolution for the sample ion in Figure 4. It starts out somewhat faster than the solar wind and thus is already suprathermal. Scattering at magnetic field fluctuations occurs approximately under conservation of energy in the plasma frame, for $\mathbf{S_1}$ the slow or shocked flow (red dot). The square of the total ion speed $v^2 = v_{||}^2 + v_{\perp}^2$ is conserved in the downstream flow, where $v_{||}$ and $v_{\perp}$ are the velocity components parallel and perpendicular to the magnetic field. As a consequence, the trajectory is a circular arc centred on the downstream flow velocity. Once the ion's parallel velocity is negative in the shock frame, it can cross the shock and reach the upstream plasma ($\mathbf{T_2}$). Here the ion is scattered on a sphere around the blue dot ($\mathbf{S_2}$) so that it reaches a positive speed in the shock frame and comes back to the downstream plasma ($\mathbf{T_3}$). Apparently, the ion gains energy with each shock crossing, as evidenced by the increasing radii of the arcs.

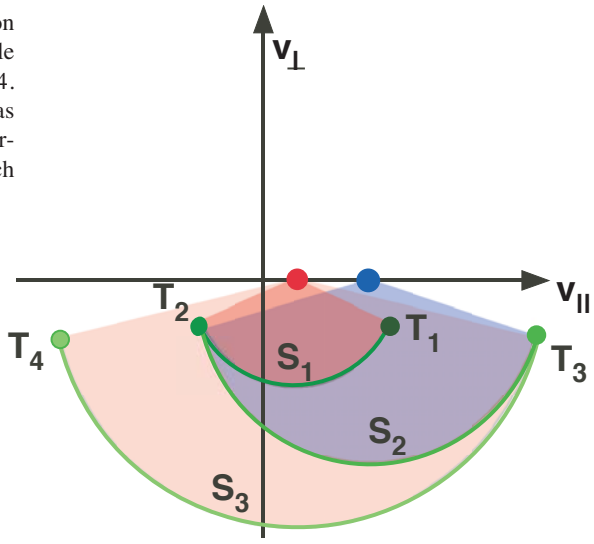


Figure 5. Schematic view of the ion velocity evolution for the sample trajectory shown in Figure 4. Increases in energy show up as increases in the radii of the scattering arcs in velocity space after each transition of the discontinuity.

As seen in Figure 4, also a smooth transition with the same overall speed change can produce the same increment of acceleration, if the ion is scattered after passing the full gradient[29]. The average momentum or energy gain in a smooth transition is less efficient though, because not all particles traverse the full speed change. The ratio of the mean free scattering length and the scale length of the gradient determines the overall acceleration efficiency. This extension of shock acceleration appears to be a natural explanation for the, at first, puzzling observation that CIRs can also accelerate particles at 1 AU, without fully developed shocks[30].

Second-order Fermi acceleration

In the presence of magnetic-field fluctuations, which provide the means of cross-shock transport, ions are also scattered back and forth between plasma waves with different velocities in the plasma frame. This leads to second-order Fermi acceleration (or acceleration of ions in turbulent plasma waves), which is also a natural consequence of particle transport[31]. If we consider for a moment two waves that propagate in the same direction, but at fast and slow speeds, respectively, we can associate these speeds with the blue and red dots in Figure 5 and treat particle acceleration accordingly. However, waves may propagate in random directions, even in opposite directions. Depending on whether a particle is scattered at a wave in a head-on or an overtaking collision determines the energy change. Particles gain momentum (and energy) in head-on scattering, but lose it in overtaking scattering, making second-order Fermi acceleration a stochastic process. However, the scattering rate is also proportional to the relative velocity of particle and scatterer. Thus head-on scattering is more frequent, leading to a net momentum and energy gain.

Compared with first-order Fermi acceleration, whose strength derives from the change in the bulk velocity (V_b) at a discontinuity, second-order Fermi acceleration depends on the random speed of scattering centres in the plasma, or the speed of the prevalent waves (Alfvén or sound speed). Therefore, first-order Fermi acceleration is usually stronger than its second-order cousin in supersonic or Alfvénic flows with discontinuities by a factor of M^2 in terms of energy gain, where M is the Mach number. However, in the absence of discontinuities second-order Fermi acceleration may become the dominant process. Relating to the pictorial view in Figure 5, the second-order process also relies on momentum change during scattering, but now in the wave frame. The increment in momentum change is determined by the wave speed relative to the plasma. Effective scattering and randomization of momentum, and thus acceleration, is achieved if more than one wave mode and/or propagation direction is present.

Acceleration efficiency and escape

Let us now ask how efficient acceleration can be, and do this while concentrating on first-order Fermi acceleration, as the arguments follow in a similar way for the second-order process. Fermi acceleration requires elastic scattering of an ion in both the upstream and downstream plasmas. This is the only way in which the plasma can "communicate" the speed difference between upstream and downstream to the ion to energize it during each round trip across the shock. There are two limits beyond which first-order Fermi acceleration may not work. Firstly, the ions are scattered too strongly in the downstream plasma, will simply be convected away, and will never return to the upstream plasma. Secondly, the ions are scattered too weakly either in the upstream or in downstream plasma. In that case they will not be energized through "ping-pong collisions" with the plasmas of different speeds in reasonable time.

To be efficiently accelerated, particles have to cross the shock or discontinuity multiple times. For a parallel shock the ions are rather mobile, as they cross in the direction of the magnetic field lines. Therefore, they finish many round trips during the lifetime of the shock. An interplanetary shock can accelerate ions that are observed near Earth during the time it travels from the Sun to Earth, typically a few days. An estimate of the average time for an ion to revert to its "parallel" speed along the mean interplanetary magnetic field is roughly 1 to 10 minutes. This is sufficiently short for the ions to undergo efficient acceleration, in agreement with observations.

If first-order Fermi acceleration were to continue unabated at a planar shock, a power law spectrum in energy E for the differential flux $j(E) = j_o \cdot (E/E_o)^{-\gamma}$ results[32], whose spectral index γ is related to the shock compression ratio r as $\gamma = (r+2)/2(r-1)$. In any realistic situation though, the energy spectra are estab-

lished in a balance between acceleration and escape from the finite acceleration region. As diffusive transport of particles usually speeds up as a function of the magnetic rigidity R, which scales in the non-relativistic limit as $R \sim v \cdot A/Q$, with atomic mass A and ionic charge Q, the spectra turn over into an exponential behaviour at high energies. This turnover occurs at lower energies for particles with higher A/Q values, as observed for heavier species[33].

Quasi-perpendicular shocks and drift acceleration

So far our models apply to a quasi-parallel shock (or compression). However, often the field is oriented oblique or even nearly perpendicular to the shock normal, which is typical for CIRs beyond 1 AU. This is a quasi-perpendicular shock, which allows also for another acceleration process, i.e. shock drift acceleration.

This process is based on the convective electric field in magnetized plasma, moving with velocity $\boldsymbol{V}_b$ perpendicular to the magnetic field $\boldsymbol{B}_0$. Protons take a right turn around the magnetic field $\boldsymbol{B}_0$ and electrons a left turn due to the Lorentz force. The charges separate, and an electric field $\boldsymbol{E}$ builds up ($E = V_b B_0$), which is perpendicular to $\boldsymbol{V}_b$ and $\boldsymbol{B}_0$. Ions with velocity $\boldsymbol{V}_b$ do not feel this field at all. With a different velocity, they feel it on their gyro-orbits alternately against and with their motion. In both cases there is no net effect on the ion energy. The situation changes at a shock. An ion that gyrates half of its orbit upstream and half downstream experiences two different field strengths, usually larger upstream. This leads to a net acceleration without scattering, while the ion drifts along the shock surface for a short distance. After this encounter, the ion usually is swept downstream with the flow, and it needs scattering to allow multiple shock encounters for further energy gain.

Multiple scattering and thus Fermi acceleration of both kinds work best when the ions can move freely between upstream and downstream, or between randomly moving scattering centres, without the need to cross magnetic field lines. At a perpendicular shock all ions are effectively swept into the shock (or compression) due to their gyro-motion, but return into the upstream region is severely hampered. In order to move back upstream the ions must now be scattered across field lines. Only a rather small fraction of the ions, which scales with the fraction $\eta = \delta B^2/B_0^2$ of magnetic energy present as fluctuations, will cross the shock again after each cycle. In a simplified way, we can express the capability to return upstream in terms of a diffusion speed across field lines $v_D = \eta v$, where v is the actual ion speed. For typically small η values, the ions need a high speed. Of course, ions may also cross oblique shocks along field lines, but the escape speed required is increased over the downstream flow speed by $1/\cos\Theta$, where Θ is the angle between the magnetic field and the shock normal. For nearly per-

pendicular shocks, this speed is enormous. As one can imagine, a high enough speed is hardest to attain at the very beginning of the acceleration.

Injection Problem

After exploring the efficiency and limits of acceleration and learning that often shock acceleration is most effective, we need to turn our attention to the start of the process. In fact there is an obstacle to shock acceleration, the injection problem, and thus far it is not solved. In the ISSI book on CIRs, the injection problem is spelled out as the major unsolved problem of shock acceleration[5]. All possible injection mechanisms are summarized, with special emphasis on perpendicular shocks, as is relevant for CIRs, but no convincing model that overcomes all problems is presented[31]. This situation has not changed since.

It is not only the scattering rate that decides on the success of the first-order Fermi process, but also the ion speed. Ion speed and scattering rate combine into the mean free path, which directly relates to ion mobility. If the mobility of an ion in the downstream plasma is sufficiently high, it has some chance of being returned to the upstream plasma against the downstream plasma flow. The mean free path usually increases with ion speed. However, this is not the only necessary condition for shock acceleration to start. Upon approaching the shock from downstream after their first scattering or from the tail of the heated distribution, the ions need a minimum speed to outrun the shock. This minimum speed with which an ion has a reasonable chance of being returned to the upstream plasma is called the *injection threshold*. It is very different for the two types of shocks, parallel and perpendicular, as it depends strongly on the angle Θ between magnetic field and shock normal.

Injection at quasi-parallel shocks

We start with the type of shock where ions have little problem to undergo a first-order Fermi acceleration. At quasi-parallel shocks, the magnetic field is almost parallel to the shock normal, i.e. perpendicular to the surface of the shock front. To initiate the multiple shock crossing an ion needs to be able to return upstream, or its minimum escape speed from the downstream region must exceed the downstream flow speed.

The sample ion depicted in Figure 4 does not seem to have difficulties to return upstream after the first scattering in the downstream region, as it is deliberately chosen as a suprathermal ion. Even when solar wind ions reach the downstream region without changing velocity at the shock, they would constitute a suprathermal population to which the above-mentioned scattering theory applies.

However, this picture is incomplete because it omits the shock potential, which slows down the vast majority of the solar wind ions (blue dot in Fig. 5) to the downstream bulk distribution (red dot). No scattering would ensue in the downstream region, and the injection problem persists for solar-wind ions that are close to the bulk distribution.

To discuss the injection for different particle populations, Figure 6 shows a compilation of various distributions in interplanetary space in a two-dimensional projection and as a one-dimensional cut. All ions that are close to the solar-wind bulk velocity, i.e. almost all of the genuine solar-wind ions (yellow) will mostly be decelerated so that they do not experience any scattering. However, there are several much more extended distributions, for which the described injection scenario works particularly well. The first are interstellar pickup ions (dark blue), whose velocity distribution is approximately described by a sphere in velocity space, centred on the solar wind, with a radius equal to the solar-wind speed. Therefore, these ions are already "suprathermal," and they have the "pole position" for acceleration. Ions that have already experienced prior acceleration, as indicated by the light blue distribution in Figure 6, have a similar advantage. This explains why both interstellar pickup ions and remnant energetic ions are found with substantially increased abundance over solar-wind ions in CIRs and at travelling shocks. Inner source pickup ions could also be viewed as suprathermal, as they are separate from the solar-wind bulk. However, they are generally slower than the solar wind. Since they do not appear efficiently accelerated at least at 1 AU, perhaps mostly particles that are already faster than the solar wind are preferentially injected.

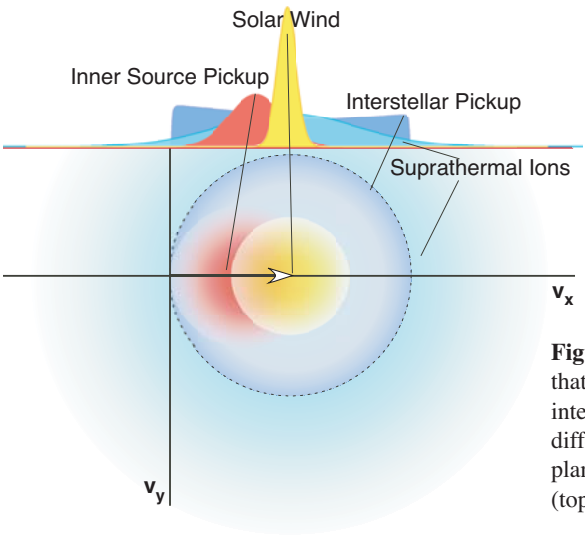


Figure 6. Velocity distributions of ions that can be injected into acceleration in interplanetary space are indicated with different colour shading in the $v_x - v_y$ plane (bottom) and as a cut in the v_x axis (top).

Injection at quasi-perpendicular shocks

Injection at a perpendicular shock is much more problematic. Let us follow the simplified view that first-order Fermi acceleration at a perpendicular shock operates for a diffusion speed ηv, which is larger than the downstream bulk speed V_d in the shock frame. For typical values of $\eta \approx 0.01$, the ion speed must be of order 5000 km/s, even if the downstream bulk speed is rather low, i.e. 50 km/s. 5000 km/s (equivalent to energies of >125 keV/amu) obviously is much higher than a typical solar-wind speed of 400 –700 km/s, and it is not clear yet how ions reach the injection speed, exemplifying the "injection problem."

The standard model of the perpendicular mean free path may be oversimplified[34], as magnetic field line "braiding" could increase the scattering efficiency by a factor 2 to 5. This would reduce the injection threshold to ≈ 25 – 60 keV/amu at CIRs, and numerical simulations revealed that at least a fraction of the interstellar pickup ions are fed into the first-order Fermi process at a perpendicular shock, but almost none of the thermal solar-wind ions. This is consistent with the observations that pickup ions are preferentially injected into acceleration at CIRs[19]. However, energetic particles with solar-wind composition are also observed at CIRs, and the assumption of increased perpendicular mean free paths has not yet been verified experimentally.

One interesting process is that of shock surfing[35,31], a special form of shock drift acceleration. In Figure 7 a pickup ion reaches the shock with very low speed, is reflected by the electric cross-shock potential, and then experiences the upstream convective electric field during a fraction of its gyro-period until it is reflected again[36]. Ultimately, such an ion reaches a speed that enables it to cross the shock, which often exceeds the injection threshold. However, the process is very idealized and only works for a shock that is perpendicular and planar over a large spatial scale and whose width is smaller than the ion gyro-radius. For typical shock widths observed in interplanetary space, the ions may reach a few 10 keV/amu, barely enough to reach the injection threshold. Again, this process works much better for pickup ions than for solar-wind ions because pickup ions populate velocity space below the bulk speed and a fraction always coincides with the shock velocity.

Alternatively, pre-acceleration by second-order Fermi acceleration has been suggested[31], which again prefers suprathermal, or typically super-Alfvénic, ion populations. Super-Alfvénic means that - in the bulk plasma frame - ions are faster than the Alfvén speed (the speed with which magnetic disturbances propagate in the solar-wind plasma regardless of frequency, $V_A \approx 50$ km/s). This also discriminates strongly against thermal solar wind and still leaves us with the question of

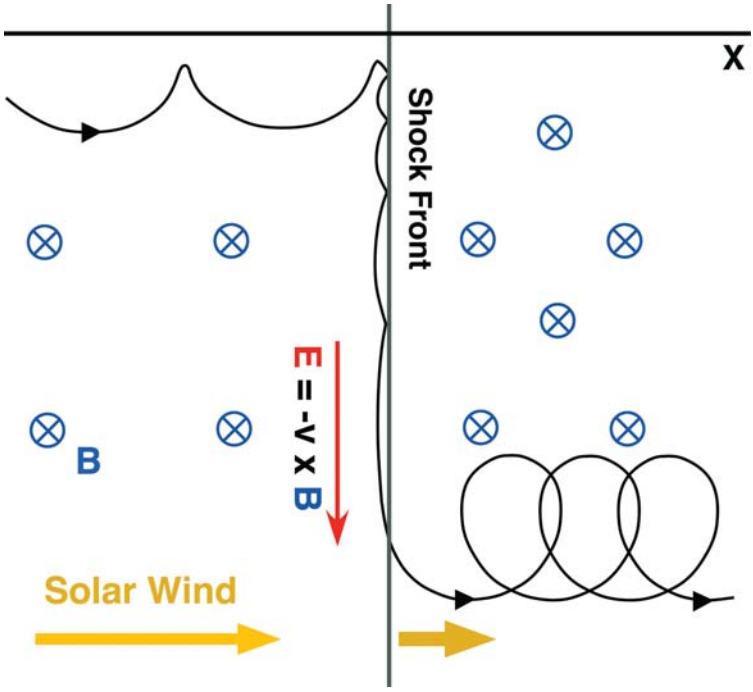


Figure 7. Schematic view of a shock surfing trajectory. An ion that arrives at the shock front with low speed in the shock frame from the left gains energy in the convective electric field in the course of several bounces.

how solar wind ions of only 1 keV/amu are pre-accelerated for injection at CIRs. Even the injection of pickup ions currently cannot be described quantitatively.

Particle Acceleration: The Bigger Picture

In summary, we have achieved a solid understanding of the basic processes that contribute to particle acceleration in the heliosphere. With a big step in the sensitivity and collection power of particle instrumentation that allows us to clearly distinguish elements, isotopes, and ionic charge states from solar-wind energies to those of cosmic rays, we have identified several different key sources that feed substantially into the energetic particle populations, such as solar wind, corona, interstellar pickup ions, possibly pickup ions from other heliospheric sources, impulsive flare plasma, and re-cycled energetic particle populations themselves. We are able to sample in detail energetic particle populations in the making at the Earth's bow shock, at interplanetary travelling shocks in CMEs, and at the shocks related to CIRs. We have sampled at least briefly the other planetary shock waves, and possibly we will have clear *in-situ* evidence from acceleration at the termination shock in the near future.

The ISSI book on CIRs paints a clear picture of the acceleration of two distinct and relatively clean sources during solar-minimum conditions, solar-wind and interstellar pickup ions, mostly unpolluted by solar energetic-particle events. This effort has led to the relatively broad consensus that Fermi acceleration is at work, most prominently at the reverse shock, and that even when no shock has yet formed a similar process is effective in the compression region, just not as strong yet. The book also contains compelling evidence from various observations that interstellar pickup ions have a clear advantage to be accelerated. The follow-on conjecture that dust related inner source ions should also have a similar advantage has led to a search that has given us a negative answer so far for observations at 1 AU. Together these pieces of evidence provide us with clues and constraints on the still vexing injection problem. The ISSI book on CIRs has compiled the current ideas on how solar-wind or pickup ions take the first step towards acceleration, but no convincing conclusions are yet on the horizon, even five years after the CIR effort. The infusion of ideas towards a resolution may come from the microscopic and multi-point study of the bow shock that is underway with Cluster, aided by extensive simulations. A workshop at ISSI has taken a current snapshot, which is also being discussed among other aspects in this volume. A compilation of the acceleration scenarios in CMEs during the solar-maximum phase with an even more complex source composition is underway in the ISSI book on CMEs[6].

The quantitative description of the acceleration regions in the inner Solar System, for which we have a plethora of *in-situ* observations, as touched upon in this article, will be the stepping stone towards understanding the largest shock in our home system, the solar-wind termination shock. By applying appropriate scaling laws, we can also transfer this knowledge to particle acceleration on the grand scale in astrophysics, i.e. to supernova shock waves and galactic winds, where we are convinced the birthplace of the much more energetic galactic cosmic rays lies. In the end, we will have to solve the remaining puzzles, such as the injection problem, through detailed studies on our own doorstep, before we can apply the new insight to regions that we can only access with remote sensing. To combine the ever-increasing body of information into a comprehensive view and to build interdisciplinary bridges between *in-situ* analysis and remote-sensing observations, ISSI provide a unique forum, which has already created and continues to create links that were not there before[37].

References

1. R. Diehl, E. Parizot, R. Kallenbach & R. von Steiger (Eds.), "The Astrophysics of Galactic Cosmic Rays", Space Sci. Ser. of ISSI Vol. 13, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.*, **99**, Nos. 1-2, 2002.
2. J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger & R. Kallenbach (Eds.), "Cosmic Rays and Earth", Space Sci. Ser. of ISSI Vol. 10, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.*, **93**, Nos. 1-2, 2000.
3. L.A. Fisk, J.R. Jopikii, G.M. Simnett, R. von Steiger & K.-P. Wenzel (Eds.), "Cosmic Rays in the Heliosphere", Space Sci. Ser. of ISSI Vol. 3, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.,* **83**, Nos. 1-2, 1998.
4. R.A. Mewaldt, R.C. Ogliore, G. Gloeckler & G.M. Mason, in: *Solar and Galactic Composition*, R. Wimmer (Ed.), *AIP Conf. Proc.,* **598**, 393, 2001
5. A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach & H. Kunow (Eds.), "Corotating Interaction Regions", Space Sci. Ser. of ISSI Vol. 7, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.*, **89**, Nos. 1-2, 1999.
6. H. Kunow, N. Crooker, J. Linker, R. Schwenn & R. von Steiger (Eds.), "Coronal Mass Ejections", Space Sci. Ser. of ISSI, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.* in prep., 2004.
7. F.B. McDonald, B.J. Teegarden, J.H. Trainor, T.T. von Rosenvinge & W.R. Webber, *Astrophys. J.,* **203**, L149, 1976.
8. G.M. Mason & T.R. Sanderson, in Ref. 5, p. 77.
9. D.V. Reames, *Space Sci. Rev.,* **90**, 413, 1999.
10. G.M. Mason, J.E. Mazur & J.R. Dwyer, *Astrophys. J.*, **525**, L133, 1999.
11. R. von Steiger *et al.*, *J. Geophys. Res.,* **105**, 27217, 2000.
12. D.V. Reames, *Astrophys J. Suppl.* **73***,* 235, 1990.
13. J.E. Mazur *et al., AIP Conf. Proc.,* **528**, 47, 2000.
14. M.I. Desai *et al., Astrophys. J.*, **588**, 1149, 2003.
15. R. von Steiger, R. Lallement & M.A. Lee (Eds.), "The Heliosphere in the Local Interstellar Medium", Space Sci. Ser. of ISSI Vol. 1, Kluwer Acad. Publ., Dordrecht, and *Space Sci. Rev.,* **78**, Nos. 1-2, 1996.
16. D. Hovestadt, G. Gloeckler, B. Klecker & M. Scholer, *Astrophys. J.,* **281**, 463, 1984.
17. E. Möbius *et al.*, *Nature,* **318**, 426, 1985.
18. G. Gloeckler *et al.*, *J. Geophys. Res.,* **99**, 17637, 1994.
19. G. Gloeckler, in Ref. 5, pp. 91-104.
20. D. Morris *et al.*, *AIP Conf. Proc.,* **598**, 201, 2001.
21. K. Bamert *et al.*, *J. Geophys. Res.,* **107**, 1130, 2002.
22. H. Kucharek *et al.*, *J. Geophys. Res.,* **108**, 8030, doi: 10.1029/2003JA009938, 2003.
23. J. Geiss, G. Gloeckler & R. von Steiger, in Ref. 15, pp. 43-52.
24. R.F. Wimmer-Schweingruber & P. Bochsler, *Geophys. Res. Lett.,* **30**, 1077, doi: 10.1029/2002GL015218, 2003.

25. E. Möbius *et al., Geophys. Res. Lett.,* **29**, 1016, 2002.
26. J.E. Mazur, G.M. Mason & R.A. Mewaldt, *Astrophys. J.,* **566**, 555, 2002.
27. M. Scholer in Ref. 5, pp. 105-114.
28. T. Sugiyama & T. Terasawa, *Adv. Space Res.*, **24**, 73, 1999.
29. J. Giacalone, J.R. Jokipii & J. Kota, *Astrophys. J.*, **573**, 845, 2002.
30. G.M. Mason *et al.*, in Ref. 5, pp 327-367.
31. M. Scholer *et al.*, in Ref. 5, pp. 369-399.
32. F.C. Jones & D.C. Ellison, *Space Sci. Rev.,* **58**, 259, 1991.
33. D.C. Ellison & R.R. Ramaty, *Astrophys. J.,* **298**, 400, 1985.
34. L.A. Fisk & J.R. Jokipii, in Ref. 5, pp. 115-124.
35. R.Z. Sagdeev, *Rev. of Plasma Phys.,* **4**, 23, 1966.
36. M.A. Lee, V.D. Shapiro & R. Sagdeev, *J.Geophys. Res.*, **101**, 4777, 1996.
37. E.M. would like to thank the International Space Science Institute and the Physikalische Institut at the Universität Bern for their hospitality during the preparation of this paper and gratefully acknowledges the support by the Hans-Sigrist Stiftung. The work was supported under NASA Grants NAG5-10890 and 12929 and INTAS Grant WP 01-270.

Interstellar and Pre-Solar Grains in the Galaxy and in the Solar System

P. Frisch[a], E. Grün[b] and P. Hoppe[c]

[a]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, USA*
[b]*Max-Planck-Institut für Kernphysik, Heidelberg, Germany and Hawaii Institute of Geophysics and Planetology, Honolulu, USA*
[c]*Max-Planck-Institut für Chemie, Mainz, Germany*

Introduction

Dust grains represent an important repository of cosmic matter, tracing stages of stellar evolution in our Galaxy. Three interdisciplinary workshops at ISSI in 1997 and 1998 studied dust grains at complementary phases of their life cycle. At these workshops the properties of astronomical interstellar dust grains (ISDG) were compared with in-situ ISDG data and pre-solar grains found in meteorites. The astronomical dust data yield grain composition and size distribution through optical, infrared (IR), and ultraviolet (UV) absorption and emission properties. In contrast, in-situ spacecraft data provide the ISDG mass distribution after correction for heliospheric interactions. Precise laboratory studies of presolar grains from meteorites yield the composition and origin of the stardust that parented the ISDGs. Bringing these views together at ISSI provided a new viewpoint of the mass distribution of ISDGs impinging on the Solar System, the gas-to-dust mass ratio and grain composition in the local interstellar cloud, grain formation in stellar atmospheres, destruction in the interstellar medium (ISM), and the composition of pre-solar versus ISDGs.

It became clear during the ISSI meetings that relating pre-solar and interstellar grains requires allowance for grain processing in space, including radiative damage, alteration, spallation, reheating, and exposure to additional compounds in dense interstellar clouds. Ed Anders noted: "There is no real contradiction between the meteoritic and astronomical data on grains. The vast majority of interstellar grains are reprocessed, homogenized [and fragmented]. The pristine circumstellar grains are chance survivors of a stochastic process that destroyed all but 10^{-4} of them." These same processes convert crystalline silicates detected in stellar atmospheres to amorphous silicates seen in interstellar space.

Several publications arose from these ISSI workshops, including a paper[1] in the Astrophysical Journal, and a special issue of the Journal of Geophysical Research (JGR), which contains 17 articles[2] delving into different aspects of these grain populations (http://www.agu.org/journals/ja/ja0005/ja105_5.html). In the following sections, we outline the state-of-knowledge at the present time.

Galactic Dust

Dust in space was discovered towards the beginning of the 20th century, with dark dust clouds dominating observations[3] (Fig. 1). The early research of Mayo Greenberg, an active participant in the ISSI workshops, helped lay the foundations for our understanding of interstellar dust grains. A consistent theoretical description of interstellar dust requires a grain mixture that varies according to the relative amounts of diffuse and dense interstellar clouds. The Local Interstellar Cloud (LIC) is a low density weakly ionized interstellar cloud[4] [n(H)~0.3 cm^{-3}, n(e)~0.1 cm^{-3}] showing larger abundances of heavy refractory

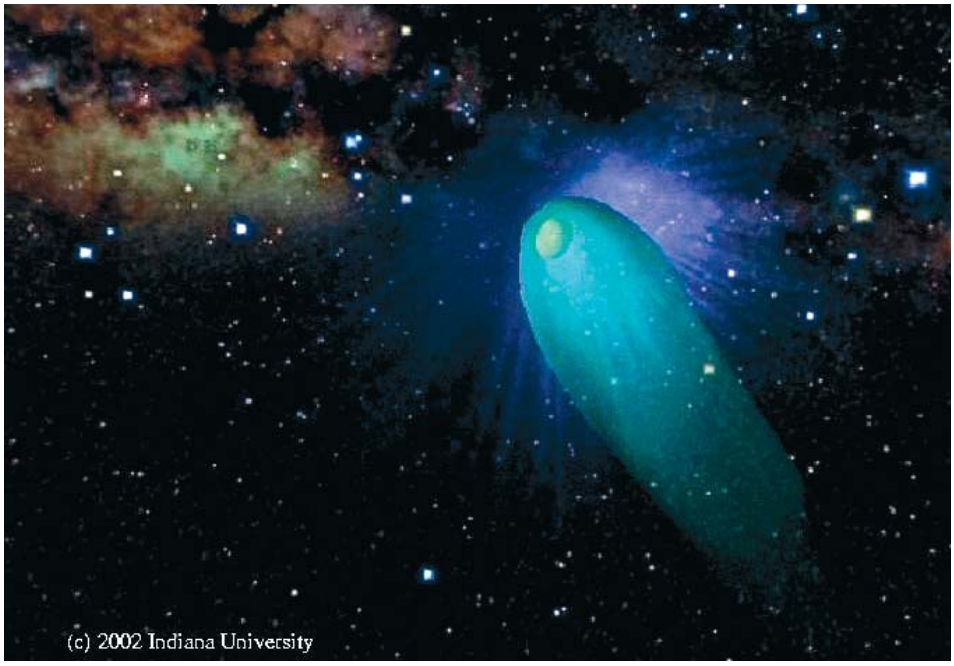


Figure 1. The heliosphere viewed against the plane of the Milky Way towards the Galactic Centre direction. Interstellar dust grains in dark clouds are seen to obscure background starlight in this composite Milky Way Galaxy image based on an A. Mellinger photograph and stars in the Hipparcos database. Dust is also found in the tenuous transparent interstellar cloud around the heliosphere. (Visualization by A. Hanson, P. Fu and P. Frisch, based on T. Linde's MHD model of the heliosphere embedded in a magnetized diffuse interstellar cloud. 3D star positions from the Hipparcos satellite.)

elements such as Fe, Ca, Mg, and Si in the gas-phase than cold dense interstellar clouds. Theoretical models of the lifetimes of interstellar dust grains indicate that LIC abundances are explained by grain destruction in violent interstellar shock fronts with velocities of 50 - 100 km s^{-1}, which shatter grains and vapourize a small fraction of the refractory elements in the grain core[5]. The interaction of galactic dust with the heliosphere depends on the grain mass and charge. The smallest ISDGs (radii ≤ 0.05 μm) are trapped by the interstellar magnetic field and diverted around the heliosphere. Larger grains penetrate the heliosphere where they are measured by interplanetary spacecraft[1, 6-8].

Astronomical grain populations

The ISSI workshops on interstellar dust inside and outside of the heliosphere, and pre-solar grains, led to the first detailed comparisons between these diverse dust data. The properties of ISDGs are ordinarily determined from observations of clouds 100 to 10 000 times denser than the LIC. Starlight is scattered, extinguished, and polarized by interstellar dust grains in the diffuse ISM, where gas densities are n_{gas}<100 cm^{-3}. An understanding of grain sizes in the LIC is required to interpret data on grains inside the heliosphere, and this data is provided by optical data from more distant stars, which reveal a power-law dependence for grain sizes (proportional to ~$a^{-3.5}$, for grain radii "a" and a = 0.005 - 0.25 μm[9]). The raw grain material required to explain observed spectral features in the ultraviolet (such as the extinction bump at 2175 Å) and infrared (such as the emission features at 9.8 μm, 18 μm, and 3.3 - 11.4 μm) includes a mixture of tiny carbonaceous particles (either PAH's or graphite, a <50 Å), and amorphous silicates[9, 10]. Denser clouds and ionized regions exhibit a notable lack of the tiny PAH grains, which appear to produce the far UV extinction, and infrared observations of molecular clouds and in-situ observations of dust inside the Solar System show that larger dust grains (a~1 μm) are present.

The low-density ISM near the Sun does not have enough dust for UV or infrared detection. However, Ulysses and Galileo discovered a population of large grains[6] with a ~ 1 μm. One puzzle is that Weingartner & Draine[11] found that these large grains are not required to model the size distributions of the tiny carbonaceous grains causing the infrared emission and the 2175 Å extinction bump. Although grain models are still uncertain, one conclusion of the ISSI workshops is that interstellar dust and gas may decouple over the lifetime of intermediate velocity clouds such as the LIC, which itself appears to have originated as a fragment of an expanding superbubble shell generated by stellar evolution in the Scopius-Centaurus Association.[1]

Classical interstellar dust grains are aligned by interstellar magnetic fields and polarize background starlight. One of the earliest tracers of very nearby interstel-

lar dust was polarized starlight observed for stars towards the galactic centre, which we now know corresponds with the location of most of the mass of the ISM within 30 pc[12]. Comparison between the wavelength distribution of polarized starlight and broad unexplained optical interstellar absorption features known as the diffuse interstellar bands (DIBs) establishes that the DIB features are carried by the tiny grains which most likely are PAHs[9,10].

Interstellar dust grains in the LIC

One of the most fundamental issues in astrophysics is the chemical composition of matter in our Universe. One of the results of our ISSI dust workshops was a new way to evaluate the ratio of the masses of the gas versus the dust proportions ($R_{g/d}$) of a diffuse interstellar cloud. If the total chemical composition of a cloud is known, and combined with abundances of elements observed in the gas phase, in principle both $R_{g/d}$ and the grain composition can be determined. This is the "missing-mass argument", and follows from a premise that the original cloud material remains together as either gas or dust throughout the cloud lifetime[1]. Classical ISM theory presumes that the assumed cloud abundance pattern is solar[13], but data on elements such as Kr, O, S, suggest lower intrinsic abundances at about ~80% solar that are reminiscent instead of the abundances of much younger B-stars[14]. We applied these missing-mass arguments to the LIC data and found $R_{g/d}$~170 if the LIC chemical composition is solar, or $R_{g/d}$~600 if it is comparable to B-star abundances[1,4]. However, when we use directly the Ulysses and Galileo data we find that $R_{g/d}$ <110 (the upper limit is used because small grains are excluded from the heliosphere). The difference between values obtained from missing-mass arguments versus in situ data indicates that larger ISDGs may decouple somewhat from the gas over the small spatial scales[15] represented by the LIC (< 0.5 pc).

The warm (6400 K) LIC shows higher abundances of refractory elements such as Ca and Fe in the gas phase than normal for cold clouds in the Milky Way Galaxy disk. Virtually all Fe in a cold cloud is in grains, while ~20% of the Fe in the warm LIC is in the gas phase. Grain destruction also preferentially erodes the silicate-rich mantle overlaying grains in dense clouds. Since the large grains remain in the LIC, either grain destruction is incomplete, or large grains are themselves replenished in space. The LIC grain composition inferred from missing-mass arguments favours a mixed oxide/silicate composition[1].

Connecting stardust and interstellar dust

Notable among the JGR special section papers is an exploration of the condensation of dust grains in stellar atmospheres, and a comparison of this grain mineralogy with interstellar dust composition[16]. This investigation represents the first effort to theoretically simulate the composition of "stardust" in terms of the condensation

pattern in stellar outflows such as asymptotic giant branch (AGB) star atmospheres, and it successfully reproduced interstellar abundances of the three elemental groups most depleted onto interstellar dust grains in cold clouds. For an initial atmosphere with solar composition, the predicted condensations of Ca and Ti (the most heavily depleted ISM elements) are consistent with the formation of Ca- and Ti-rich oxides at high temperatures (>1500 K). Moderately depleted elements (Cr, Co, Fe, V, Ni) condense as olivine and metal alloys at 1250 - 1500 K. The mildly depleted elements Mg and Si condense first as olivine above 1350 K (Mg/Si = 2), but convert to orthopyroxene at lower temperatures (Mg/Si = 1) as more Si condenses out. Ebel[16] also found that most Fe condenses into metals, with minimal Fe in silicates. This silicate condensation pattern is consistent with the preferential erosion of Si-rich grain surfaces seen in the ISM[1]. The correspondence between the three depletion groups in the ISM and condensation phases in stellar outflows suggests that refractory element abundance patterns of interstellar dust grains are related to grain formation in stellar atmospheres.

Interstellar Dust in the Solar System

The only direct observation of ISDGs close to the Sun is the weak polarization of 36 Oph (~6 pc) from magnetically aligned grains[12]. From observations of nearby interstellar gas we know that the Solar System passes currently through a shell of material that is located at the edge of the Local Bubble. It emerged from the interior of this bubble within the past 10^5 years. The local interstellar cloud may have been ejected by supernova explosions from the molecular clouds and star-forming regions in the Scorpius-Centarus Association. It is clear that in-situ sampling of dust from this cloud would greatly help us to understand the nature and processing of dust in various galactic environments, and cast new light on the chemical composition and homogeneity of the interstellar medium.

More than a decade ago, interstellar dust was positively identified inside the planetary system. After its flyby of Jupiter, the dust detector on board the Ulysses spacecraft detected impacts predominantly from a direction that was opposite to the expected impact direction of interplanetary dust grains (Fig. 2). It was found that, on average, the impact velocities exceeded the local Solar System escape velocity, even if radiation pressure effects were neglected[6]. Subsequent analysis showed that the motion of the interstellar grains through the Solar System was parallel to the flow of neutral interstellar hydrogen and helium gas, both travelling at a speed of 26 km/s. The interstellar dust flow persisted at higher latitudes above the ecliptic plane, even over the poles of the Sun, whereas interplanetary dust is strongly depleted away from the ecliptic plane.

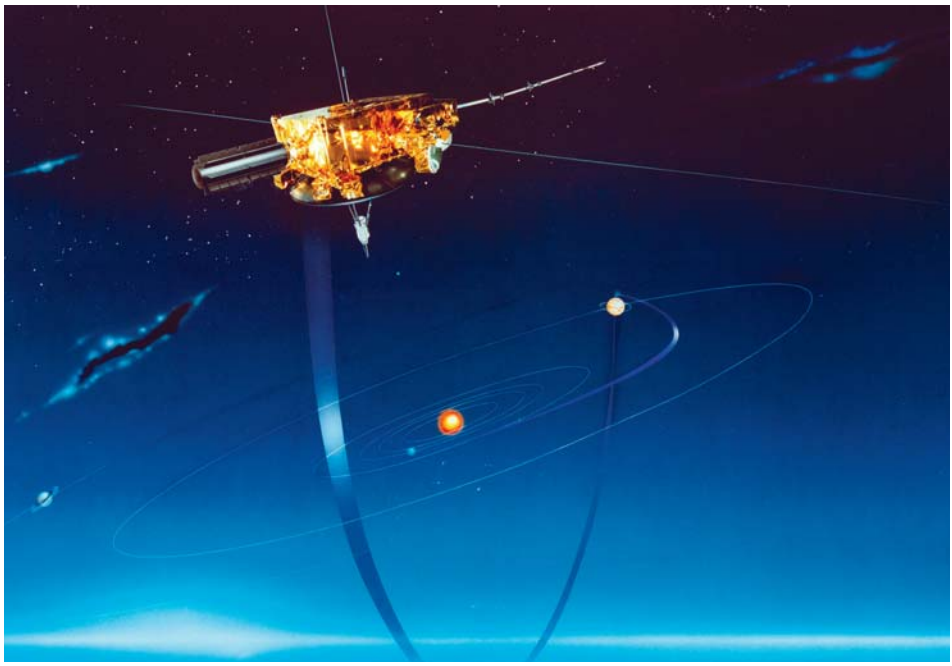


Figure 2. The Ulysses spacecraft in its highly inclined orbit above the planetary system is in a unique position to detect interstellar grains passing through the heliosphere. (From ESA-NASA Ulysses Project Team.)

Since that time, Ulysses has monitored the stream of interstellar dust grains through the Solar System at higher latitudes. In mid-1996, a decrease by a factor 3 in the interstellar dust flux density was observed[17]. This decrease was attributed to the increased filtering of small grains by the solar-wind magnetic field during solar-minimum conditions[18]. Since early 2000, Ulysses has detected the earlier higher interstellar-dust flux levels again[19]. Interstellar dust had initially been identified outside 3 AU up to Jupiter's distance. However, refined analyses[20,21] showed that both Cassini and Galileo recorded several 100 interstellar grains in the region between 0.7 and 3 AU from the Sun. Even in the Helios dust data, interstellar grains were identified down to 0.3 AU distance from the Sun[21].

Size distribution

The radii of clearly identified interstellar grains[6] range from 0.05 μm to above 1 μm with a maximum at about 0.3 μm. The deficiency of small grain masses ($a < 0.3$ μm) compared to astronomically observed ISD is not solely introduced by the detection threshold of the spacecraft instruments, but indicates a depletion of small interstellar grains in the heliosphere.

There are significant differences in the particle sizes that were recorded at different heliocentric distances. Mass-distribution measurements revealed a lack of small (<0.3 μm) ISD grains inside 3 AU heliocentric distance[22]. Measurements by Cassini and Galileo in the distance range between 0.7 and 3 AU showed that interstellar particles were bigger than 0.5 μm, with increasing masses closer to the Sun. The flux of these big particles did not exhibit temporal variations due to the solar-wind magnetic field like the flux of smaller particles observed by Ulysses. The trend of increasing masses of particles continues, as demonstrated by Helios measurements, which recorded particles of ca. 1 μm radius down to 0.3 AU. These facts support the idea that the ISDG stream is filtered by the radiation pressure as well. It is concluded that interstellar particles with optical properties of grains consisting of astronomical silicates or organic refractory materials are consistent with the observed radiation pressure effect[23].

Recently, even bigger (40 μm) interstellar meteors have been reliably identified by their hyperbolic speeds in radar meteor observations[24-26]. The flow direction of these larger particles varies over a much wider angular range than that of small (sub-micron- sized) grains observed by spacecraft. Baggaley[25] identified a general background influx of extra-Solar System particles from southern ecliptic latitudes with enhanced fluxes from discrete sources. Meteor observations with the Arecibo radar are more sensitive than the above: Meisel *et al*.[26] found a flux of several micron-sized interstellar meteor particles radiating from the direction of the Geminga supernova.

There are important consequences from the big particle population in the local diffuse interstellar medium. While the particles observed by spacecraft couple to the interstellar medium on length scales of < 1 parsec via electromagnetic interactions, more massive grains couple to the gas over much longer scales of 100 to 1000 pc[15]. Therefore, big interstellar meteor particles travel unaffected over much longer distances and may come directly from their source region.

Pre-solar Grains in Meteorites and IDPs

Historical background and astrophysical implications

Our Solar System formed from the collapse of a molecular cloud about 4.57 billion years ago, possibly triggered by a supernova (SN) explosion[27,28]. The release of gravitational energy led to the evaporation of a large fraction of the dust grains in the molecular cloud and much of the nucleosynthetic memories carried by these grains was erased by chemical and isotopic equilibration. A small fraction of the dust grains, however, survived in relatively cool regions of the solar nebula and were incorporated into planetary bodies. In comets and small

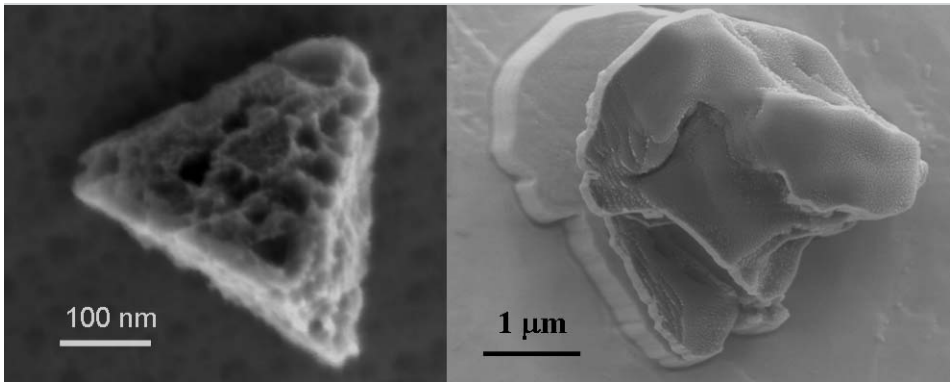


Figure 3. Pre-solar grains separated from primitive meteorites. Left: SiC. Right: Corundum [Max-Planck-Institut für Chemie, Mainz.]

asteroids, these pre-solar grains escaped destruction by planetary metamorphism and they are finally carried to the Earth by meteorites and interplanetary dust particles (IDPs).

Pre-solar grains are recognized by their anomalous isotopic compositions. The first pre-solar minerals, namely diamond and silicon carbide (SiC), were isolated from primitive meteorites in 1987 by Ed Anders and co-workers at the University of Chicago[29,30]. The noble gases played a key role in the identification of carbonaceous pre-solar grains, not only for diamond and SiC, but also for graphite, which was discovered several years later. In the following decade, pre-solar oxides such as corundum (Al_2O_3) and spinel ($MgAl_2O_4$) and silicon nitride (Si_3N_4) were found in acid-resistant residues from primitive meteorites by single-grain isotopic studies in the ion microprobe. Pre-solar silicates (variable composition) were identified only recently. The search for them was complicated because the harsh chemical treatments used to prepare acid-resistant residues destroy silicates. The invention of the new-generation NanoSIMS 50 ion microprobe makes it now possible to search in slices of meteorites and IDPs for in-situ pre-solar dust with sizes in the sub-micrometre range, and the application of this method has been essential for the discovery of pre-solar silicates. Examples of pre-solar grains separated from meteorites are shown in Figure 3.

The laboratory study of pre-solar grains has opened a new window in astronomy. It was quickly realized that the pre-solar grains must have formed around evolved stars (Fig. 4) or in the ejecta of SN explosions, i.e. they represent a sample of stardust that can be analyzed with high precision in the laboratory. Isotopic and structural studies have provided a wealth of information on stellar nucleosynthesis and evolution, mixing in SN ejecta, Galactic chemical evolution,

grain formation in stellar environments, and the types of stars that contributed dust to our Solar System.

Table 1 lists pre-solar minerals, their abundances, sizes, and stellar sources. In the following sections we will briefly summarize the isotopic properties and stellar sources of presolar silicon carbide, graphite, silicon nitride, corundum, spinel, and silicates. More detailed information can be found in recent review papers[31,32].

Table 1. Pre-solar minerals in meteorites and IDPs

Mineral	***Abundance (ppm)***	***Size (μm)***	***Stellar source***
Diamond	1000	0.002	Supernovae
SiC mainstream	10	0.2-10	AGB stars
SiC X	0.1	0.2-10	Supernovae
Graphite	1	1-10	Supernovae, AGB stars
Silicon nitride	0.001	1	Supernovae
Corundum	0.1	0.2-5	RGB & AGB stars
Spinel	10	0.2-5	RGB & AGB stars
Silicates in IDPs	1000	0.1-1	RGB & AGB stars
Silicates in meteorites	100	0.1-1	RGB & AGB stars

Silicon carbide, graphite, and silicon nitride

Silicon carbide is the best-studied pre-solar mineral phase. Based on the isotopic compositions of C, N and Si and the abundance of radiogenic ^{26}Mg (from the radioactive decay of ^{26}Al) SiC was divided into six different populations. Most abundant are the mainstream grains (about 90% of the total SiC), which are believed to have formed in the winds of 1-3 $M_\odot$ AGB stars. Such stars are also the source for some of the graphite grains.

The isotopic compositions of a rare sub-group of pre-solar SiC grains, the so-called X grains, and of most graphite and Si_3N_4 grains show imprints of advanced stellar nucleosynthesis, and those grains are most likely from SN. The presence of the decay products of short-lived radioactive ^{44}Ti (half-life 60 years) and ^{49}V (half-life 11 months) indicates that the grains incorporated matter from the innermost regions of the exploding star and that grain growth occurred on a time scale of several months.

Corundum, spinel and silicates

Similar to the light elements in SiC and graphite, the O-isotopic compositions of pre-solar corundum and spinel range over many orders of magnitude. The O-isotopic signatures suggest that most of these grains formed in the winds of 1-4 $M_\odot$ red giant branch (RGB) and AGB stars. Grains from SN are apparently rare. To

Figure 4. Hubble Space Telescope image of the Cat's Eye Nebula. The evolved AGB star (centre) ejected its mass in a series of pulses at 1500 year intervals. These convulsions created concentric dust shells, making a layered, onion-skin structure around the star (ESA, NASA, HEIC and the Hubble Heritage Team, STScI/AURA).

date, only one corundum grain (out of ~400 identified pre-solar corundum and spinel grains) shows a strong isotopic enrichment in ^{16}O, the predominantly expected signature of SN grains.

Among the pre-solar silicates identified to date are olivines, pyroxenes, and so-called GEMS (glass with embedded metal and sulphides). Abundances of pre-solar silicates are much higher in IDPs than in primitive meteorites (Table 1). The O-isotopic signatures of most grains are compatible with those of the majority of the pre-solar corundum and spinel grains, suggestive of formation in RGB and AGB stars. Similar to SiC and the refractory oxides, SN grains are apparently rare among pre-solar silicates (one out of ~50 identified pre-solar silicate grains).

Outlook

Once it became evident that galactic ISDGs are accessible to in-situ detection and even to sample-return to Earth, the Stardust mission was proposed to analyze and return samples of ISDGs together with samples of cometary dust[33]. In January 2004, Stardust made a fast flyby of comet Wild 2. En route to the comet, and on the return path to Earth, Stardust collected interstellar dust and analyzed it with the Cometary and Interstellar Dust Analyzer instrument, CIDA. This instrument provided the first high-mass-resolution analyses of a few tens of presumably interstellar grains[34]. CIDA is a time-of-flight mass spectrometer with a mass resolution $M/\Delta M > 100$ that analyses ions generated by dust particles that impact a target at speeds of 20 to 40 km/s. It was concluded that the main constituents of interstellar grains are organic with a high oxygen and low nitrogen content. They suggest that polymers of derivatives of the quinine type are consistent with all impact ion spectra recorded. Analyses of the returned samples will provide more information on the composition of ISDGs.

Based on the experience gained by previous in-situ dust measurements in space, a new approach of accurate dust trajectory measurement (a few percent in speed and a few degrees in angle) together with high-resolution chemical analysis of the same interstellar grain is being developed. Recent instrument developments allow us to combine sensors of these capabilities into a single dust telescope[35], designed for being carried into space by a dust observatory spacecraft in order to measure the composition of large numbers of ISDGs in interplanetary space. Despite the great advances made in the last years in the understanding of the heliospheric dust environment, there remain many important questions to be answered. Significant compositional information will be gained from future in-situ measurements, but even more can be learned if this dust is collected and brought to the laboratory, where the most advanced instrumentation can be used for its analysis. Sample-return from Earth orbit is, of course, much easier than sample return from distant worlds. However, in order to separately collect interplanetary and interstellar dust in Earth orbit, a preceding detailed analysis of the various dust-flow components by a dust telescope is necessary.

The chemical and isotopic compositions and physical properties of ISDGs provide important information on stellar evolution and nucleosynthesis, the physics and chemistry of the interstellar medium, and on Solar System formation and evolution. Precise data on interstellar grains inside our heliosphere offers the potential to revolutionize our understanding of the Milky Way Galaxy.

References

1. P.C. Frisch *et al*., *Astrophys. J*., **525**, 492, 1999.
2. Interstellar Dust and the Heliosphere, *J. Geophys. Res.,* **105**, pp. 10237-10417, 2000.
3. E.F. Van Dishoeck & A.G.G.M. Tielens, in J.A.M. Bleeker, J. Geiss & M.C.E. Huber (Eds.), The Century of Space Science, Kluwer Academic Publ., Vol. 1, p. 607, Dordrecht, 2001.
4. P.C. Frisch & J.D. Slavin, *Astrophys. J.,* **594**, 844, 2003.
5. A. Jones, in Ref. 2, p. 10257.
6. E. Grün *et al*., *Astron. Astrophys.,* **286**, 915, 1994.
7. I. Mann & H. Kimura, in Ref. 2, p. 10317.
8. T.J. Linde & T.I. Gombosi, in Ref. 2, p. 10411.
9. J.S. Mathis, in Ref. 2, p. 10269.
10. J. Lequex, in Ref. 2, p. 10249.
11. J.C. Weingartner & B.T. Draine, *Astrophys. J.,* **548**, 296, 2001.
12. J. Tinbergen, *Astron. Astrophys.,* **105**, 53, 1982.
13. N. Grevesse & A.J. Sauval, *Space Science Reviews,* **85**, 161, 1998.
14. T. Snow, in Ref. 2, p. 10239.
15. E. Grün & M. Landgraf, in Ref. 2, p. 10291.
16. D. Ebel, in Ref. 2, 10363.
17. M. Landgraf & E. Grün, in D. Breitschwerdt, M.J. Freyberg & J. Trümper (Eds.), The Local Bubble and Beyond, Lecture Notes in Physics, Vol. 506, Springer Heidelberg, p. 381, 1998.
18. M. Landgraf, in Ref. 2, p. 10303.
19. M. Landgraf, H. Krüger, N. Altobelli & E. Grün, *J. Geophys. Res.,* **108** (A10), 8030, 2003.
20. N. Altobelli *et al*., *J. Geophys. Res.,* **108** (A10), 8032, 2003.
21. N. Altobelli, Private communication, 2004.
22. M. Landgraf, W.J. Baggaley, E. Grün, H. Krüger & G. Linkert, in Ref. 2, p. 10343.
23. M. Landgraf, K. Augustsson, E. Grün & B.A.S. Gustafson, *Science,* **286**, 2319, 1999.
24. A.D. Taylor, W.J. Baggaley & D.I. Steel, *Nature,* **380**, 323, 1996.
25. W.J. Baggaley, in Ref. 2, pp. 10353.
26. D.D. Meisel, D. Janches & J.D. Mathews, *Astrophys. J.,* **567**, 323, 2002.
27. R. Kallenbach, W. Benz & G.W. Lugmair (Eds.), "From Dust to Terrestrial Planets", Space Science Series of ISSI Vol. 9, Kluwer Academic Publ., Dordrecht, and *Space Sci. Rev.,* **92**, Nos. 1-2, 2000.
28. A. Boss & H.A.T. Vanhale, in Ref. 27, p. 13.
29. R.S. Lewis, M. Tang, J.F. Wacker, E. Anders & E. Steel, *Nature,* **326**, 160, 1987.
30. T. Bernatowicz *et al*., *Nature,* **330**, 728, 1987.
31. E. Zinner, in K. K. Turekian, H.D. Holland & A.M. Davis (Eds.), "Treatise in Geochemistry", Elsevier, Oxford and San Diego, p. 17, 2004.
32. P. Hoppe & E. Zinner, in Ref. 2, p. 10371.
33. D.E. Brownlee *et al*., *J. Geophys. Res.,* **108**, 8111, 2003.

34. F.R. Krueger, W. Werther, J. Kissel & E.R. Schmid, *Rapid Commun. Mass Spectrom.,* **18**, 103, 2004.
35. E. Grün et al., in Ref. 2, p.10403.
36. P. Frisch would like to thank NASA for support through grants NAG 5-13107 and NAG 5-11005.

Comets and Their Interstellar Connections

W.F. Huebner[a] and K. Altwegg[b]

[a]*Southwest Research Institute, San Antonio, Texas, USA*
[b]*Physikalisches Institut, Universität Bern, Bern, Switzerland*

Although comets are among the smallest bodies in our Solar System, they nevertheless have an inordinately important scientific value. They represent material that had its origin in a dark molecular cloud from which our Solar System emerged. Comets provide insight into the history of our Solar System and connect the present-day composition of the planets, their satellites, and the Sun with the composition of the ancient molecular cloud, even though it has long since disappeared. They give insight into the conditions that prevailed in the Solar System during its formation and into the processes responsible for its current heterogeneity.

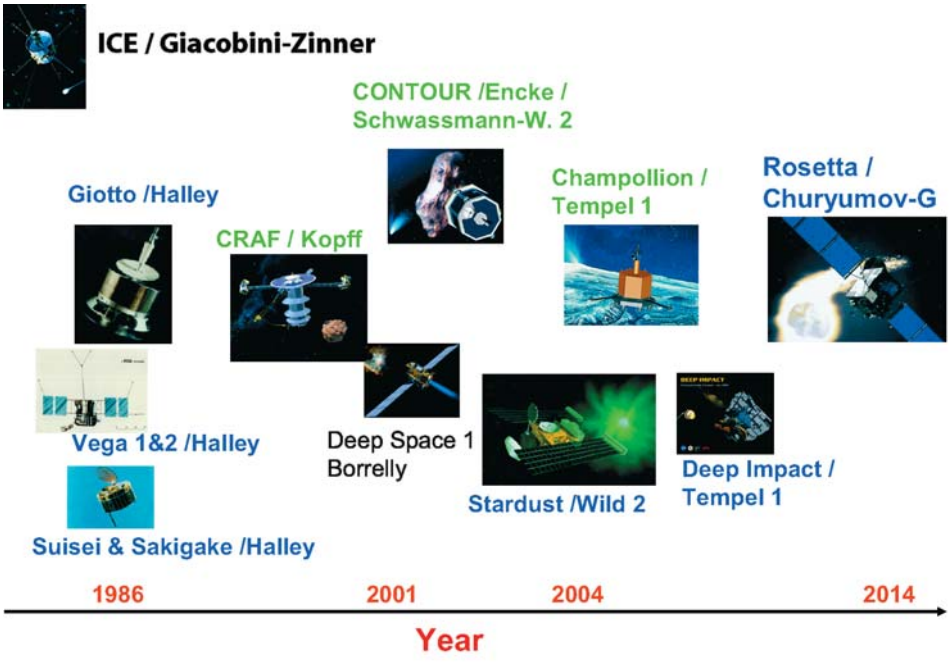


Figure 1. Comet missions from 1986 to the present. Blue: dedicated comet missions; black: missions where comet science is a secondary goal, green: cancelled or failed missions.

The reappearance of comet 1P/Halley in 1985 and the appearance of two bright comets towards the end of the last century, Hale-Bopp (C/1995 O1) and Hyakutake (C/1996 B2), triggered enormous scientific interest in these small icy bodies and in their relationship with the interstellar material. Many comet space missions were planned (see Fig. 1). However, after the multi-spacecraft visits to comet 1P/Halley in 1986, only a few smaller missions materialized. They left a gap of almost 30 years between the 1P/Halley encounter and the next comprehensive comet mission, Rosetta, which is now on its long journey to comet 67P/Churyumov-Gerasimenko. The ISSI Comet Workshop in 1998 led to the founding of a comet programme at ISSI with working groups, teams, and workshops in order to bridge this gap and to pass knowledge to the next generation of comet scientists that will be active at the time when Rosetta reaches its target in 2013.

Comets and the History of Solar-System Materials

The history of cometary material is complex; it is shown schematically in Figure 2. Atomic nuclei heavier than H and He are synthesized in the interiors of stars and released into the interstellar medium at particular stages of stellar evolution. Molecules form and condensation into solid grains occurs when densities are high and temperatures fall, as happens in stellar envelopes (producing high-temperature condensates), molecular clouds (producing low-temperature condensates), and proto-stellar disks. Different stellar sources release matter with different isotopic signatures that are preserved in grains. Comets are

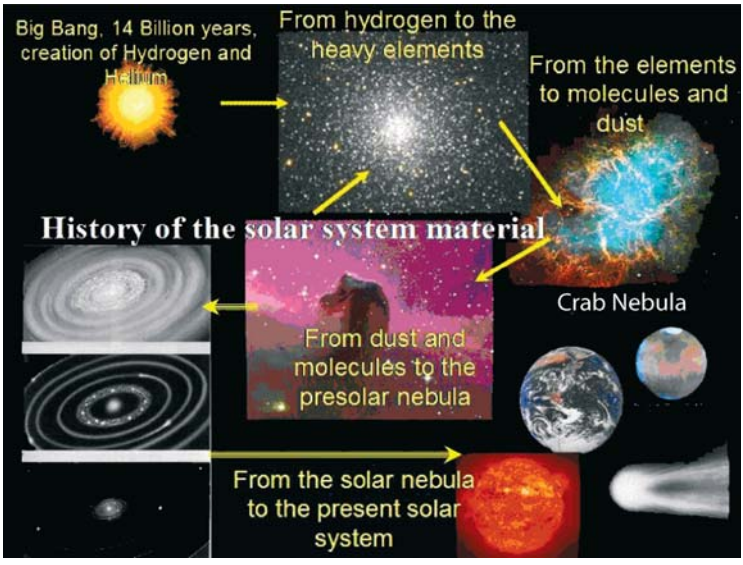


Figure 2. Sketch of the history of Solar System materials.

expected to have preserved a larger variety of grains with the original stellar signature than meteorites. So far, such isotopic evidence is sparse. A few grains with high and variable $^{13}C/^{12}C$ were found in the coma of comet Halley[1] indicating multiple stellar origins. More results are expected from the Stardust mission samples that are now on their way back to Earth and from the Rosetta mission, now on its way to comet 67P/Churyumov-Gerasimenko. In interstellar molecular clouds, these grains serve as nuclei for the condensation of more volatile molecular species. Upon irradiation with UV light and high-energy particles, the grains are processed further. About 4.6 Gy ago, a galactic local proto-stellar cloud partially collapsed forming our Solar System. Fred Whipple was one of the first to recognize the importance of comets for the history of our Solar System. In the last decades of the last century it was accepted that comets represent the best-preserved material to be found in the Solar System and that some molecules found in comets can be traced back to the interstellar medium. This allows us to study the processes that led from the molecular cloud, through agglomeration in the solar nebula, to the present constituents of the comet nuclei.

Composition and Origin of Cometary Material

Until about 1980, it was assumed that comet nuclei contained primarily frozen gases of H_2O, NH_3, and CH_4. Comet models based on this assumption, even when supplemented with a few other minor species, failed to explain the radicals and ions identified in spectra of comet comae. One of the first models to come close to explaining the observed composition of comet comae was based on frozen interstellar molecules in the nucleus, as proposed by Biermann *et al*.[2] As reviewed in several conference proceedings, much progress has been made in finding new molecular species in comets and in modeling and understanding comets. One of these conferences was the ISSI workshop on "Composition and Origin of Cometary Material"[3] in 1998. From this workshop, three main recommendations emerged. We will review progress made based on some of these recommendations.

Workshop Recommendations. The summary of the workshop on "Composition and Origin of Cometary Material" recommended[4]:

1. Comet data acquisition, analysis, and evaluation.
   - 1.1 Acquire data by remote sensing over large heliocentric distance ranges with modern techniques in all wavelength regions.
   - 1.2 Acquire data by *in situ* measurements.
   - 1.3 Acquire data from sample returns.
   - 1.4 Acquire data from comet simulation experiments.
   - 1.5 Develop global models of comet activity.

2. Establish a comet source book for data on material properties on a website for easy access.
   2.1 Establish a working group to deal with basic material properties.
   2.2 Atomic and molecular cross-sections and rate coefficients.
   2.3 Sputtering and heat desorption from solids.
   2.4 Contact laboratories to make measurements, and theoretical groups to make calculations.
3. Establish an interdisciplinary working group to model the collapse of an interstellar cloud. Include formation of the accretion shock in the outer regions of the solar nebula where comets form.
   3.1 Use comet data as a guide to constrain the models.
   3.2 Investigate the proto-planetary disks.
   3.3 Pursue the emerging field of Edgeworth-Kuiper belt objects.

We summarize below the actions taken to fulfil several of these recommendtions.

The Interstellar Connection

Molecules. New searches for molecular species in comets have been conducted (see, for example, Crovisier *et al*.[5,6]). Some searches were successful in identifying new mother molecules and isotopes; others resulted in placing upper limits on the abundances of potentially new cometary molecules. Table 1 summarizes the present status of identified cometary molecules and compares them with molecules identified in protostars (PS) and in dark interstellar clouds (DISC). We do not list ions in this table because the differences between the interstellar radiation field and that of the Sun are too large to make such a comparison meaningful. We could use the same argument for radicals as photo-dissociation products. However, in this case it is difficult to decide which radical is the result of photo-dissociation and which is the result of chemical reactions. We also do not list atoms or metallic or siliceous products derived from the dust of Sun-grazing comets.

Homologous series of molecules like C_nH and HC_nN will become excellent indicators to trace back cometary material to its source, be it molecular clouds or the solar nebula, when more cometary data will be available.

A more meaningful comparison of cometary with interstellar (extrasolar) molecules results if we list their relative abundances and include interstellar ices, which is done for some molecules presented in Figure 3. Such data are now becoming available[7]. Figure 3, based on Crovisier[8], illustrates a quantitative similarity between inter-stellar ice and cometary ice. The interstellar data is from

Table 1. Comparison of Identified Cometary and Interstellar Neutral Molecules

Molecule	Comet	PS	DISC	Molecule	Comet	PS	DISC	Molecule	Comet	PS	DISC
Homonuclear Molecules				HC_4H^*				CH_3NH_2			
H_2				HC_6H^*				CH_2NH			
N_2				$H_2C_4^r$				HNCO			
S_2				$H_2C_6^r$				NH_2CHO			
C_2				$C_6H_6^*$				NH_2CH_2COOH		?	
$C_3^{*,r}$				*l*-C_3H^r				NO^r			
C_4^r				*c*-C_3H^r				HNO			
$C_5^{*,r}$				*l*-$C_3H_2^r$				N_2O			
Oxidised carbon bearing				*c*-$C_3H_2^r$				*Sulphur bearing*			
CO				CH_3C_4H				SH^r			
C_2O				$H_2C_2H_2^*$				H_2S			
C_3O				CH_3CCH				CS			
C_5O				*Nitrogen bearing*				C_2S			
HCO^r				NH^r				NS^r			
H_2CO				NH_2^r				SiS			
CH_2CO				NH_3				C_3S			
CO_2				CN^r				C_5S			
HCOOH				C_3N^r				OCS			
CH_3OH				C_5N^r				H_2CS			
CH_3COOH				PN				HNCS			
CH_2CHOH				SiN^r				CS_2			
CH_3CH_2OH				HCN				SO			
$OH(CH)_2OH$				HNC				SO_2			
CH_2CHO				HC_3N				CH_3SH			
HC_2CHO				HC_5N				*Metal bearing*			
CH_3CHO				HC_7N				FeO		?	
CH_3OCHO				HC_9N				AlF^*			
CH_2OHCHO				$HC_{11}N$				$AlCl^*$			
c-CH_2OCH_2				$NaCN^*$				$NaCl^*$			
CH_3OCH_3				$MgCN^*$				KCl^*			
CH_3COCH_3				$MgNC^*$				*Others*			
$C_2H_5OCH_3$				$AlNC^*$				OH^r			
Hydrocarbons				$SiCN^*$				H_2O			
CH^r				H_2CN^r				HF		?	
CH_2^r				$HCCN^r$				HCl			
CH_3^r				CH_2CN^r				$CP^{*,r}$			?
CH_4^*				CH_3CN				SiH^r		?	
C_2H_6				CH_3NC				SiC^*			
C_2H^r				CH_2CHCN				SiO			
C_4H^r				CH_3C_2CN				*c*-SiC_2			
C_5H^r				CH_3CH_2CN	?			*c*-SiC_3			
C_6H^r				CH_3C_4CN		?	?	SiH_4^*			
$C_7H^{*,r}$				NH_2CN				$C_4Si^{*,r}$			
C_8H^r				HCCNC							
HC_2H				HNCCC							

* Only in envelopes of evolved stars ? Tentative identification

c- Cyclic molecule *l*- Linear molecule r-Radicals

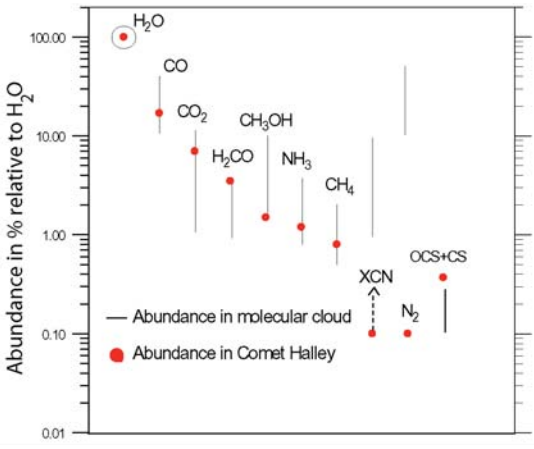


Figure 3. Comparison of molecular abundances in interstellar ices and in comet comae (after Ref. 8)

infrared spectra of icy grains in dense clouds, and the cometary data is from coma gases sublimed from the nucleus and dust. Except for N_2, which condenses only at very low temperatures and which therefore was probably never incorporated in comets or has been lost in the last 4.6 billion years, the similarity between interstellar and cometary ice is quite striking.

In making the comparisons of interstellar (and in particular dark cloud) molecules with Solar System (cometary) molecules, we must also remember that the Sun is a G2 star for which the relative abundance of O to C is about 2. Thus, the comparisons between the compositions of comets and protostars (including dark interstellar clouds) are only meaningful for protostars with O/C ≈ 2. Finally, the protosolar molecules listed in Table 1 are from observations of high-mass protostars. High-mass protostars evolve differently from low-mass protostars, such as the Sun. However, low-mass protostars are too faint for detailed molecular observations with current instrumentation.

We note that with few exceptions (N_2, S_2, CS_2, and C_2H_6), all molecules identified in comets have also been identified in protostars and in many cases in dark interstellar clouds. The four exceptions are molecules with no permanent dipole moment, and therefore no rotational emissions or absorptions in the radio range of the spectrum. Since radio astronomy is the primary tool for detecting interstellar molecules, these four molecules are very difficult to detect in interstellar space. The list of identified cometary molecules is thus consistent with the list of molecules that have their origin in protostars or dark interstellar clouds. It suggests that all cometary molecules may have traversed interstellar space and survived accumulation into the solar nebula. It also suggests that other interstellar molecules may be identified in cometary spectra.

Radicals and Noble Gases. There are two identified molecular radicals that appear to be present in cometary ices, namely CH_2 and C_4H [9,10]. C_4H is well-known from radio astronomy to be present in molecular clouds**,** and CH_2 is a relatively abundant species in chemical models of dense interstellar clouds.

These radicals are unlikely to have been present in sufficient quantities to be incorporated in ices frozen from a hydrogen-rich solar nebula. They are much more likely to have been incorporated into ices frozen on interstellar grain mantles at very low temperatures. This hypothesis requires that these icy mantles are preserved in transiting the accretion shock during their accumulation into the solar nebula.

So far, there is no reliable measurement of the abundance of noble gases in comets. However, there exist upper limits[10,11] for neon/H_2O < 1.5×10^{-3} and argon/oxygen[12] < 2.4×10^{-4} for some comets, which show that the noble gases are depleted compared with the Solar System values. This is not surprising for neon, which can be trapped in amorphous ice only at temperatures below 20 K. However, the existence of amorphous ice in comets is still controversial. The deficiency in argon points to a formation temperature of the nucleus of more than 60 K. These conclusions on the formation temperatures rest on the assumption that the laboratory experiments on ices accurately reflect the trapping of gases in cometary nuclei, which may not be valid.

Isotopes in Comets. Isotope ratios can provide additional clues to the physical conditions that prevailed during the formation of cometary volatiles. The most important is the D/H ratio in different molecules. A summary of all deuterium measurements in comets was compiled by Kallenbach *et al*.[13]

All comets for which data are available exhibit a similar D/H ratio in H_2O (i.e. HDO/H_2O), enriched by about a factor of two relative to terrestrial water and approximately one order of magnitude larger relative to the protosolar value. The D/H ratio in cometary HCN (i.e. DCN/HCN) is seven times higher than the value in cometary H_2O. All comets for which such data are available are Oort-cloud comets. Data for Kuiper-belt comets are not yet available. Species-dependent D-fractionations occur at low temperatures and low gas densities via ion-molecule or grain-surface reactions and cannot be explained by pure solar-nebula chemistry. Cometary volatiles appear therefore to preserve, at least partially, the interstellar D fractionation.

The observed D abundances set a lower limit to the formation temperature of 30±10 K. Similar numbers can be derived from the ortho-to-para ratio in cometary water, from the absence of neon in cometary ices, and the presence of S_2. So far, all cometary D/H measurements refer to bulk compositions, and it is conceivable that significant departures from the mean value could occur at the grain-size level. Strong isotope effects caused by chemical reactions in the coma can be excluded for H_2O and HCN. A comparison of the cometary $(D/H)_{H_2O}$ ratio with values found in the atmospheres of the outer planets is consistent with the

Table 2. Other Isotope measurements in different comets

Isotope			Comet
[$H^{12}CN/H^{13}CN$]	=	109 ± 22	Hale-Bopp[16]
[$H^{12}CN/H^{13}CN$]	=	111 ± 12	Hale-Bopp[17]
[$HC^{14}N/HC^{15}N$]	=	330 ± 98	Hale-Bopp[16]
[$HC^{14}N/HC^{15}N$]	=	323 ± 46	Hale-Bopp[17]
[$C^{14}N/C^{15}N$]	=	140 ± 30	Hale-Bopp, C/2000 WM1[18]
[$H_2{}^{16}O/H_2{}^{18}O$]	=	518 ± 45	Halley[19]
[$H_2{}^{16}O/H_2{}^{18}O$]	=	470 ± 40	Halley[20]
[$C^{32}S/C^{34}S$]	=	27 ± 3	Hale-Bopp[17]
[$^{32}S/^{34}S$]	=	23 ± 6	Halley[21]

long-held idea that the gas planets formed around icy cores with a high cometary D/H ratio, and subsequently accumulated significant amounts of H_2 from the solar nebula with a low protosolar D/H. Robert *et al*.[14,15] reviewed the currently available models and measurements of D/H in our Solar System.

The measurements of isotopes other than deuterium in the volatile part of comet comae are rare (see Table 2). So far, all isotope ratios reported for comets, apart from deuterium, are compatible with the terrestrial values. There is one exception, and that is the recently reported measurements of $^{14}N/^{15}N$ in comet C/2000 WM1 and comet Hale-Bopp[18] observed in the optical wavelength range. Both display the same isotopic abundance ratio, about 1 nitrogen-15 atom for each 140 nitrogen-14 atoms ($^{14}N/^{15}N = 140 \pm 30$). This value is about half of the terrestrial value (272). It is also very different from the result obtained by means of radio measurements of Comet Hale-Bopp ($^{14}N/^{15}N = 330 \pm 75$)[16,17]. Optical and radio measurements concern different molecules (CN and HCN, respectively), and this isotopic anomaly must be explained by some differentiation mechanism. The astronomers conclude that part of the cometary nitrogen is trapped in macromolecules in organic-dust-particle components.

An ISSI workshop in 2002 on "Solar System History from Isotopic Signatures of Volatile Elements" tried to put together all elements that determine the isotopic ratios in the different bodies of our Solar System.[13] It became clear that more work is needed on the outer Solar System in order to get a handle on the turbulent mixing in the early Solar System. The best accessible representatives of this region are comets. The Rosetta mission, being on its way to rendezvous with comet 67P/Churyumov-Gerasimenko, will be able to determine ratios for different isotopes in different parent molecules.[22] Comet 67P/Churyumov-Gerasimenko is one of the Edgeworth-Kuiper belt objects that are hard to investigate with remote sensing because of their usually small sizes. Comparing this comet with the Oort-cloud Comets 1P/Halley, Hale-Bopp, and Hyakutake will help to understand the isotopic heterogeneity observed in the Solar System.

The Interstellar Medium, the Solar Nebula, and Comet Formation

Observations of comets reveal information about the structure and composition of their nuclei. This in turn provides clues about the thermodynamic conditions and composition of the solar nebula. Most neutral molecular species in comets are also in the interstellar medium. The D/H ratio measured in a few comets is consistent with values expected from interstellar chemistry. This suggests that interstellar molecules were not chemically transformed in the solar nebula accretion shock in the region where comet nuclei formed. The hydrogen ortho-to-para ratio in cometary H_2O and NH_2 suggests nucleus formation at temperatures of about 30 K. Conversely, detection of crystalline silicate grains in comet comae indicates that the dust has been exposed to temperatures of about 900 K. One theory suggests that turbulent mixing and heating of dust in the inner solar nebula, followed by transport into the comet-forming zone, may be responsible for the crystallization of silicates. However, this also requires heating of the gas that entrains and transports the dust. Heating the gas would change its interstellar composition. Joint Discussion 14 at the XXVth General Assembly of the IAU in Sydney, Australia, initiated clarification of the conflicting evidence: survival of the low-temperature interstellar composition in the solar nebula in spite of the accretion shock and the detection of crystalline features in cometary dust.

From Comet Observation to Comet Nucleus Composition

So far, most of the knowledge we have about comet nucleus structure and composition stems from observations of comet comae, mostly by remote sensing. The bright coma of an active comet close to the Sun effectively obliterates the innermost coma and the nucleus from observations. On the other hand, when comets are far away from the Sun and there is almost no coma, comet nuclei are very faint objects and hard to investigate. There exist very few measurements from *in situ* observations (Comet 1P/Halley, Comet 81P/Wild2, Comet 19P/Borrelly). However, even there the measurements were not performed on the nucleus itself, but rather in the coma, a few hundred kilometres from the nucleus. Even when the Rosetta spacecraft reaches its target, the amount of data that will be gathered by the lander on the nucleus will be limited. It will also be limited because it concerns only one single comet. It is therefore mandatory to work with reliable models that connect the coma observations to the nucleus features. Figure 4 shows a sketch of the different regions of a comet. The original composition of the comet, present in the nucleus, will be changed and fractionated by diffusion processes, sublimation, chemical reactions with

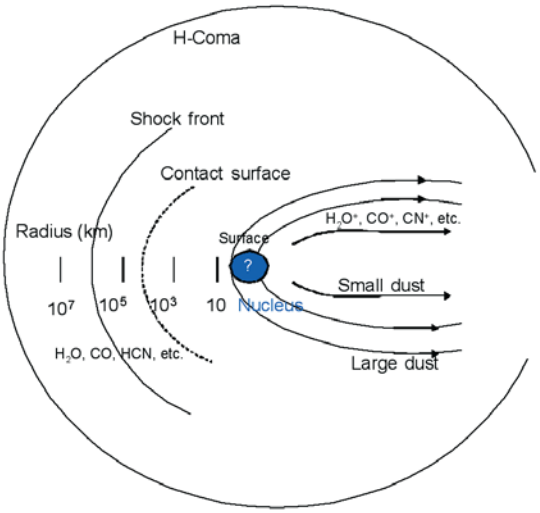


Figure 4. Sketch of the different regions of a comet coma. The H-coma is larger than the Sun.

neutrals and dust, photo-dissociation, interaction with solar-wind particles, etc. From the structure of the nucleus to the sublimation processes, from the innermost region of the coma out to the shock front there are a large number of processes responsible for the observed coma composition. In order to calculate back from the coma composition to what is really inside the comet nucleus, models, laboratory measurements, and observations are needed.

*Data for Comet Nuclei***.** A major goal of comet research is to determine conditions in the solar nebula based on the chemical composition and structure of comet nuclei. Until a comet mission examines the nucleus in situ, the composition of the nucleus is mostly deduced from observations of the coma. However, the composition of the coma changes[23] with heliocentric distance, *r*. Thus, the nucleus composition must be determined from analysis of coma mixing ratios as a function of *r*. New observing technology and early detection of some very active comets now permit coma mixing ratios to be determined over a large range of heliocentric distances. Even so, abundance ratios of species in the nucleus are very difficult to determine.

In coma analyses, three sources for the gas produced must be considered[24]. (1) the surface of the nucleus (releasing mostly water vapour and dust), (2) the interior of the porous nucleus (releasing many species more volatile than water), and (3) the distributed source (releasing gases from ices and organic poly-condensates trapped and contained in coma dust). Molecules diffusing inside the nucleus are sublimated from ice by heat transported into the interior. The mixing ratios in the coma are modelled assuming various chemical compositions and structural parameters of the spinning nucleus as it moves in its orbit from large heliocentric distance through perihelion. An investigation of comet-nucleus models has been undertaken by the ISSI Comet Nucleus Team.[25]

An extensive chemical reaction network has been assembled for coma analysis. Our original network[26] has been expanded to consist of molecules and molecular ions composed of H, He, C, N, O, Ne, S, Na, Mg, P, Si, Cl, K, Ar, and Fe. Photo-rate coefficients for dissociation, ionization, and dissociative ionization for various phases of activity of the Sun have been made available on a website: http://amop.space.swri.edu.

The ISSI Comet Nucleus Team[25] investigated and explored five algorithms for calculating the heat and gas diffusion in comet nuclei. Effects of amorphous ice[27-29] are often invoked to explain observed phenomena related to comet outgassing, outbursts, and splitting. However, Kouchi *et al*.[30] showed that formation of amorphous ice depends not only on low temperatures, but also on the speed of condensation. The density of the vapour phase of H_2O must be sufficiently high for condensation to take place rapidly, so that H_2O molecules do not have time to reorient themselves during condensation to form a crystalline lattice. They concluded that amorphous water ice could not form in the solar nebula.

The thermal conductivity of amorphous ice is much lower than that of crystalline ice. Klinger[27] derived a theoretical expression from classical phonon theory

$$\kappa = (1/4)\ v_s\ l_{ph}\ c\ \rho$$

where $v_s = 2.5 \times 10^3$ m/s is the speed of sound in ice, $l_{ph} = 5 \times 10^{-10}$ m is the phonon mean free path, and c and ρ are the specific heat and density of the H_2O ice, respectively. However, Kouchi *et al*.[28] measured conductivities $\kappa = 0.6 \times 10^{-5}$ to 4.1×10^{-5} W m^{-1} K^{-1} (valid in the range from $T = 125$ to 135 K) that are several orders of magnitude lower. Haruyama *et al*.[31] in their work of radiogenic heating of comet nuclei in the Oort cloud, adopted the values from Kouchi *et al*.[28], assuming the heat conductivity is proportional to temperature. Tancredi *et al*.[32] used a geometric mean between the values of Klinger and Kouchi, $\kappa = 7.1 \times 10^{-5}$ T W m^{-1} K^{-1} for their nucleus models. In another experiment, Kouchi and coworkers confirmed their earlier values and found that the phase transition from amorphous to crystalline ice is endothermic if the water ice has other gases trapped. Andersson and Suga[29] found an experimental value for the conductivity of nonporous, low-density amorphous water ice that is similar to the value derived by Klinger $\kappa \approx 0.6$ W m^{-1} K^{-1} in the temperature range $T = 70$ to 135 K. The thermal conductivity appears to also depend on the density of the amorphous ice. Figure 5 illustrates the enormous discrepancies in the conductivity values. Clearly, a reinvestigation of the thermal conductivity is in order.

The temperature-dependent thermal conductivity coefficient of the porous mixture in comet nuclei is usually derived from the conductivities of the different

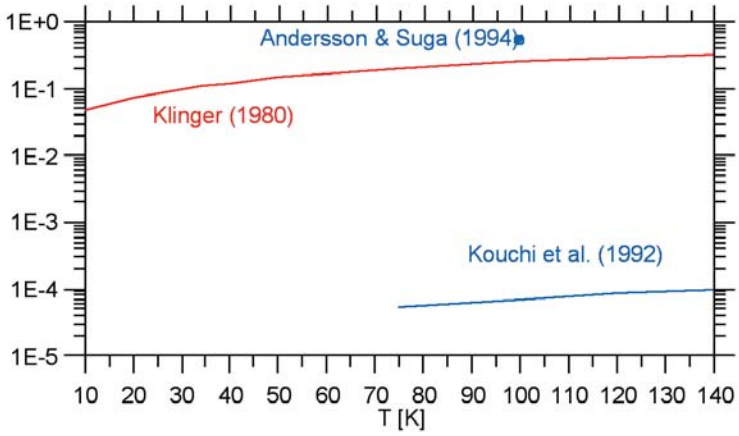


Figure 5. Comparison of thermal conductivities of amorphous ice averaged from measurements of Kouchi *et al*.[28] (blue line), a measurement of Andersson and Suga[29] (blue dot), and calculated values from Klinger[27] (red line). The results differ by more than three orders of magnitude.

dust and ice components, taking into account the porosity of the medium. As for the conductivity of crystalline ice, most modellers agree with the value introduced by Klinger.[27]

Global Models for Comet Activity. Another one of the current ISSI projects concentrates on improving our understanding of the comet's coma, the comet/solar-wind interaction, cometary ion chemistry, and cometary high-energy processes, through a combination of state-of-the-art modelling with analysis of observations of cometary environments. Comet coma models using MHD[33], 3D hybrid[34], and 3D direct Monte Carlo[35] simulations will be combined with a general comet environment model. This will allow a single model to be used to follow the development of the coma from the innermost part just above the nucleus surface out to the shock front and beyond. The understanding of comets and cometary processes gained from the modeling studies will be important preparatory work for future comet missions in general, and the Rosetta mission in particular.

In addition, the results will be very useful to missions other than Rosetta. Before Rosetta even gets close to the comet, there will be many interesting applications where the model can not only be tested, but also used for interpreting data. In particular, one can apply the results to *in-situ* measurements obtained by Giotto at Halley, Ulysses observations of Hyakutake's tail, X-ray observations, and optical and UV observations of comets. This will help the comet community, particularly in Europe, to bridge the long time until the Rosetta encounter.

Conclusions

The 1998 workshop at ISSI triggered a broad activity on cometary science. The follow-up on the recommendations of the ISSI workshop are in progress. The relationship of cometary molecules to interstellar molecules is now on a much firmer basis. Several tables of unidentified comet spectral lines exist[36-40]. The interstellar molecules listed in Table 1 should be used as a guide to identify these lines. Once Rosetta reaches its target, comet 67P/Churyumov-Gerasimenko, the models developed within the framework of ISSI will help us to understand the in-situ measurements and will therefore help to guarantee the maximum scientific return.

References

1. E.K. Jessberger & J. Kissel, in: "Comets in the Post-Halley-Era", R. Newburn (Ed.), Kluwer Academic Publishers, Dordrecht, 1989.
2. L. Biermann, P.T. Giguere & W.F . Huebner, *Astron. Astrophys.*, **108**, 221, 1982.
3. K. Altwegg, P. Ehrenfreund, J. Geiss & W.F. Huebner (Eds.), "Composition and Origin of Cometary Materials", SSS of ISSI, Vol. 8, Kluwer Academic Publ., Dordrecht, 1999, and *Space Sci. Rev.*, **90**, Nos. 1-2, 1999.
4. K. Altwegg, P. Ehrenfreund, J. Geiss, W.F. Huebner & A.-C. Levasseur-Regourd, in Ref. 3, p. 373.
5. J. Crovisier, D. Bockelée-Morvan, P. Colom, N. Biver & D. Despois, *Astron. Astrophys.*, **418**, 1141, 2004.
6. J. Crovisier *et al.*, *Astron. Astrophys.*, **418**, L35, 2004.
7. P. Ehrenfreund & S.B. Charnley, *Ann. Rev. Astron. Astrophys.*, **38**, 427, 2000.
8. J. Crovisier, *Faraday Discussions,* **109**, 437, 1998.
9. K. Altwegg, H. Balsiger & J. Geiss, *Astron. Astrophys.*, **290**, 318, 1994.
10. J. Geiss, K. Altwegg, H. Balsiger & S. Graf, in Ref. 3, p. 253.
11. V.A. Krasnopolsky, *Science,* **277**, 1488, 1997.
12. H.A.Weaver *et al.*, *Astrophys. J.*, **576,** L95, 2002.
13. R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T. C. Owen & F. Robert (Eds.), "Solar System History from Isotopic Signatures of Volatile Elements", SSS of ISSI, Vol. 16, Kluwer Academic Publ., Dordrecht, 2003, and *Space Sci. Rev.*, **106**, Nos. 1-4, 2003.
14. W. Benz, R. Kallenbach & G.W. Lugmair (Eds.), "From Dust to Terrestrial Planets", SSS of ISSI, Vol. 9, Kluwer Academic Publ., Dordrecht, 2000, and *Space Sci. Rev.*, **92**, Nos. 1-2, 2000.
15. F. Robert, D. Gautier & B. Dubrulle, in Ref. 14, p. 201.
16. L.M. Ziurys *et al.*, *Astrophys. J.*, **527**, L67, 1999.

17. D.C. Jewitt, H.E. Matthews, T.C. Owen & R. Meier, *Science,* **278**, 90, 1997.
18. C. Arpigny *et al*., *Science,* **301**, 1522, 2003.
19. H. Balsiger, K. Altwegg & J. Geiss, *J. Geophys. Res*., **100**, 5827, 1995.
20. P. Eberhardt, M. Reber, D. Krankowsky & R.R. Hodges, *Astron. Astrophys.,* **302**, 301, 1995.
21. K. Altwegg, "Sulfur in Comet Halley", Habilitationsschrift, University of Bern, 1996.
22. K. Altwegg, in: *Astrophys. Space Sci. Lib*., **311**, 257, 2004.
23. W.F. Huebner & J. Benkhoff, in: "Composition and Origin of Cometary Material", K. Altwegg, P. Ehrenfreund, J. Geiss & W.F. Huebner (Eds.), *Space Sci. Rev*., **90**, 117, 1999.
24. A.A. De Almeida, W.F. Huebner, J. Benkhoff, D.C. Boice & P.D. Singh, *Rev. Mex. Astron. Astrof.,* **4**, 110, 1996.
25. Comet Nucleus Team, "Heat and Gas Diffusion Models of Comet Nuclei". To be published, 2005.
26. H.U. Schmidt, R. Wegmann, W.F. Huebner & D.C. Boice, *Comp. Phys. Comm*., **49**, 17, 1988.
27. J. Klinger, *Science,* **209**, 271, 1980.
28. A. Kouchi, J.M. Greenberg, T. Yamamoto & T. Mukai, *Astrophys. J.,* **388**, L73, 1992.
29. O. Andersson & H. Suga, *Phys. Rev. B*, **50**, 6583, 1994.
30. A. Kouchi, T. Yamamoto, T. Kozasa, T. Kuroda & J.M. Greenberg, *Astron. Astrophys*., **290**, 1009, 1994.
31. J. Haruyama, T. Yamamoto, H. Mizutani & J.M. Greenberg, *J. Geophys. Res*., **98**, 15079, 1993.
32. G. Tancredi, H. Rickman & J.M. Greenberg, *Astron. Astrophys.,* **286**, 659, 1994.
33. T.I. Gombosi, D.L. De Zeeuw, R.M. Häberli & K.G. Powell, *J. Geophys. Res*., **101**, 15233, 1997.
34. T. Bagdonat & U. Motschmann, *J. Comp. Phys*., **183**, 470, 2002.
35. M.R. Combi, K. Kabin, D.L. De Zeeuw, T.I. Gombosi & K.G. Powell, *Earth, Moon, Planets*, **79**, 275, 1999.
36. J.H. Valk *et al*., *Astrophys. J*., **388**, 621, 1992.
37. M.E. Brown, A.H. Bouchez, H. Spinrad & C.M. Johns-Krull, *Astron. J*., **112**, 1197, 1996.
38. S. Wyckoff, R.S. Heyd & R. Fox, *Astrophys. J*., **512**, L73, 1999.
39. H. Zhang, G. Zhao & J.Y. Hu, *Astron. Astrophys*., **367**, 1049, 2001.
40. J. Clairemidi, P. Bréchignac, G. Moreels & D. Pautet, *Planet. Space Sci*., **52**, 761, 2004.

Chronology and Physical Evolution of Planet Mars

W.K. Hartmann[a], D. Winterhalter[b] and J. Geiss[c]

[a]*Planetary Science Institute, Tucson, Arizona, USA*
[b]*Jet Propulsion Laboratory, Pasadena, California, USA*
[c]*International Space Science Institute, Bern, Switzerland*

Introduction

About half the size of Earth, Mars has desert landscapes coloured red by the rusting of iron-bearing minerals, similar to those of America's southwestern deserts. As early as 1800, telescopes showed polar ice caps and shifting clouds, and the idea of a distant Earth-like planet began to emerge. Further studies showed streaky markings and dusky patches that darken and lighten during a Martian year. Astronomers realized that Mars is colder than Earth, since it is farther from the Sun, but many thought that the changing markings must be vegetation. By the 1880s, the Italian Giovanni Schiaparelli drew the streaks as an enigmatic network of narrow straight lines, which he called "canali." The nature of these streaks caused much discussion.

Around 1895, the colorful American astronomer Percival Lowell offered a new theory to explain the observations. Lowell started his argument by noting correctly that Mars itself had evolved. Any initial thick atmospheric gases, such as water vapour, would have slowly leaked off into space because of Mars's low gravity, causing the current air to be thin and dry. In spite of these losses, said Lowell, Mars was an abode for life. Not only was there vegetation, but also civilization! Much of the vegetation was cultivated crops. Lacking water, the Martians had built a system of canals to bring water from polar ice caps across the arctic plains to the warmer equator, where the Martians lived. The streaky markings, which Schiaparelli and Lowell both perceived as straight lines, were bands of vegetation – crops – being grown along the sides of the canals.

Lowell's theory electrified the public for decades – but it had the scientific drawback of being completely wrong. As early as the 1920s, astronomers accumulated evidence that the Martian air is even thinner and drier than Lowell thought, and that his straight-line canals were only streaky alignments of dusky patches.

The pendulum of Martian theory swung much farther from Lowell's scenario in 1965 when Mariner 4 returned the first close-up snapshots during a quick flyby. These few fuzzy images showed no cities or canals, but rather moon-like craters. Analysts concluded that Mars is merely a Moon-like world with a little air to blow the dust around.

The Early Space Missions – Volcanoes, Riverbeds, and Sterile Soil

The first-ever orbital mapping of the entire planet, by Mariner 9 in 1971-1972, revealed that the dusky markings and Lowellian canals were patches and streaks of dust, oriented according to prevailing winds. Mariner 9 found volcanoes, lava flows, sand dunes, layered sediments, canyons, polar ice, craters, fractures, clouds, and fogs. The highest volcano, named Olympus Mons (Mt. Olympus, after the mountain that supposedly housed the Greek gods), towers some 25 km above the mean Martian surface. Still more astounding, Mariner 9 found this dusty planet, with no known liquid water, to be laced with dry riverbeds.

Now the scientific pendulum swung back to questions that still dominate Martian science today. How could a frozen, desert planet once have had rivers? How long ago did the water flow and where did it go? Were earlier conditions more Earth-like? Could life have started there after all?

The first three successful landers on Mars shed some light on these issues. On 20 July 1976 – seven years to the day after Apollo 11 astronauts put the first human footprints on the Moon – Viking 1 became the first human-built device to land safely on Mars and send back data. Viking 2 followed a few weeks later, landing on a similar, but more northerly plain called Utopia. Pathfinder and its small rover, Sojourner, landed at the mouth of the Ares Vallis river channel in 1997. The mission of these landers was to take the first surface photos, analyze the Martian soil and rocks, and look for life. Pictures at all three sites showed beautiful but barren desert landscapes with scattered rocks and dunes. The soils turned out to be sterile, with no organic molecules (to an accuracy of a few parts per billion). Modest abundances of sulphates and salts in the soils suggested that they might have been exposed to salty water that evaporated. No compelling evidence for life was found. The soils did show some curious chemical reactions when exposed to nutrients, but chemists concluded that these reactions were caused not by life, but by unusual soil chemistry, due to exposure of the soil to the Sun's ultraviolet radiation. Pathfinder showed the rocks to be essentially basalts and other igneous rocks. The harsh surface conditions and lack of surface organics convinced most scientists by the late 1970s that no life currently exists on the Martian surface.

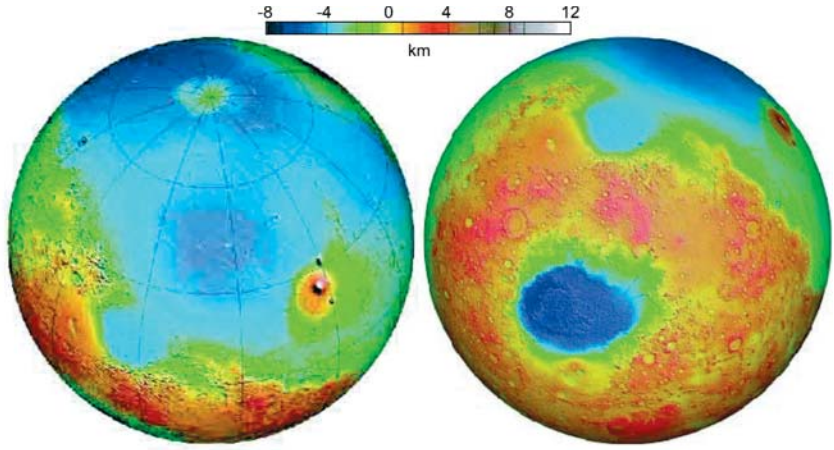


Figure 1. A global map of Mars compiled from the Mars Global Surveyor Laser Altimeter data (Courtesy NASA, JPL, MOLA Science Team). Visible is the dichotomy, with one hemisphere dominated by low plains and the other by higher, ancient cratered uplands. The colour scale depicts elevation referenced to some standard, with dark blue being the lowest elevations.

On a planetary scale, the most striking geological feature found by the early missions is the Martian crustal dichotomy[1] (Fig. 1). Its boundary can be roughly described as a great circle inclined ~35° to the equator[2]. Plains to the north of the boundary are typically younger and smoother than the ancient, heavily cratered uplands to the south. Recently, data from the Mars Global Surveyor (MGS) has shown the dramatic difference in elevation between the two hemispheres, as much as 6 km in some areas[3]. The crustal dichotomy is indicative of a significant event or series of events, early in the history of Mars, one that has had a fundamental influence on the subsequent evolution of the interior and surface. Determining the origin of the dichotomy is still a problem to be solved.

Chronologies of the Three Worlds: Earth, Moon and Mars

Martian observations indicate not only ancient, heavily cratered uplands, but also sparsely cratered plains, flows of young lavas, dry river channels in varying states of preservation, finely layered, nearly uncratered polar deposits, tectonic fractures cutting through sparsely cratered plains, and other such features. One important factor in understanding the development of the planet is to gain at least a crude understanding of not just the relative ages, but also the absolute ages of these various geological units. A time scale is necessary to answer fundamental questions such as: When was the most recent volcanism? How long ago did Martian rivers flow? Were there repeated major fluvial periods? How much water and other gases have escaped into space, how much underground ice is

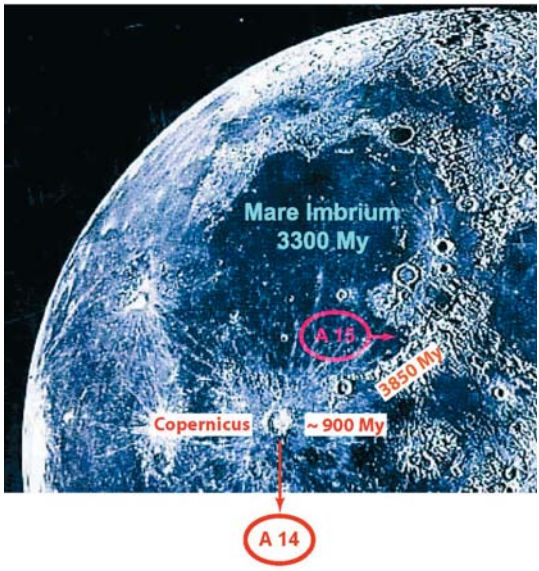


Figure 2. North-western region of the lunar front side with three important time markers: the date of the impact excavating the Imbrian Basin 3850 My (million years) ago and the age of the basalts in Mare Imbrium at 3300 My were obtained from Apollo 15 and 16 sample analyses[4,5]. Radiometric ages were obtained for the geological areas around all Apollo landing sites and also around the areas of Luna sample-return missions. An age for the crater Copernicus of about ~900 My was obtained from analyses of glass pieces collected not in the immediate vicinity of the crater, but 400 km away at the Apollo 14 landing site. The age of the glass pieces is correct, but their assignment to the Copernicus crater should be confirmed.

still there, and what is the relative importance of geothermal heating versus climatic change in the episodes when it was released as liquid water?

The history of terrestrial geology and lunar science shows that a reliable time scale is the backbone of the geosciences. In the 19th century, a worldwide relation was found between fossil species and sedimentary rock strata on Earth. This established a chronology of geologic epochs from the Cambrian to the present, forced recognition of a very ancient age for the Earth, and led to the concepts of geological and biological evolution. The discovery of radioactivity at the end of the 19th century paved the way for establishing Earth's absolute chronology, covering the Precambrian in addition to the classical epochs from the Cambrian to the Quaternary.

On Mars and the Moon, relative time scales are based not on fossils, but on stratigraphy and surface features, in particular on the density of impact craters. In the case of the Moon, the cratering chronology is calibrated with the radio-

metric ages of lunar samples collected and documented by the Apollo astronauts and Soviet sample-return probes, at known landing sites (Fig. 2). The best-calibrated epoch in lunar history is from ~4100 to ~3100 million years (My) ago, when the large lunar basins were excavated and then filled with mare basalts[4,5].

Chronology of Mars

For Mars, three main geological epochs have been defined by measuring densities of impact craters (number/km^2) from photography obtained by orbiting spacecraft[6]. In the absence of Martian samples with documented geographic origin, the Martian time scale is calibrated by comparison with the lunar time scale, taking into account theoretical estimates of the difference in cratering rates between the Moon and Mars[7]. At the ISSI workshop on "Chronology and Evolution of Mars"[4] in February 1999, existing methodology and models relevant to this absolute crater-based Martian chronology were discussed and compared[7-9], and a consensus was reached on a time scale for the Martian crater chronology. Figure 3 shows the assigned ages of the Martian geologic epochs as a function of numbers of craters larger than 1 km on units of various stratigraph-

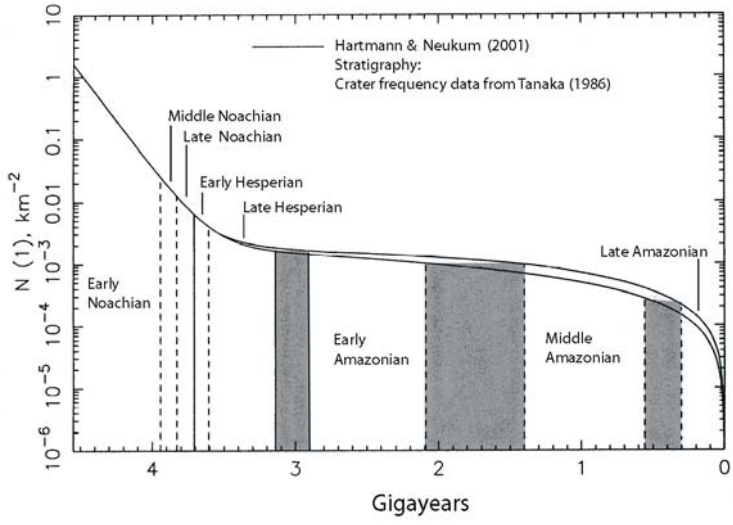


Figure 3. Mars cratering chronology model of Hartmann and Neukum[8]. N(1), the crater density of all craters larger than 1 km, is plotted against inferred age. The two curves are derived with slightly different assumptions. The diagram shows estimated ages of Martian geological epochs. Grey bands give the uncertainties due to the difference between the two models and emphasize that uncertainties are greatest in the middle part of Martian history, where the curve is relatively flat. The difficulty will be overcome by future space missions that will return samples from appropriate locations on Mars and the Moon.

ic age. The crater count system is much less accurate than radiometric dating, yet, significantly, certain Mariner 9 and Viking-era crater-count efforts revealed that at least some Martian lava plains formed as recently as the last few hundred million years[6], a conclusion controversial at the time, but later found to be consistent with radiometric ages of Martian meteorites[10] (see below).

The Martian Meteorites

By the early 20th century, the chemistry and mineralogy of a few rare stony meteorites indicated that they did not fit into the classification scheme used for all the other meteorites. Three classes of these rare stones were identified: the Shergottites, Nakhlites, and Chassignites (named after the place of fall of the first example in each class) or, in short, SNC-meteorites. When radiometric dating of meteorites began in the 1950s, most meteorites were dated at more than 4000 My, compatible with an asteroidal origin. A Shergottite, however, gave an age of only a few hundred million years. Such young rocks could hardly have formed by igneous processes in an asteroid, because asteroids cooled several billion years ago. Then in the 1980s, researchers recognized that the SNC rocks

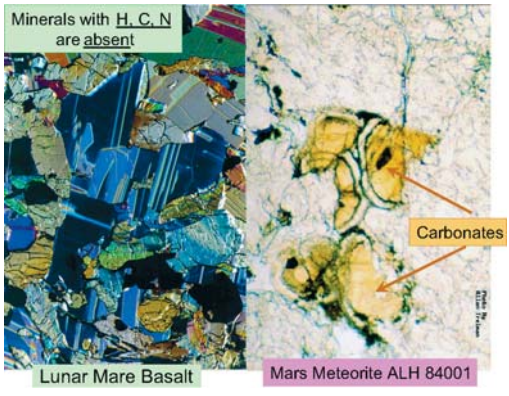


Figure 4. Comparison of lunar and Martian sample thin sections. *Left*: A lunar mare basalt lacking any evidence for volatiles (sample collected by Apollo 11 astronauts in Mare Tranquilitatis). *Right*: Martian igneous cumulate ALH 84001, containing carbonates left by evaporation of mineral-rich water. (Courtesy Beda Hofmann, Naturhistorisches Museum, Burgergemeinde Bern, Switzerland)

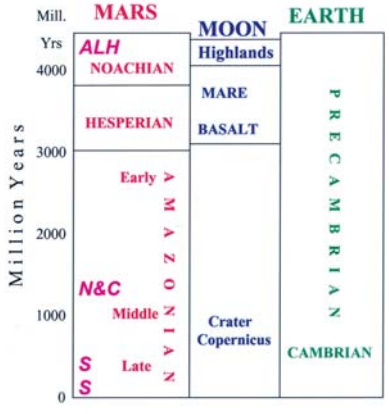


Figure 5. Comparison of the geologic time scales of the three best-studied terrestrial worlds. On Mars, three major epochs (Noachian, Hesperian and Amazonian) have been defined by Tanaka[6] using comparison of impact-crater densities. Radiometric ages have been determined for the known classes (S, N, C, ALH) of Mars meteorites.

included gas that exactly matched the Martian atmosphere, and that they had been blasted off Mars by asteroid impacts.

Today, the Martian origin of SNC meteorites is undisputed. About thirty of them are presently known. One (ALH 84001) is a very ancient igneous cumulate rock dating from 4500 My ago – apparently a piece of the original crust of Mars. All the others are relatively young metamorphic igneous rocks, dating from 170 My to 1300 My ago. Their launch from Mars implies that they resided within near-surface (<100 m depth?) coherent layers, according to current impact models[11]. The rocks confirm not only late magmatism on Mars, but also late aqueous activity, since they often exhibit signs of exposure to water, including deposits of salts and carbonates in fractures. Figure 4, for example, contrasts a thin section of an absolutely dry lunar basalt with a thin section of the 4500 My-old Martian cumulate, with prominent carbonates left by evaporation of mineral-rich water. The youngest Martian igneous meteorite, dated at 170 My, points to magmatic activity within the last few percent of Martian time.

As seen in Figure 5, these results allow us to begin interplanetary comparison between the chronologies of the three best-studied terrestrial worlds, namely Earth, Moon, and Mars. In the case of the Moon, an impact-cratered surface was modified by volcanism that virtually ended 3000 My ago (consistent with the Moon's small size and consequent rapid cooling). Earth has suffered continuing plate-tectonic activity, volcanism, and fluvial erosion that completely replaced the ancient surface with young materials. Mars, an intermediate case, has both extremes. The discovery of the 4500 My Martian crustal fragment ALH 84001, shown in Figure 4, is especially exciting because no rocks of Earth's original crust have survived plate-tectonic destruction processes, and the Moon's original magma ocean crust is buried under a kilometre or more of "mega-regolith" fragmental material. Thus, Mars appears to be the only planet with accessible exposures of intact original crust.

Future Improvements of the Martian Chronology

The time scale in Figure 3 has been in use as the standard Martian chronology since the ISSI workshop in 1999. A glance at this figure shows that measured crater densities give a relatively good time resolution for the oldest and the youngest epochs. It is, however, poor for ages from about 1000 My to about 3000 My. Various efforts would improve this situation:

1. Reduce the uncertainty in the lunar chronology: Part of the curve in Figure 3 depends on interpretations of the age of Copernicus from Apollo samples

picked up at large distance from Copernicus. Carefully selected lunar rock samples from large, young lunar craters would be extremely helpful in calibrating the curve in Figure 3. Sample return from the western regions of Oceanus Procellarum would also be desirable, because some lava flows in these regions have been interpreted as younger than 3000 My. They could help calibrate the diagram.

2. Determine the absolute ages of the middle and early Amazonian, and of the late Hesperian regions on Mars. This would involve sample-return missions landed on carefully chosen geologic units of well-determined stratigraphic age. Crater counts on those units would then directly calibrate Figure 3.

3. Identification of the source area for specific Martian meteorites might be possible. For example, the four Nakhlites and Chassignites analyzed so far have identical radiometric ages of 1300 My. Moreover, all four have ejection ages of 11 My[10], suggesting that they were ejected from Mars by one and the same impact. If the place of origin of these meteorites on the surface of Mars could be identified, it would allow calibration of Figure 3. Once the curve is calibrated, it will be possible to derive approximate absolute ages, within perhaps 20% uncertainty, by counting craters in any extended geologic unit on Mars.

Water on Mars

Earth's abundant water is believed to be the key to the origin of life on our planet. Life formed in the oceans or hot springs, according to currently dominant theories. If early Mars had rivers and volcanoes, did it once have lakes, seas, and/or hot springs, too? By the 1980s, researchers realized Martian water is hidden in three places: frozen water in the polar caps, water bound molecule-by-molecule in mineral crystals, and massive amounts of water frozen underground, as in Earth's arctic tundra. Early studies[12] showed that upper-latitude impact craters ejected not dry dust and rocks, but a muddy slurry, indicating penetration into ice-rich layers. This effect was used to map the depth to ice as a function of latitude[13], and also to investigate the possibility that ice flow has modified landforms and contributed to ancient "terrain softening" effects. Mars Odyssey, an orbiter that arrived in 2001, confirmed the existence of abundant underground ice, within a few feet of the surface, at high latitudes[14]. Thus, a revolutionary change in our view of Mars happened by the 1990s, indicating that while Mars has a dry and dusty surface, it is actually a wet planet in the sense of having lots of water – mostly hidden underground in solid form. This is a step towards explaining the ubiquitous river channels discovered on Mars by Mariner 9 in 1972.

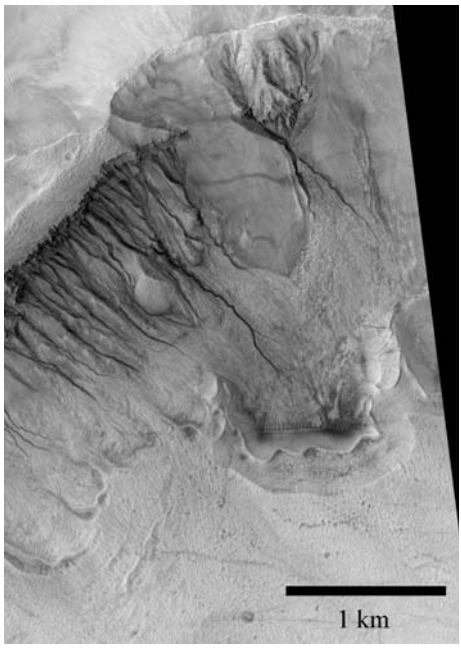


Figure 6. Example of a gullied hillside in a Martian crater at 39°S and 166°W. Gullied hillsides are common and generally on young surfaces lacking in impact craters, indicating a very youthful age within the last percent or so of Martian time. (Courtesy Daniel C. Berman, Planetary Science Intitute; Mars Global Surveyor image, NASA/JPL, Malin Space Science Systems)

There is geomorphic and chemical evidence for the shaping of many Martian land forms and geochemical weathering by water. Of particular interest, in terms of the planet's water history, are the indications that many Martian surface features have been exposed at one time or another to briny waters[15]. For example, aqueous-alteration products have been documented in the 1300 My-old Lafayette meteorite, where water exposure was dated as being within the last 700 My[16]. Thus, this rock was exposed to water not when it formed during igneous activity, but in an episode at least 600 My later. Furthermore, as seen in Figure 6, the walls of many mid- and upper-latitude Martian craters reveal very young, water-caused erosion[17], virtually identical in scale and morphology to gullies on basaltic hillsides of Iceland[18].

The problem with Martian surface water flow is how to maintain the liquid state once it starts to flow. The present-day Martian climate is sub-freezing most of the time, but temperatures occasionally rise above the freezing point. The atmosphere has a surface pressure that ranges from only 3 millibars in the high areas to as much as 10-15 millibars in the lowlands, compared to 1000 millibars at sea level on Earth. The 6 millibars is a "magic number": at high Martian elevations with pressure below 6 millibars, liquid water spontaneously bubbles away. In low-elevation regions with pressures above 6 millibars, it is more stable, but still evaporates very fast. Thus, water would seem to have a hard time flowing far under the present conditions, since it tends to freeze and dissipate at the same time. Paradoxically, however, a Martian river could extend its erosive lifetime by forming a frozen ice layer on the surface, shielding the water underneath from rapid evaporation. More importantly, high salt content helps. Sufficiently salty brines can stay fluid down to -40°C. Salty water is likely on Mars, because when water runs across soils or percolates through them, it dis-

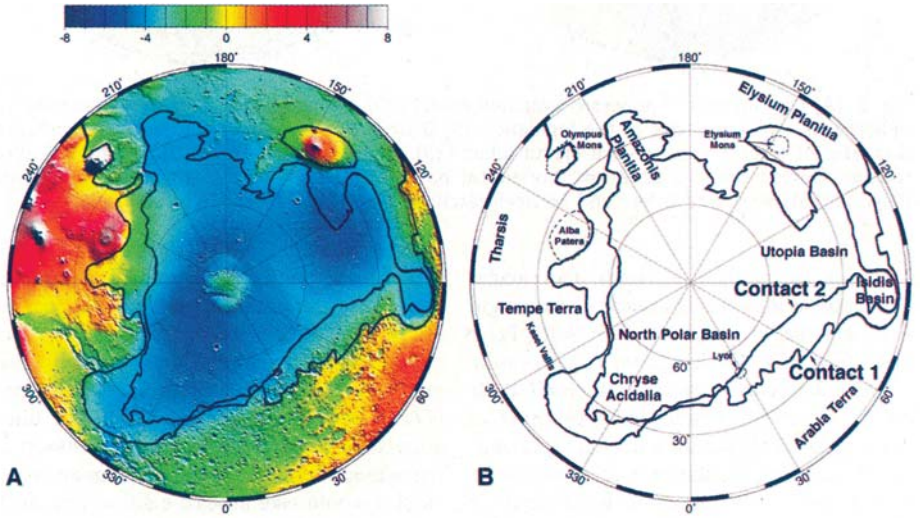


Figure 7. Polar view of Mars from Mars Global Surveyor. (Courtesy NASA/JPL, MOLA Science Team and J. W. Head et al.Science 286, 2134, 1999). The colour bar is given in kilometres. Dark blue shows the lowest present-day altitudes. The black solid lines labeled Contact 1 and Contact 2 have been interpreted to be shorelines[20]. The case for an ancient ocean shore is stronger for Contact 2, because it represents an equipotential line in good approximation[21], and because the area at the altitude below is remarkably smooth. Assuming that the present topography is approximately valid for the time when Contact 2 was formed, the volume contained by this shoreline would correspond to a global water layer of 100 metres thickness[21]

solves salts and other minerals. The mineral deposits in Martian meteorites (see Fig. 4) are thus important in confirming that Martian water was salty.

There is also the likelihood that the earliest Mars had an atmosphere sufficiently dense to maintain liquid water on the surface. Chemical and isotopic analysis of the Martian atmosphere indicates that hundreds of bars of carbon dioxide were probably emitted by Martian volcanoes. Mars Global Surveyor showed that Mars' protective global magnetic dynamo ceased about 4000 My ago (see below), allowing the solar wind unhindered access to the upper atmosphere. Rough calculations[19] show that the solar wind would erode a 1-bar atmosphere to the present levels in that time. Thus, current data strongly suggest that Mars has lost much of an initially much-thicker atmosphere.

An ultimate question about Martian water involves the maximum amount present on the surface at any one time. One view, held by some, is that there was enough to fill northern lowland basins and create an ocean. In the 1980s and 1990s, researchers mapped possible shoreline features and proposed ancient oceans[20]; later topographic mapping by the MOLA instrument on Mars Global

Figure 8. Landscape on Meridiani Planum taken by Opportunity Rover in 2004 (Credit: NASA/JPL/Cornell). Sulphate-rich rocks, hematite spherules in the soils, flat topography, weak layered rocks, and characteristics of the degraded craters visible from orbit all suggest that this region may be a large ancient lake bed. The plain is a depressed flat region with ripple relief of a few centimetres height. The rim of Endurance crater is on the horizon, at a distance of a few hundred metres.

Surveyor proved that some of these shoreline features lie at a uniform elevation, strengthening the idea that they may have marked the edge of an ancient sea[21]. Figure 7, for example, shows a polar view of Mars with dark blue marking the lowest topography and red and white the highest. The two heavy black "contacts" mark the suggested outline of a transient early north polar sea on Mars, according to this camp. Pervasive sediments, salts, and underground ice deposits may be relics of such ancient seas.

Full-fledged oceans remain controversial. Nonetheless, an exciting step came in 2004 when the Spirit and Opportunity rovers landed at two sites where researchers had deduced that water ponded in ancient times. The most provocative news was detection of fundamentally new Martian rock types at both sites – thinly bedded, apparent sediments containing up to 40% sulphates. At the Opportunity site, a remarkably flat area (see Fig. 8), previously suggested as a possible ancient lake bed[22], where unusual deposits of the iron-oxide mineral hematite had been seen from orbit, thinly bedded, sulphate-rich sediments were seen in every crater that punched through the surface. At the Spirit site, on the floor of Gusev crater, the floor itself was covered with broken igneous rocks, but the Spirit rover reached old hills protruding above the floor, and these also had layered, sulphate-rich rocks.

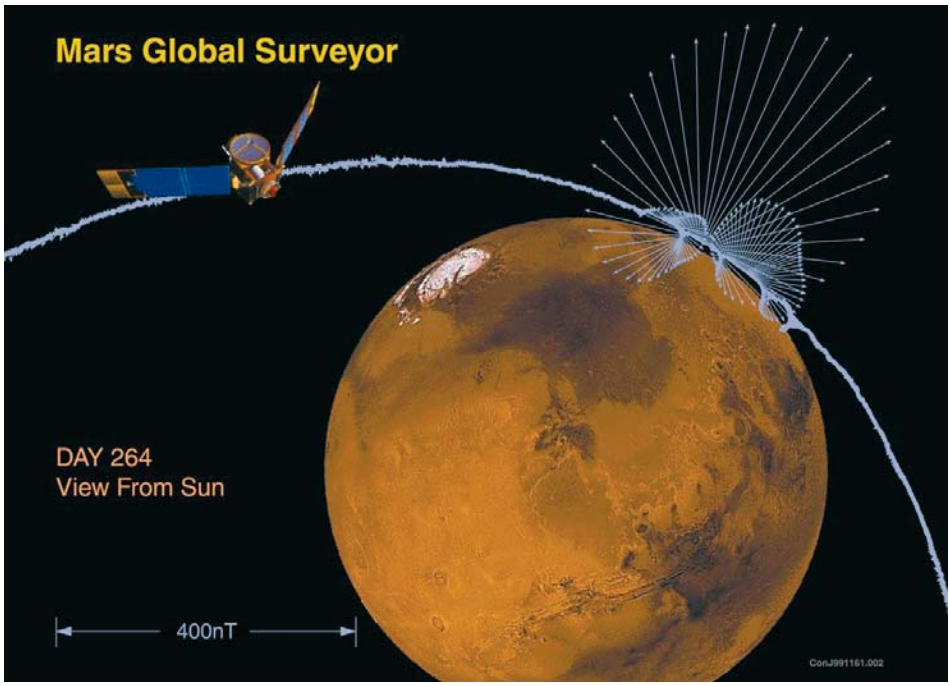


Figure 9. Projection of the MGS spacecraft trajectory and observed magnetic field (B) for MGS periapsis pass 6, 1997 day 264. The magnetic field is illustrated at 3 s intervals by a scaled vector projection of B originating from the spacecraft position at such times. Near periapsis, the small-magnitude solar-wind magnetic field (the closely spaced vectors appear as an undulating thick line along the trajectory) is replaced by the strong radial vectors from the surface sources[25].

Mars' Early Chronology – The Magnetic Connection

Given that the question of an internal magnetic field is of fundamental importance to the understanding of Mars' formation and thermal evolution, and of the evolution of Mars' atmosphere, surprisingly few spacecraft sent to Mars were equipped with instrumentation for such investigations. Of the nine or so orbiters that achieved Mars orbit, only two returned useful data about the magnetic field and about the plasma environment near Mars: Phobos 2, and particularly the more recent Mars Global Surveyor (MGS). The magnetometer data from these two spacecraft were integrated into a consistent picture of the magnetic environment of Mars at an ISSI workshop in October 2001, the results of which were published in the Space Science Series of ISSI[23].

The MGS magnetic field investigation provided the first unambiguous detection of the magnetic field associated with Mars[24] (Fig. 9). Measurements made early

in the mission confirmed that Mars does not currently possess a significant global magnetic field, yet, at the same time, detected strong, small-scale crustal magnetic sources associated with the ancient, heavily cratered terrain. This revealed that Mars had an internal active dynamo in its past that is now extinct[24].

Only a few crustal magnetic sources occur in the young, northern lowland plains (where the crust is thin). Most of them lie to the south of the dichotomy boundary in ancient, densely cratered southern highlands, and they extend to about 60 deg south of this boundary (see Fig. 10). No strong association between individual craters and magnetic sources has been found, and no magnetic sources were detected over Tharsis, Elysium, Valles Marineris, or any of the other major Martian volcanic edifices. There is an interesting correlation, however, between the location of the sources and the modification of the ancient crust by impact basins. The Hellas and Argyre basins, formed during bombardment in the Middle Noachian epoch (~3900 My ago), are devoid of magnetic sources. The absence of crustal magnetism in these basins and their surroundings probably implies that the Mars dynamo did not operate when these impact basins formed and locally disrupted the magnetized crust[24]. The reasoning is that had these structures formed in the presence of a large ambient field, they would likely have acquired intense remanent magnetization as the rock cooled below the Curie temperature – but none is observed. One conclusion would be that the dynamo had ceased operating just a few hundred million years after the formation of the planet.

An alternative interpretation would have the large impact basins form prior to the onset of the dynamo[26]. This interpretation stipulates that the magnetization of the southern highlands arose mainly from heating and cooling events that postdate the era of large impacts and basin formation. This approach is driven mainly by the notion that the early onset and cessation of the dynamo would be difficult to reconcile with a dynamo driven by solidification of an inner core (as is thought to be the case for Earth)[25].

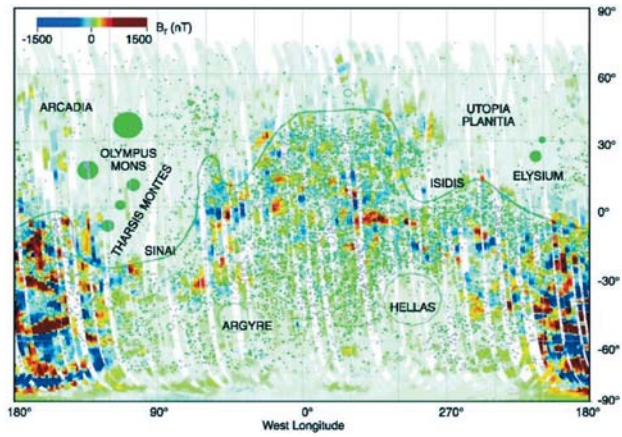


Figure 10. Map of magnetic disturbances on Mars, from Mars Global Surveyor data. Magnetic sources in the ancient, southern cratered highlands indicate magnetization of the primordial crust by a now-defunct global magnetic field (From Acuña *et al*. 1999[24]).

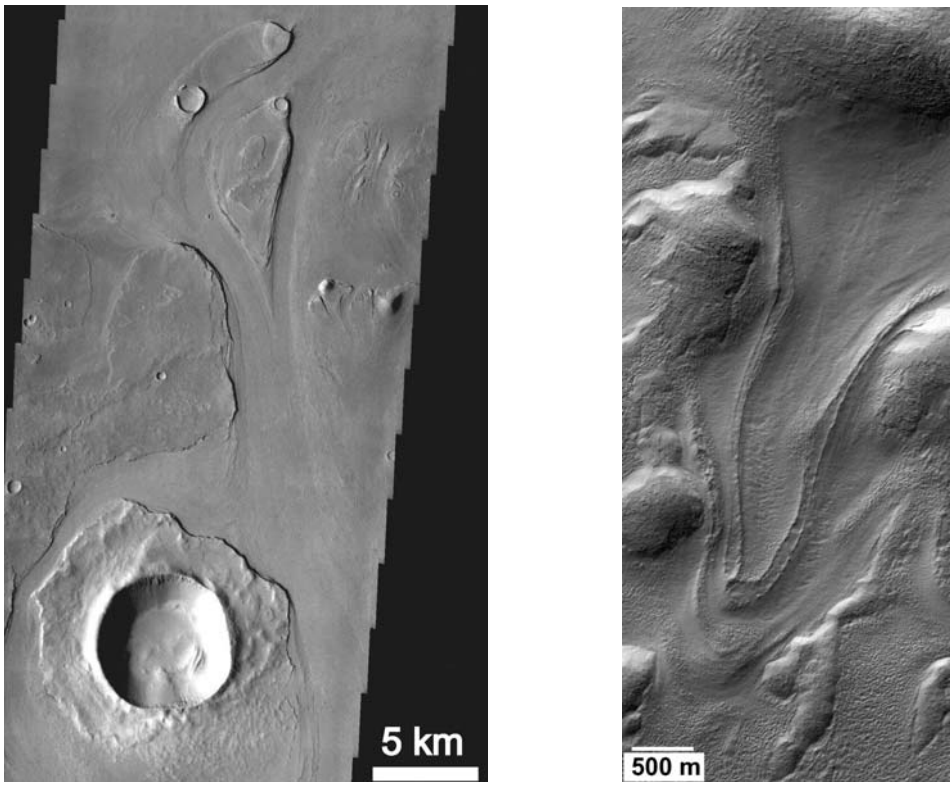


Figure 11. *Left:* River channel system of Marte Vallis, showing extensive erosion of young, sparsely cratered volcanic plain. Low crater densities indicate that this major channel system formed relatively recently in geologic time. *Right:* Example of tongue-shaped glacier-like flow features on the north inner wall of an unnamed Martian crater at 38°S and 247°W. Scarcity of impact craters on tongues and mantle-like valley fill implies recent formation. The formative process is not known, but might involve flow of ice-rich materials deposited during climatic episodes associated with obliquity excursions in the last few tens of Myrs (Mars Global Surveyor images: NASA, JPL, Malin Space Science Systems, Planetary Science Institute).

There is little doubt that a dynamo existed in the past and ceased to operate in early Martian history. But apparently there is as yet no unambiguous observational constraint on the exact timing of the dynamo action. Whatever the cause is, the dichotomy in magnetization of the Mars crust is perhaps the most significant clue in the Martian crustal evolution. That the crust of Mars is still intensely magnetized, and the fact that the magnetization displays a coherence over hundreds of kilometres, persisting for billions of years, requires an iron-rich crust with a magnetic mineralogy that can acquire and preserve, over aeons, a large remanent field. Together with inferences on the composition of the Mars mantle from studies of the SNC meteorites[27], this implies an increased oxidation

state relative to mantle-derived rock, consistent with assimilation of an aqueous component at crustal depths[25]. In recent years, many more MGS results on the crustal magnetism of Mars have been published and interpreted, providing increasingly important constraints on the evolution of the planet[28].

The discovery of an extinct dynamo on Mars should renew the interest in other weakly magnetized bodies in the Solar System. A partial magnetization of the lunar crust was discovered long ago, and an extinct dynamo was considered as a possible explanation[29]. Among the satellites of Jupiter, Ganymede is the most likely to have an internal magnetic field that could result from a dynamo, either extinct or even active to this day[30].

New Views of Mars

The two ISSI Martian volumes emphasize a once-radical view of Mars as a planet with a complex history of volcanic, fluvial, sedimentary, and magnetic activity, and some level of ongoing, active geological processes in recent geologic time. This view has been supported and still more radically extended by the most recent work, including the European Mars Express mission, and many recent reports refer to papers in the ISSI volumes.

Putting the various lines of evidence together, many researchers agree that most Martian riverbeds – the "outflow channels" and "valley networks" – formed long ago, when the Martian air was thicker and climatic conditions were different from those of today. Evidence from recent rovers suggests that during that time fluvial activity was so abundant that temporary lakes accumulated inside some craters. Connections, as yet poorly understood, may exist among phenomena such as the early largest impacts, convection patterns in the mantle, hemispheric asymmetries and the Tharsis dome, and the cessation of the magnetic field.

While most Martian river channels date from the first third of Martian time, according to the crater-count-based chronology, at least one of the Martian river channel systems, named Marte Vallis, is eroded into unusually young lava plains. This system is shown in Figure 11 (left), which emphasizes the paucity of craters. Apparently this river system formed within the last 200 My, possibly within the last 10 My – certainly within the last few percent of Martian history[31]. This is consistent with the meteorite evidence, cited earlier, for at least modest water mobility within the last few hundred million years. Water may be mobilized sporadically when geothermal activity melts underground ice.

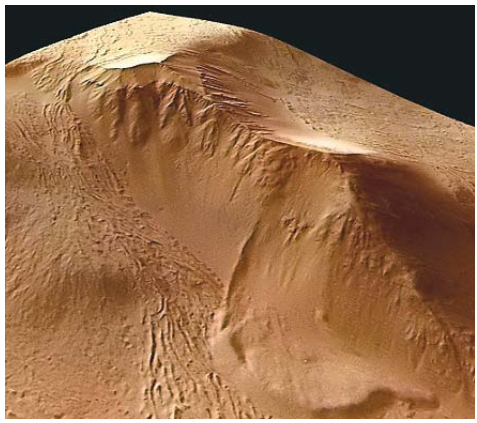


Figure 12. New imagery from the European Mars Express mission utilizes orbital stereo photography, allowing researchers to make 3D models and view terrain from different directions. This view shows part of a scarp on the giant volcano, Olympus Mons. Climate modelling by Francois Forget and French colleagues suggests massive ice deposition in this area during obliquity excursions, and some researchers believe that the slump-like feature off the wall (right) is a young glacial flow covering striated young lavas (left).
(Courtesy ESA/DLR/FU Berlin, G. Neukum).

Research in 2001-2004 revealed that not only the geology, but also the Martian climatic environment is dynamically active and changeable[32]. Mars today has an axial tilt of 25.2 degrees, close to the 23.5 degree value that causes Earth's seasons. (This tilt is measured relative to the planet's orbital plane.) While Earth's tilt remains nearly constant due to certain forces exerted by our Moon, Mars's tilt currently wanders from 0 degrees to values above 50 degrees, on a geologically "short" timescale of about 10 My. Costard and co-authors[33] and others used this to study insolation variations on gullied equator-facing versus pole-facing crater walls and to explain deposition of probable ice-rich mantles that seem to have been deposited at upper latitudes during recent climatic excursions[34].

Calculations of the obliquity during more remote epochs indicated values as high as 81 degrees, with the most common value being approximately 40 degrees[35]. This raises the stunning realization that the Mars we see today is not the Mars that has existed in the past. Today both poles have permanent ice caps of frozen water and semi-permanent caps of CO_2, but in the past, when the polar axis nodded toward the Sun as much as 40-45 deg, sunlight on the summer pole was intense enough to "burn off" all the ice. Water-vapour contents increased by factors as much as 50-100, enhancing deposition of frost and ice at low latitudes. Snow may have been much more common; spring thaws may have led to more water runoff than anyone believed possible a few years ago. Exemplifying the unexpected features are glacier-like tongues on the wall of a southern crater in a region northeast of Hellas[18], as shown in Figure 11 (right). The absence of craters on the slope (as with the Malin-Edgett gully features) suggests a very young age, consistent with ice deposition during recent obliquity cycles.

Mars Express has added stereo images of possible ice-flow features (Fig. 12), and this orbiter also confirmed local deposits of hydrated minerals such as gypsum (hydrated calcium sulphate). The new evidence opens the door to the ques-

tion of whether life might have started on our next-door planet. This is a perfect scientific question, because either answer is so profound. If we confirm that simple bacterial life started on Mars, even if it became extinct later, that will be the first time humans know that life started on another planet. On the other hand, if we search Mars and conclude that life never started, in spite of the water, then maybe we are more alone in the Universe than we thought. Either answer affects our understanding of our place in the cosmic realm.

References

1. W.K. Hartmann, *J. Geophys. Res.*, **78,** 4096, 1973.
2. T.A. Mutch & R.S. Saunders, *Space Sci. Rev.,* **19,** 3, 1976.
3. H. Frey, S.E. Sakimoto & J. Roark, *Geophys. Res. Lett.,* **25,** 4409, 1998; D.E. Smith & M.T. Zuber, *Geophys. Res. Lett.,* **25,** 4397, 1998; D.E. Smith *et al.*, *Science,* **279,** 1686, 1998; D.E. Smith *et al.*, *Science,* **284,** 1495, 1999.
4. R. Kallenbach, J. Geiss & W.K. Hartmann (Eds.), "Chronology and Evolution of Mars", Space Science Series of ISSI Vol. 12, Kluwer Academic Publ., Dordrecht, 2001, and *Space Sci. Rev.,* **96**, Nos. 1-4, 2001.
5. D. Stöffler & G. Ryder, in Ref. 4, p. 9.
6. K.L. Tanaka, *J. Geophys. Res. Supp.,* **91,** E139, 1986.
7. B.A. Ivanov, in Ref. 4, p. 87.
8. W.K. Hartmann & G. Neukum, in Ref. 4, p. 165.
9. W.K. Hartmann *et al.*, "Basaltic Volcanism on the Terrestrial Planets (Basaltic Volcanism Study Project)", p. 1050, 1981.
10. L.E. Nyquist *et al.*, in Ref. 4, p. 105.
11. H.J. Melosh, "Impact Cratering: A Geologic Process", Oxford Univ. Press, New York, 1989.
12. R.O. Kuzmin, "Lunar Planet. Sci. Conf.", Vol. XI, abstracts, Houston: Lunar and Planetary Institute, p. 585, 1980; S.W. Squyres, M.H. Carr, *Science,* **231,** 249, 1986.
13. S.W. Squyres, S.M. Clifford, R.O. Kuzmin, J.R. Zimbelman & F.M. Costard, in: H.H. Kieffer, B.M. Jakosky, C.W. Snyder & M.S. Matthews (Eds.), "Mars", The University of Arizona Press, Tucson 1992, p. 523.
14. W.V. Boynton *et al.*, *Science,* **297,** 81, 2002.
15. J.C. Bridges *et al.*, in Ref. 4, p. 365.
16. T.D. Swindle *et al.*, *Meteor. Planet. Sci.,* **35,** 107, 2000; C.-Y. Shih, L.E. Nyquist, Y. Reese & H. Wiesmann, "Lunar Planet. Sci. Conf.", Vol. XXIX, abstract 1145, Houston: Lunar and Planetary Institute, 1998.
17. M. Malin & K. Edgett, *Science,* **288,** 2330, 2000.
18. W.K. Hartmann, J. Anguita, M.A. de la Casa, D.C. Berman & E.V. Ryan, *Icarus*, **149**, 37, 2001; W.K. Hartmann, T. Thorsteinsson & F. Sigurdsson, *Icarus,* **162,** 259, 2003.
19. D.A. Brain, "6th Int. Conf. on Mars", Abstract 3241, Cal. Tech.: Pasadena, 2003.

20. T.J. Parker *et al*., *J. Geophys. Res*., **98,** 11061, 1993; D.E. Smith *et al*., *Science,* **284**, 1495, 1999J. W. Head *et al*., in Ref. 4, p. 263; M.H. Masson *et al*., in Ref. 4, p. 333.
21. J.W. Head *et al*., *Science,* **286**, 2134, 1999;
22. W.K. Hartmann, O. Popova & I. Nemtchinov, "Lunar Planet. Sci. Conf.", Vol. XXXIV, Abstract 1815, Houston: Lunar and Planetary Institute, 2001; W.K. Hartmann, J. Anguita, M.A. de la Casa, D.C. Berman & E.V. Ryan, *Icarus,* **149**, 37, 2001.
23. D. Winterhalter, M. Acuña & A. Zakharov (Eds.), "Mars Magnetism and its Interaction with the Solar Wind", Space Science Series of ISSI, Vol. 18, and *Space Science Reviews,* **111,** Nos. 1-2, 2004.
24. M.H. Acuña *et al*., *Science,* **284,** 790, 1999.
25. Figure 9 and conclusions reported here are from the article by J.E.P. Connerney, M.H. Acuña, N.F. Ness, T. Spohn & G. Schubert in Ref. 23, p. 1.
26. G. Schubert, C.T. Russell & W. B. Moore, *Nature,* **408**, 666, 2000.
27. M. Wadhwa, "Lunar Planet. Sci. Conf.", Vol. XXXI, abstract 1966, Houston: Lunar and Planetary Institute, 2000; H. McSween *et al*., 11th Annual V. Goldschmidt Conf., 3012, 2001.
28. J.E.P.Connerney, M.H. Acuña, N.F. Ness, D.L. Mitchell & R.P. Lin, Lunar and Planetary Science XXXV, Abstract 1114, Houston: Lunar and Planetary Science Institute, 2004; R.J. Lillis, D.L. Mitchell, R.P. Lin, J.E.P. Connerney & M.H. Acuna, *Geophys. Research Lett.,* **31**, L15702, 2004
29. S.K. Runcorn, *Nature,* **304,** 589, 1983.
30. M.G. Kivelson *et al*., *Nature,* **384,** 537, 1996; D.J. Southwood & M.G. Kivelson, *J. Geophys. Res.,* **106,** 6123, 2001.
31 D.C. Berman & W.K. Hartmann, *Icarus,* **159,** 1, 2002; D.M. Burr, J.A. Grier, A.S. McEwen & L.P. Keszthelyi, *Icarus,* **159,** 53, 2002.
32. J. Laskar, B. Levrard & J.F. Mustard, *Nature,* **419,** 375, 2002.
33. F. Costard, F. Forget, N. Mangold & J.P. Peulvast, *Science,* **295,** 110, 2001.
34. J.F. Mustard, C.D. Cooper & M.K. Rifkin, *Nature,* **412,** 411, 2001.
35. J. Laskar, M. Gastineau, F. Joutel, B. Levrard & P. Robutel, "Lunar Planet. Sci. Conf.", Vol. XXXV, Abstract 1600, Houston: Lunar and Planetary Institute, 2004.
36. W.K.H. acknowledges support from NASA grant NAG5-12217, and also from the staff of the Planetary Science Institute in preparing this paper.

The Search for Extrasolar Planets

S. Zucker and M. Mayor

Observatoire de Genève, Sauverny, Switzerland

During the recent decade, the question of the existence of planets orbiting stars other than our Sun has been answered unequivocally. About 150 extrasolar planets have been detected since 1995, and their properties are the subject of wide interest in the research community. Planet formation and evolution theories are adjusting to the constantly emerging data, and astronomers are seeking new ways to widen the sample and enrich the data about the known planets. In September 2002, ISSI organized a workshop focusing on the physics of "Planetary Systems and Planets in Systems"[1]. The present contribution is an attempt to give a broader overview of the researches in the field of exoplanets and results obtained in the decade after the discovery of the planet 51 Peg b.

The existence of planets orbiting other stars was speculated upon even in the 4th century BC, when Epicurus and Aristotle debated it using their early notions about our world. Epicurus claimed that the infinity of the Universe compelled the existence of other worlds. After the Copernican Revolution, Giordano Bruno wrote: *"Innumerable suns exist; innumerable earths revolve around these suns in a manner similar to the way the seven planets revolve around our Sun"*.

Aitken[2] examined the observational problem of detecting extrasolar planets. He showed that their detection, either directly or indirectly, lay beyond the technical horizon of his era. The basic difficulty in directly detecting planets lies in the brightness ratio between a typical planet and its host star, a ratio that can be as low as 10^{-8}. Using Jupiter as a typical example, we expect a planet to orbit at a distance of the order of 5 Astronomical Units (AU) from its host star. At a relatively small distance of 5 parsecs from Earth, this would mean an angular separation of 1 arcsecond. Therefore, with present technology, it is extremely difficult to directly image any extrasolar planet inside the overpowering glare of its host star.

The first way to get around this problem is to look for the minute motion of the host star, which is induced by the gravitational attraction of the planet. In principle, this motion could have been detected through astrometry, i.e. by monitoring the celestial position of the star. However, again using Jupiter as an exam-

ple, the apparent motion it induces on the Sun has a semi-major axis of 5 milliarcseconds, if viewed from a distance of 10 parsecs. Such a small astrometric signature is also extremely difficult, if not impossible, to detect.

This orbital motion of the host star can nevertheless be detected through the method of *radial velocities*. In this method, the observers exploit the Doppler effect in order to measure the radial component of the star's motion (i.e. its line-of-sight component). A periodic variation in the radial velocity may hint at an orbital motion and the possible existence of a planet near the star. The first claim of a very-low-mass companion detected using this method was the companion of the star HD114762. The amplitude of the radial-velocity variation was about 600 m s^{-1}, and the companion mass was found to be around 10 Jupiter masses[3,4] (M_J). Although the existence of this object is well-established, the question of its planetary nature is still debated. Alternatively, it could be a *brown dwarf* – an intermediary object between a planet and a star. The amplitude of the variation was barely detectable using the techniques available at the time. The detection of smaller planet candidates had to wait for the development of instruments that could measure *precise* radial velocities.

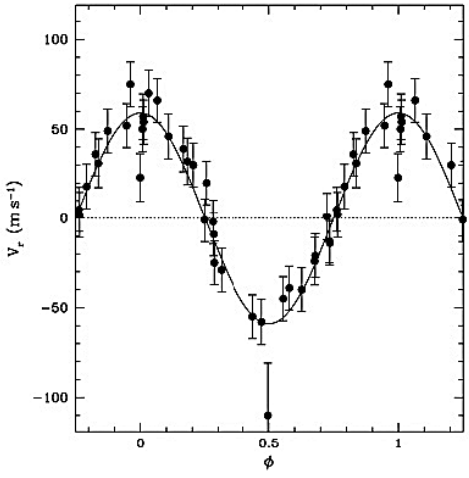


Figure 1. The "phase-folded" radial-velocity curve of 51 Peg. The period is 4.23 days and the planet mass implied by the amplitude is 0.44 Jupiter masses. The x-axis (ϕ) refers to the normalized phase (between 0 and 1) (original plot from the paper by Mayor & Queloz[5]).

The major breakthrough occurred in 1995, when Mayor & Queloz[5] announced the first discovery of an extrasolar planet orbiting the star 51 Pegasi, by using precise radial velocities (Fig. 1) obtained with the ELODIE spectrograph in the Haute-Provence Observatory. The planet has a mass of 0.44 M_J and an orbital period of 4.23 days, which means it orbits at a distance of 0.05 AU from its host star. Thus, this planet turns out to be very different from planets in our Solar System. All previous formation theories, which were based purely on the Solar System data, had to be re-examined, because the existence of a giant planet in such close proximity to a star was supposed to be impossible. 51 Peg b (the planet around 51 Peg), together with other 51 Peg-like planets (also nicknamed "Hot Jupiters"), formed the most serious challenge to the planet-formation theories, and gave rise to the notion of planetary migration.

Since 1995 precise radial-velocity measurements have been routinely performed by several groups, the most prominent ones being the California Group, using iodine-cell spectrographs[6], and the Geneva group, using fibre-fed spectrographs[7]. Those two methods were the major independent technological breakthroughs that enabled precise radial velocities to be measured.

Currently, the two techniques of radial-velocity and photometric-transit measurements (see below) have contributed most of the observational knowledge on extrasolar planets around main-sequence stars. However, there are also other techniques, which are worth mentioning here. Such is the case with the planets detected around pulsars[8,9]. In principle, they were detected by means of their radial velocities, but the radial velocities were not measured by the usual spectroscopic means, but by the precise timing of the pulsar's pulses. Pulsars are neutron stars, i.e. they are no longer main-sequence stars, but remnants of supernovae. Given the extremely violent process that forms pulsars, the existence of planets around them becomes an intriguing issue, which is completely different from the main quest after planets orbiting viable main-sequence stars.

Another important issue is the search for planets by using gravitational lensing. Due to a known phenomenon in General Relativity, when a faint relatively close star (the "lens") passes in front of a very distant star (the "source"), the light from the source undergoes a strong amplification. In case there is a planet orbiting the "lens" star, it may be detected by its effect on the amplification curve[10]. There are already several promising detections[11], but follow-up studies of the detected candidates are not feasible. This renders the main contribution by these surveys a statistical one, regarding the frequency of planets in the Galaxy.

Properties of the Extrasolar Planets

Currently, about 150 extrasolar planets have been detected. This number, although not overwhelming, is enough to make some preliminary observations regarding the characteristics of this growing population. Obviously, these findings have affected the development of theories concerning the formation and evolution of planets in general, and the Solar System in particular. In the following paragraphs we discuss the most prominent features of the population of extrasolar planets.

Mass distribution

Since the very early days of the search for extrasolar planets, a central research theme was the definition of planets, especially a definition that distinguishes them from stars. The most straightforward criterion, which remains the most common-

ly used one, is simply the object mass. The hydrogen-burning borderline between stars and substellar objects, at 0.08 solar masses ($M_{\odot}$), is already well known and understood. A similar limit was sought that would apply for planets. This was set at the so-called "Deuterium burning limit", at 13 M_J[12]. This arbitrary limit has nothing to do with upper mass limit of objects formed by agglomeration in accretion disks. The tail of the mass distribution of very-low-mass companions suggests that "planets" could exist with masses as large as 17 or 20 M_J. The question remains, whether the mass distribution of the detected planets indeed shows two distinct populations of planets and stars. The evidence is mounting that this is indeed the case. As several researches have shown[13,14], it seems that the mass regime between 20 M_J and 0.08 $M_{\odot}$ is under populated (Fig. 2). Thus, the two populations on both sides of this "brown-dwarf desert"[15,16] may rightfully be considered two physically distinct populations.

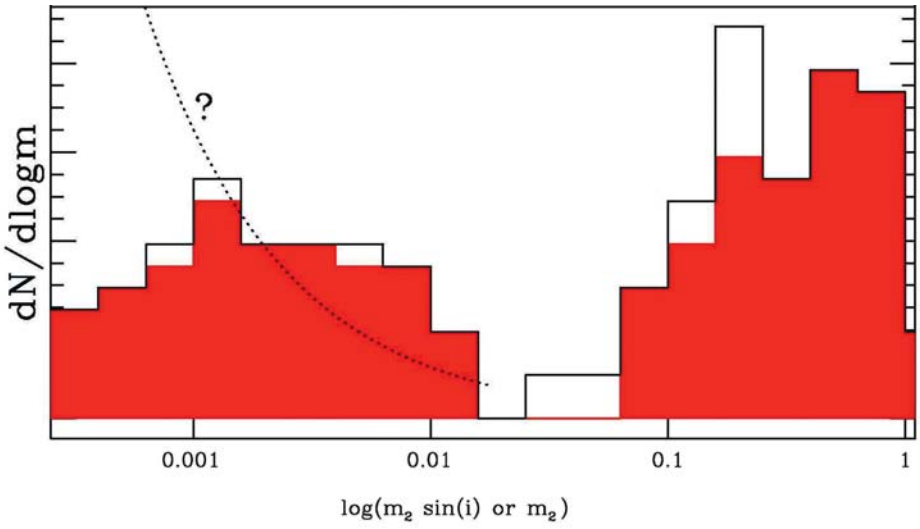


Figure 2. The mass distribution of companions to solar-type stars. The empty bars represent known planetary companion masses (m_2), while the filled bars represent minimal masses ($m_2 \sin i$) when the inclination is unknown. The dotted line represents an attempt to correct for an observational selection effect that is biased against low companion masses. Note the absence of planets at masses between 0.02 and 0.08 solar masses – "the Brown-Dwarf Desert" (from Udry & Mayor[17]).

51 Peg b, "Hot Jupiters" and migration

As already mentioned in the previous section, the detection of 51 Peg b was a major surprise. The paradigm concerning the evolution of planets and planetary systems can be traced back to works by P.S. Laplace and I. Kant in the 18th century. According to the modern form of this paradigm, the Solar System was

formed by a rotating disk of gas, dust and rocks, where the solids agglomerated into protoplanets and eventually into planets. Beyond a certain distance from the Sun (the "snow line"), water ice (and other similar "icy" materials) existed in abundance. Their existence allowed the formation of much larger protoplanets – large enough to rapidly accrete very massive atmospheres, and develop into the Solar System giant, Jovian planets. This scenario explained the segregation we see in the Solar System between terrestrial planets inwards from the "snow line", and Jovian planets beyond this line.

The detection of 51 Peg b and all the other "Hot Jupiters" contradicted the "snow line" paradigm. It seems that giant planets ("Jupiters") can be found very close to their host stars, much closer than the "snow line". The currently accepted way to solve this contradiction is by introducing the *migration* mechanism. Planetary migration is the hypothetical process by which the planet, still embedded in the protoplanetary disk, loses angular momentum through tidal interactions between the planet, the disk and the star[18-20]. The details of this mechanism are quite elaborate, and are still studied, but it is already certain that it can bring a planet that was formed outside the "snow line" into close proximity with the host star, thus explaining the existence of "Hot Jupiters".

Mass-period distribution

The existence of "Hot Jupiters" implied the possible presence of massive planets very close to their host stars. According to Kepler's laws, close orbital distances imply short periods. We thus expect to find very massive planets with very short periods. Since the radial-velocity technique is mostly sensitive to short periods and massive planets, we expect a very dense population in this area of the mass-period diagram. However, this is not the case, as has been shown by Zucker & Mazeh[21], Udry *et al*.[22] and Pätzold & Rauer[23]. They proved that there is a statistically significant dearth of *very* massive planets in very short periods. This phenomenon can serve as a hint that the migration process is maybe less effective for very massive planets[24,25], or maybe that at very short distances the planet "spills over" part of its mass into the central star[26].

Orbital eccentricities

The second extrasolar planet, 70 Virginis b, was detected by Marcy & Butler[27] using the Iodine-cell spectrograph in the Lick Observatory. 70 Vir b turned out to have a considerable orbital eccentricity of 0.4. The eccentricity record is currently held by HD80606 b, with a value of 0.93[28]. Such high orbital eccentricities were another challenge facing conventional theories about planets. The matter in protoplanetary disks was assumed to orbit the central star in circular Keplerian orbits, and the planets were supposed to reflect this primordial behaviour of the disk by having relatively circular orbits, similar to those in the

Solar System. In order to account for the observed high eccentricities, several models were suggested, and it seems that no single model alone can explain all cases. Thus, models to produce highly eccentric orbits evoked interactions with distant stellar companions[29,30], interactions in multi-planetary systems[31,32], or tidal interactions with the protoplanetary disk itself[33].

Stellar metallicities

The chemical composition of the planet-hosting star is intimately related to that of the primordial molecular cloud where the star was born. The common astrophysical index of *metallicity* is a measure of the relative amount of heavy elements (i.e. heavier than hydrogen and helium) in the stellar atmosphere. Santos *et al*.[34] have shown that the frequency of planets is strongly related to the metallicity of the host stars (Fig. 3). This can plausibly be explained by the need to have enough solid material and ices in order to form planets in the protoplanetary disk. The exact details of the influence of metallicity on the planet formation process are still obscure, but the existence of this influence is already well-established.

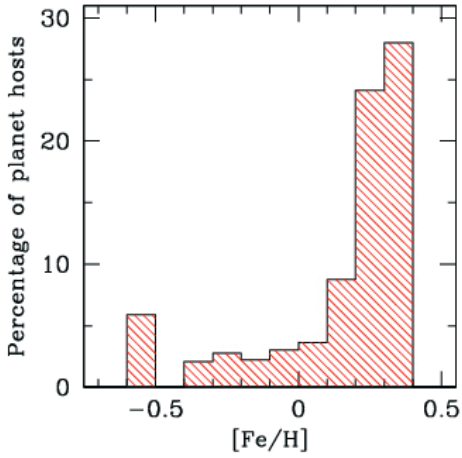


Figure 3. Percentage of planet-hosting stars found amid the stars in the CORALIE sample, as a function of stellar metallicity. [Fe/H] is a logarithmic measure of the heavy-element abundance, where a value of 0 corresponds to the solar value (Figure taken from Santos *et al*.[34]).

Planetary systems and resonances

The Solar System is the home of nine planets. The existence of more than one planet around a host star is easily explained by the protoplanetary disk paradigm. We thus expect extrasolar planets to appear in multiple planet systems as well[35]. The first detection of an extrasolar planetary system was a system of three planets orbiting the star υ Andromedae[36]. The most recent example is the detection of a third planet in the planetary system around the star μ Arae (HD160691) by Santos *et al*.[37], using the HARPS spectrograph at the 3.6 m telescope at La Silla. Interestingly, this planet is also the smallest detected to date, with a sub-Neptunian mass of 14 Earth masses (see Fig. 4).

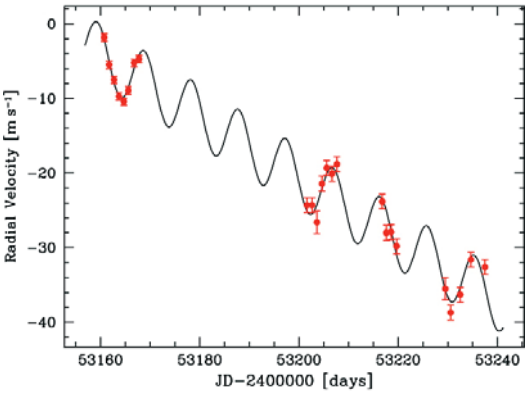


Figure 4. Radial-velocity curve of μ Arae. The time is measured in Julian days – a common astronomical time unit that corresponds to one day. The line represents the best fit to the data, obtained with the sum of a periodic Keplerian orbit and a linear trend. The linear trend represents the effect of the outer long-period companions in the system (from Santos *et al*.[36]).

In several cases, a couple of planets are arranged in a configuration where the two orbital periods form a mean-motion *resonance*. Such is the case, for example, in the planetary system around GJ876, where the two periods (30.1 and 60.0 days) form a 2:1 resonance[38] (Fig. 5). The Solar System exhibits many examples of such resonances, e.g. the Neptune/Pluto 3:2 resonance. The resonance phenomenon serves as a kind of laboratory for testing theories of orbital stability and evolution[39,40].

Multiplanetary systems clearly impose strong constraints on scenarios for the formation of planetary systems. Planets that orbit one of the components of binary stars could provide additional constraints. Currently, about twenty planets have been shown to exist in orbit around components of the binaries. The orbital characteristics of these planets seem to differ from those orbiting single stars[14]. ISSI workshop "Planetary Systems and Planets in Binaries" held in 2002, was dedicated to studying these very issues[1].

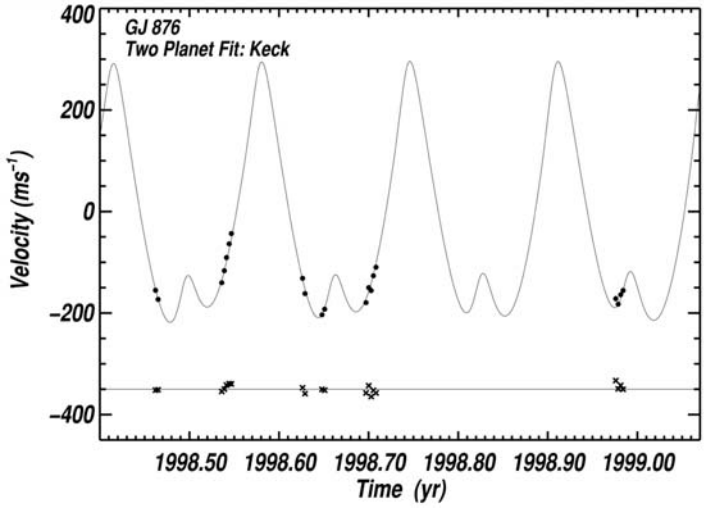


Figure 5. Radial-velocity measurements of GJ 876 as a function of time. The residuals to the fit are presented in the lower part of the plot (from Marcy *et al*.[38]).

Extrasolar Planetary Transits

Detecting planets through radial-velocity variation has one basic drawback. Since the star motion is recorded only along one spatial axis, the line of sight, some ambiguity remains when determining the planet dynamical characteristics. The ambiguity is between the planet mass and the so-called "orbital inclination", i.e. the angle between the line of sight and the orbital angular momentum vector. Instead of deriving both these quantities from the orbit, only $M \sin i$ is observable, where M is the planet mass and i is the inclination. It can easily be shown that values of sin i that are closer to unity are much more probable. Therefore it is usually assumed that $M \sin i$ is reasonably close to the true planet mass. Nevertheless, in order to fully characterize the planet, additional information is needed.

The need for additional information about extrasolar planetary orbits pushed astronomers to look for more ways to detect and study them, other than radial velocities. This led to the first discoveries of planetary *transits*. The phenomenon is well-known in the context of the Solar System, where Venus and Mercury may cross the line of sight between the Earth and the Sun. The recent Venus transit occurred on 8 June 2004, and attracted considerable public attention and media coverage (Fig. 6). Obviously, extrasolar planetary transits cannot be observed with the same detail. The only observable effect is a periodic weak dimming of the starlight, due to obscuration of part of its surface by the planet. Obviously, the probability of having a planet in such a geometric configuration that would allow planetary transits is not very high. For close-in planets, with periods of a few days (like "Hot Jupiters"), this probability is about 10%. For wider orbits this probability reduces considerably[10].

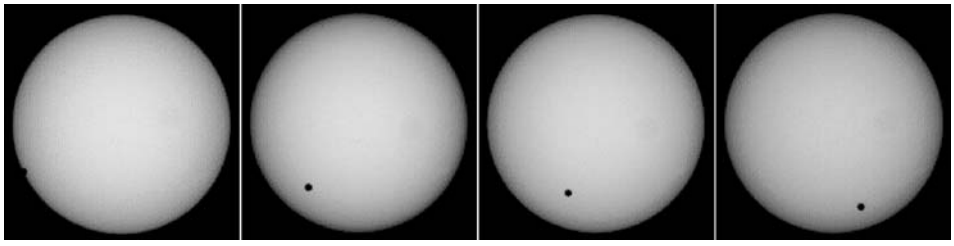


Figure 6. Images of the latest Venus transit on June 2004, showing the transit's various phases. The images were taken by the astronomers of the Rimavskej Sobote Observatory, Slovakia.

HD209458 b

The first extrasolar planetary transits that were discovered were those of HD209458 b. Mazeh *et al*.[41] first discovered the planet in the traditional radial-

velocity technique. Soon after the radial-velocity detection, Charbonneau *et al*.[42] and Henry *et al*.[43] detected the periodical dimming of the light, with exactly the same period as the radial-velocity variations, of 3.52 days (Fig. 7a). The two teams detected the transits using small and relatively cheap telescopes, which demonstrated that achieving the required photometric precision by ground-based observations was realistic.

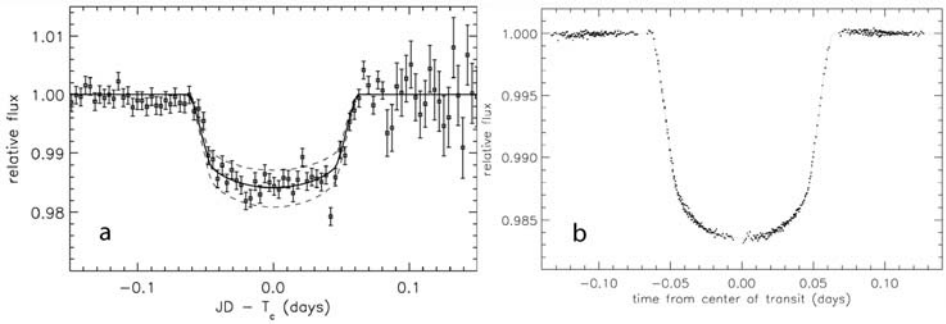


Figure 7. Measured transit light curves of HD209458. The left panel (a) shows ground-based measurements by Charbonneau *et al*.[42]. The right panel (b) shows the very precise light curve obtained by the Hubble Space Telescope. This panel is taken from the paper by Brown *et al*.[44].

HD209458 b, being the first transiting extrasolar planet, demonstrates the scientific potential of transits. Obviously, the mere fact that transits occur constrains the orbital inclination, thus solving the $M\sin i$ ambiguity. On top of that, the transit depth (i.e. the amount by which the stellar light dims during the transit) is closely related to the radius of the planet. Thus, it adds information that cannot be acquired from radial-velocity measurements[10]. Knowing the mass and the radius of the planet, we can even derive its density and surface gravity!

Space-based observations

The importance of the data derivable from the transit shape soon led the researchers to utilize the superb photometric precision of the Hubble Space Telescope (HST)[44]. The result was the very precise light curve shown in Figure 7b. It gave rise to a very precise derivation of the planet radius – 1.35 ± 0.06 Jupiter radii (R_J). It was rather surprising to find out that the planet was more extended than Jupiter, while its mass was about 0.69 M_J. It seemed to be inflated, probably because of its proximity to the central star[45].

Besides allowing a more accurate light curve, HST also provided a new and exciting kind of measurement, which constituted the first *direct* evidence about the planet. Charbonneau *et al*.[46] performed spectrophotometric observations of the transit in a bandpass centred at a sodium absorption feature, and in other

bands. The observations showed that the transit was significantly deeper in the sodium-feature bandpass than in the other bands. The difference was attributed to absorption by the planetary atmosphere, through which part of the stellar light had to pass during the transit. Vidal-Madjar *et al*.[47] performed similar observations, meant to detect other chemical species. They found that oxygen and carbon were present in the extended upper atmosphere of the planet, probably in an escaping state.

In 2004, a network of small ground-based telescopes detected planetary transits in the light curve of a fainter star (11th magnitude), which was then called TrES-1[48]. On the very last days before submitting the final version of this manuscript, Charbonneau *et al*.[49] reported another breakthrough. Using the Infrared Space Telescope 'Spitzer', they have measured the secondary eclipse of TrES-1, i.e. the dimming of the infrared luminosity when the planet itself is occulted by the star. This measurement allows, for the first time, a direct estimation of the temperature of a "Hot Jupiter". This exciting result also demonstrates the unique possibilities offered by space observatories in the domain of extrasolar planets.

OGLE observations

After the first detection of transits, numerous surveys to look for others were initiated by several groups. One such survey is the Optical Gravitational Lensing Experiment (OGLE), a project that monitors dense stellar fields in search of another exotic effect – gravitational lenses[50]. After the detection of HD209458 b, the OGLE team decided to dedicate part of their observing time for planetary-transit searches. Eventually, more than 130 transit candidates were detected in this framework[51-54]. However, a transit-like light curve alone is not sufficient for establishing the existence of a planet, since it can be produced by a faint stellar companion rather than a planet. In order to prove the planetary nature of the eclipsing object, the radial-velocity information is needed after all. The OGLE project targets relatively distant stars, which renders precise radial velocities much more difficult to obtain than in the relatively close stars of the radial-velocity surveys. Nevertheless, four OGLE candidates were already proven to be planets using radial velocities[55-58].

Currently, already six transiting planets are known, four of them from the OGLE project. The data about planetary radii and masses is thus enough for producing a preliminary mass-radius diagram (Fig. 8). The diagram demonstrates the consistency of the theories about planets with the available data. HD209458 b is seen to be exceptionally inflated relative to the other transiting planets[58].

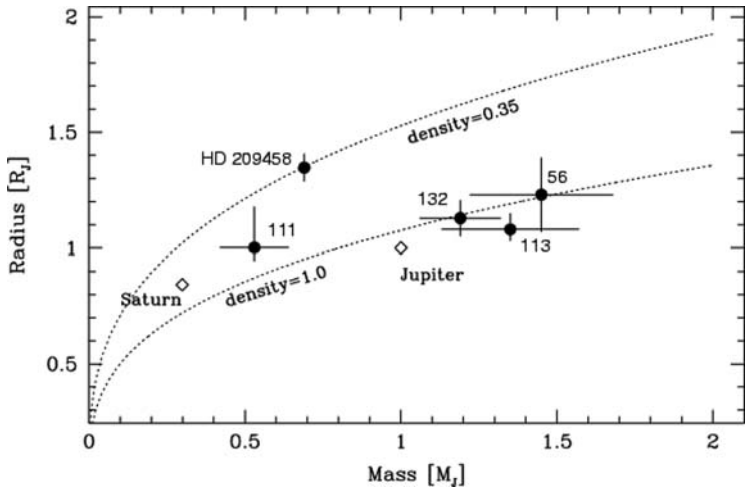


Figure 8. The mass-radius relation for the six currently known transiting extrasolar planets. Jupiter and Saturn are indicated for comparison. The two dotted lines represent lines of constant densities (from Pont *et al*.[58]).

Planned Space Missions

By now it is quite safe to assume that ground-based radial-velocity observations are suitable for detecting Jupiter-sized and even Saturn-sized planets orbiting the nearby stars, at distances of tens of parsecs. Figure 9 depicts the detectable parts of the mass-separation diagram by the radial-velocity method. The very precise spectrographs needed for precise radial velocities are probably still too difficult to maintain and manipulate onboard spacecraft. Jupiter-sized transiting planets cause dimmings of the stellar light of about 1%, which also has already been proven to be detectable by ground telescopes. In order to explore the other realms of the parameter space, dedicated space missions are planned.

Photometric missions

Two space missions designed for detecting planetary transits are expected to be launched in the near future. Free from "seeing" problems caused by turbulences in the Earth's atmosphere, those space telescopes are expected to monitor dense stellar fields in search of planets. Those missions are the French-led CoRoT satellite, expected to be launched in 2006[59], and the American Kepler mission, currently scheduled for 2007[60]. Besides providing many more detections of Jupiter-sized planets, they are also expected to detect Earth-sized planet candidates. Our experience from ground-based transit detections shows that the main challenge ahead is to provide radial-velocity confirmations for the planetary

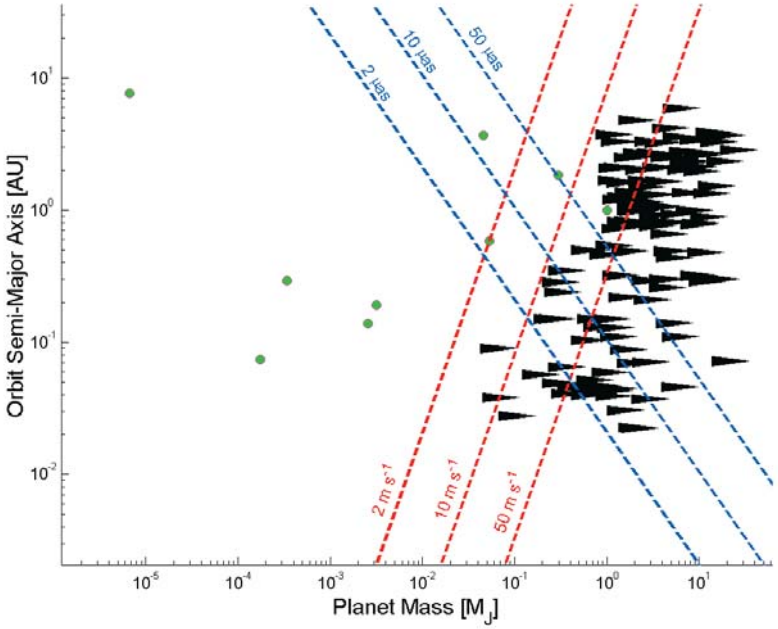


Figure 9. Detection limits of radial velocities and astrometry. The triangles represent the masses and separations of the known extrasolar planets. The triangles are elongated to account for the inclination ambiguity. The circles represent the masses and orbits of the Solar System planets. The ascending lines represent the detection limits of the radial-velocity method, with limiting amplitudes of 2, 10 and 50 m s^{-1}. The descending lines represent the expected limits of astrometry, for a star at a distance of 10 parsecs. The assumed astrometric accuracies are 2, 10 and 50 microarcseconds.

nature of the detected candidates, which is especially difficult for the smaller planets. A technological precursor for those two missions was the Canadian MOST satellite, designed for asteroseismology studies but requiring the same kind of photometric precision[61]. MOST is already yielding valuable measurements, and serves as a kind of proof-of-concept for CoRoT and Kepler[62].

Astrometric missions

As was previously explained, planets detected by radial velocities are prone to the $M\sin i$ ambiguity, which is solved if the planet transits the star. This happens only rarely, especially for planets that orbit the star in a wide orbit. However, in those cases the motion performed by the star may be large enough to be recorded as a trajectory in the two-dimensional plane of the sky. Measurements designed to detect this motion belong at the astronomical field of *astrometry*. In the early 1990s, a dedicated astrometric mission, called Hipparcos, was launched

by ESA[63]. Hipparcos, operating for about 3 years, had an astrometric precision of the order of milli-arcseconds. This was not enough to detect the motion caused by any known extrasolar planet. However, the fact that this motion was *not* detected served to put upper limits on the planetary masses, thus proving the sub-stellar nature of some of them[64-66]. It became clear that a much finer precision is needed in order to be able to detect the stellar orbits caused by the presence of planets. Two astrometric space missions are currently planned to reach the required astrometric precision: the American Space Interferometry Mission (SIM) (expected launch 2009)[67] and the European Gaia mission (expected launch 2010)[68]. Those two missions are designed to monitor the astrometric position and movement of tremendously large numbers of objects, with precisions of the order of micro-arcseconds. Thus, they will also detect the motions caused by planets orbiting those stars, this time without the notorious $M\sin i$ ambiguity. Figure 9 shows the expected detection capability of the astrometric space missions.

Darwin and TPF

Both NASA and ESA are contemplating the launch of similar extremely ambitious space missions, which are expected to be able to directly image extrasolar Earth-sized planets. Those missions are ESA's Darwin[69] and NASA's Terrestrial Planet Finder (TPF)[70]. The idea is to use an array of several medium-sized telescopes as an interferometric array, where the light from all the telescopes will be combined, interference patterns will be studied, and the effective size of the telescope will be much larger, in several respects. This will enable extremely precise astrometric measurements, but also very-high-resolution imaging, allowing the separation of the light we intercept from very close objects, previously unseparable. The expected scientific yield can be very exciting. Examination of the light reaching us from the planets themselves will unprecedentedly broaden our understanding of their physics and chemistry.

Darwin and TPF may very well be the closest approach ever towards understanding the origin and evolution of our Solar System and maybe even of life itself. In fact, the subject of trying to remotely detect and characterize life on extrasolar planets is also emerging as a seriously studied subject, where interesting and innovative ideas are proposed. The discipline of astrobiology is gradually developing, preparing for the long-awaited moment when it becomes practical, and a dedicated ISSI team has already been assembled[71]. Undoubtedly major scientific and philosophical issues will surface, in ways we can now only imagine.

References

1. S. Udry, W. Benz & R. von Steiger (Eds.), "Planetary Systems and Planets in Systems", Space Science Series of ISSI Vol. 19, Kluwer Academic Publ. Dordrecht, and *Space Sci. Rev.*, in press, 2005.
2. R.G. Aitken, *Astron. Soc. Pac. Leaflet,* **112**, 98, 1938.
3. D.W. Latham, T. Mazeh, R.P. Stefanik, M. Mayor & G. Burki, *Nature,* **339**, 38, 1989.
4. T. Mazeh, D.W. Latham & R.P. Stefanik, *Astrophys. J.,* **466**, 415, 1996.
5. M. Mayor & D. Queloz, *Nature,* **378**, 355, 1995.
6. A. Cumming, G.W. Marcy & R.P. Butler, *Astrophys. J.,* **526**, 890, 1999.
7. A. Baranne *et al., Astron. Astrophys. Suppl. Ser.,* **119**, 373, 1996.
8. A. Wolszczan & D.A. Frail, *Nature*, **355**, 145, 1992.
9. A. Wolszczan, *Science*, **264**, 538, 1994.
10. P. Sackett, in: "Planets Outside the Solar System: Theory and Observations", NATO ASI 189, J.-M. Mariotti & D. Alloin (Eds.), Kluwer Academic Publ., Dordrecht, p. 189, 1999.
11. I.A. Bond *et al., Astrophys. J. Lett.,* **606**, L155, 2004.
12. A. Burrows *et al., Astrophys. J.,* **491**, 856, 1997.
13. A. Jorissen, M. Mayor & S. Udry, *Astron. Astrophys.,* **379**, 992, 2001.
14. S. Zucker & T. Mazeh, *Astrophys. J.,* **562,** 1038, 2001.
15. G.W. Marcy & R.P. Butler, *Publ. Astron. Soc. Pac.,* **112**, 137, 2000.
16. J.L. Halbwachs, F. Arenou, M. Mayor, S. Udry & D. Queloz, *Astron. Astrophys.,* **355**, 581, 2000.
17. S. Udry & M. Mayor, in: "First European Workshop on Exo/Astrobiology", ESA SP-496, P. Ehrenfreund, O. Angerer & B. Battrick (Eds.), ESA Publ. Div., Noordwijk, p. 65, 2001.
18. P. Goldreich & S. Tremaine, *Astrophys. J.,* **241**, 425, 1980.
19. W.R. Ward, *Icarus,* **126**, 261, 1997.
20. C.E.J.M.L.J. Terquem, in Ref. 1.
21. S. Zucker & T. Mazeh, *Astrophys. J. Lett.,* **568**, L113, 2002.
22. S. Udry *et al., Astron. Astrophys.,* **390**, 267, 2002.
23. M. Pätzold & H. Rauer, *Astrophys. J. Lett.,* **568**, L117, 2002.
24. R.P. Nelson, J.C.B. Papaloizou, F. Masset & W. Kley, *Mon. Not. R. Astron. Soc.,* **318**, 18, 2000.
25. D. Trilling, J. Lunine & W. Benz, *Astron. Astrophys.,* **394**, 241, 2002.
26. D. Trilling *et al., Astrophys. J.,* **500**, 428, 1998.
27. G.W. Marcy & R.P. Butler, *Astrophys. J. Lett.,* **464**, L147, 1996.
28. D. Naef *et al., Astron. Astrophys.,* **375**, L27, 2001.
29. M. Holman, J. Touma & S. Tremaine, *Nature,* **386**, 254, 1997.
30. S. Udry, A. Eggenberger & M. Mayor, in Ref. 1.
31. D.N.C. Lin & S. Ida, *Astrophys. J.,* **477**, 781, 1997.
32. H.F. Levison, J.J. Lissauer & M.J. Duncan, *Astron. J.,* **116**, 1998, 1998.
33. P. Artymowicz & S.H. Lubow, *Astrophys. J.,* **467**, L77, 1996.
34. N.C. Santos, G. Israelian & M. Mayor, *Astron. Astrophys.,* **415**, 1153, 2004.
35. W. Kley, in Ref. 1.

36. R.P. Butler *et al*., *Astrophys. J.,* **526**, 916, 1999.
37. N.C. Santos *et al*., *Astron. Astrophys.,* **426**, L19, 2004.
38. G.W. Marcy *et al*., *Astrophys. J.,* **556**, 296, 2001.
39. E.J. Rivera & J.J. Lissauer, *Astrophys. J.,* **558**, 392, 2001.
40. G. Laughlin & J.E. Chambers, *Astrophys. J.,* **551**, L109, 2001.
41. T. Mazeh *et al*., *Astrophys. J.,* **532**, L55, 2000.
42. D. Charbonneau, T.M. Brown, D.W. Latham & M. Mayor, *Astrophys. J.,* **529**, L45, 2000.
43. G.W. Henry, G.W. Marcy, R.P. Butler & S.S. Vogt, *Astrophys. J.,* **529**, L41, 2000.
44. T.M. Brown, D. Charbonneau, R.L. Gilliland, R.W. Noyes & A. Burrows, *Astrophys. J.,* **552**, 699, 2001.
45. A. Burrows *et al*., *Astrophys. J.,* **534**, L97, 2000.
46. D. Charbonneau, T.M. Brown, R.W. Noyes & R.L. Gilliland, *Astrophys. J.,* **568**, 377, 2002.
47. A. Vidal-Madjar *et al*., *Astrophys. J.,* **604**, L69, 2004.
48. R. Alonso *et al*., *Astrophys. J.,* **613**, L153, 2004.
49. D. Charbonneau *et al*., *Astrophys. J*., accepted for publication.
50. A. Udalski, M. Kubiak & M. Szymanski, *Acta Astron.,* **47**, 319, 1997.
51. A. Udalski *et al*., *Acta Astron.,* **52**, 1, 2002.
52. A. Udalski *et al*., *Acta Astron.,* **52**, 115, 2002.
53. A. Udalski *et al*., *Acta Astron*., **52**, 317, 2002.
54. A. Udalski *et al*., *Acta Astron.,* **53**, 133, 2003.
55. M. Konacki, G. Torres, S. Jha & D.D. Sasselov, *Nature,* **421**, 507, 2003.
56. F. Bouchy *et al*., *Astron. Astrophys.,* **421**, L13, 2004.
57. M. Konacki *et al*., *Astrophys. J.,* **609**, L37, 2004.
58. F. Pont *et al*., *Astron. Astrophys.,* **426**, L15, 2004.
59. C. Moutou *et al*., in Ref. 1.
60. W.J. Borucki *et al*., *Proc. SPIE,* **4854**, 129, 2003.
61. S. Rucinski, K. Carroll, R. Kuschnig, J. Matthews & P. Stibrany, *Adv. Space Res.,* **31**, 371, 2003.
62. J. Matthews *et al*., *Nature,* **430**, 51, 2004.
63. C. Turon, *Rev. Mod. Astron.,* **9**, 69, 1996.
64. D. Pourbaix, *Astron. Astrophys.,* **369**, L22, 2001.
65. D. Pourbaix & F. Arenou, *Astron. Astrophys.,* **372**, 935, 2001.
66. S. Zucker & T. Mazeh, *Astrophys. J.,* **562**, 549, 2001.
67. J.C. Marr, *Proc. SPIE,* **4852**, 1, 2003.
68. M.A.C. Perryman, *Astrophys. Space Sci.,* **280**, 1, 2002.
69. C.V.M. Fridlund, *Adv. Space Res.,* **34**, 613, 2004.
70. A. Legér, *Adv. Space Res.,* **25**, 2209, 2000.
71. P. Ehrenfreund *et al*., *Rep. Prog. Phys.,* **65**, 1427, 2002.
72. S.Z. acknowledges the support by the European RTN "The Origin of Planetary Systems" (PLANETS, Contract number HPRN-CT-2002-00308) in the form of a fellowship.

List of Authors

K. Altwegg
Physikalisches Institut, Universität Bern, Bern, Switzerland
e-mail: altwegg@phim.unibe.ch

A. Balogh
The Blackett Laboratory, Imperial College, London, UK
e-mail: a.balogh@imperial.ic.ac.uk

J. Beer
Swiss Federal Institute for Environmental Science and Technology, Duebendorf, Switzerland
e-mail: beer@eawag.ch

R.-M. Bonnet
International Space Science Institute (ISSI), Bern, Switzerland
e-mail: rmbonnet@issi.unibe.ch

L. Colangeli
INAF, Astronomical Observatory of Capodimonte, Naples, Italy
e-mail: colangel@na.astro.it

L.A. Fisk
Department of Atmospheric, Oceanic and Space Sciences University of Michigan, Ann Arbor, USA
e-mail: lafisk@umich.edu

P. Frisch
Department of Astronomy and Astrophysics, University of Chicago, Chicago, USA
e-mail: frisch@oddjob.uchicago.edu

C. Fröhlich
Physikalisch-Meteorologisches Observatorium and World Radiation Center, Davos, Switzerland
e-mail: cfrohlich@pmodwrc.ch

J. Geiss
International Space Science Institute (ISSI), Bern, Switzerland
e-mail: Geiss@issi.unibe.ch

G. Gloeckler
Department of Physics and IPST, University of Maryland, USA
e-mail: gg10@umail.umd.edu

E. Grün
Max-Planck-Institut für Kernphysik , Heidelberg, Germany and
Hawaii Institute of Geophysics and Planetology, Honolulu, USA
e-mail: Eberhard.Gruen@mpi-hd.mpg.de

W.K. Hartmann
Planetary Science Institute, Tucson, Arizona, USA
e-mail: hartmann@psi.edu

P. Hoppe
Max-Planck-Institut für Chemie, Mainz, Germany
e-mail: hoppe@mpch-mainz.mpg.de

W.F. Huebner
Southwest Research Institute, San Antonio, Texas, USA
e-mail: whuebner@swri.edu

B. Hultqvist
Swedish Institute of Space Physics, Kiruna, Sweden
e-mail: hultqv@irf.se

V. Izmodenov
Department of Aeromechanics and Gas Dynamics, Faculty of Mechanics and
Mathematics, Lomonosov Moscow State University, Russia
e-mail: izmod@ipmnet.ru

R. Kallenbach
International Space Science Institute (ISSI), Bern, Switzerland
e-mail: kallenbach@issi.unibe.ch

R. Lallement
Service d'Aéronomie CNRS, Verrières-le-Buisson, France
e-mail: rosine.lallement@aerov.jussieu.fr

G.M. Mason
Department of Physics and IPST, University of Maryland, USA
e-mail: gmmason@umd.edu

K.G. McCracken
Institute for Physical Science and Technology, University of Maryland, USA
e-mail: jellore@hinet.com.au

F.B. McDonald
Institute for Physical Science and Technology, University of Maryland, USA
e-mail: fm27@umail.umd.edu

M. Mayor
Observatoire de Genève, Sauverny, Switzerland
e-mail: michel.mayor@obs.unige.ch

R.A. Mewaldt
California Institute of Technology, Pasadena, USA
e-mail: rmewaldt@aegir.srl.caltech.edu

E. Möbius
Space Science Center and Dept. of Physics, University of New Hampshire, Durham, USA
e-mail: Eberhard.Moebius@unh.edu

G. Paschmann
International Space Science Institute (ISSI), Bern, Switzerland
e-mail: paschmann@issi.unibe.ch

R.Z. Sagdeev
Department of Physics, University of Maryland, USA
e-mail: rzs@ew1.umd.edu

D. Sibeck
NASA/Goddard Space Flight Center, Greenbelt, USA
e-mail: david.g.sibeck@nasa.gov

T. Terasawa
Department of Earth and Planetary Science, University of Tokyo, Tokyo, Japan
e-mail: terasawa@geoph.s.u-tokyo.ac.jp

R.A. Treumann
Max-Planck-Institute for Extraterrestrial Physics, Garching, and
Dept. of Geosciences, Ludwig-Maximilians-University, Munich, Germany
e-mail: tre@mpe.mpg.de

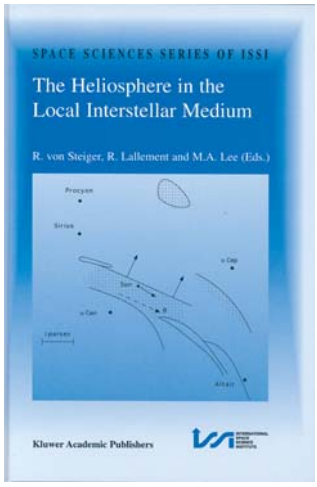
R. von Steiger
International Space Science Institute (ISSI), Bern, Switzerland
e-mail: vsteiger@issi.unibe.ch

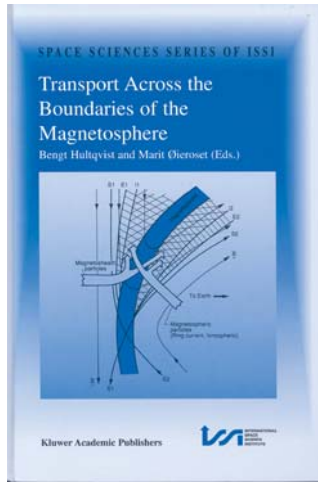
D. Winterhalter
Jet Propulsion Laboratory, Pasadena, California, USA
e-mail: daniel.winterhalter@jpl.nasa.gov

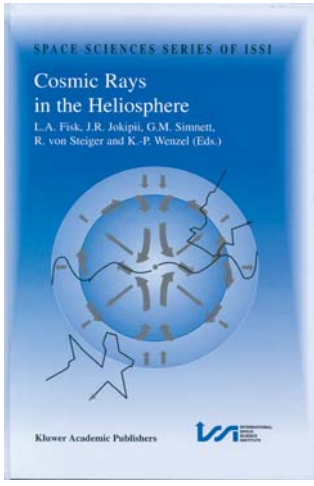
L. Zelenyi
Space Research Institute (IKI), Russian Academy of Sciences, Moscow, Russia
e-mail: lzelenyi@iki.rssi.ru

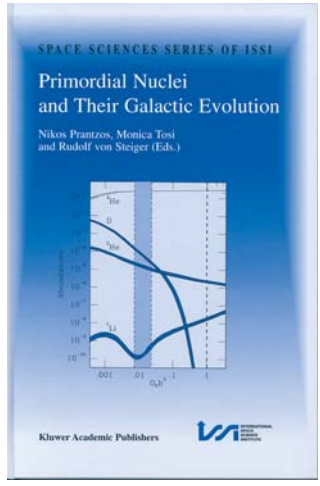
S. Zucker
Observatoire de Genève, Sauverny, Switzerland
e-mail: shay.zucker@wisemail.weizmann.ac.il

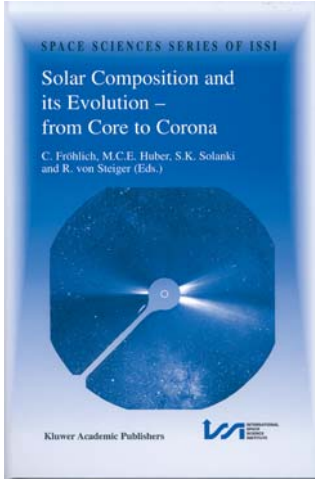
ISSI volumes

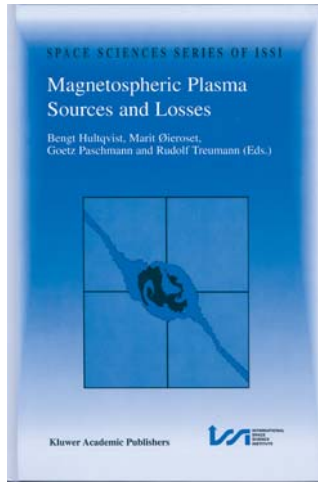

SPACE SCIENCES SERIES OF ISSI
The Heliosphere in the Local Interstellar Medium
R. von Steiger, R. Lallement and M.A. Lee (Eds.)
Kluwer Academic Publishers

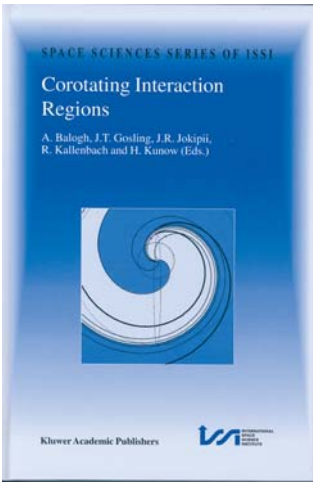

SPACE SCIENCES SERIES OF ISSI
Transport Across the Boundaries of the Magnetosphere
Bengt Hultqvist and Marit Øieroset (Eds.)
Kluwer Academic Publishers

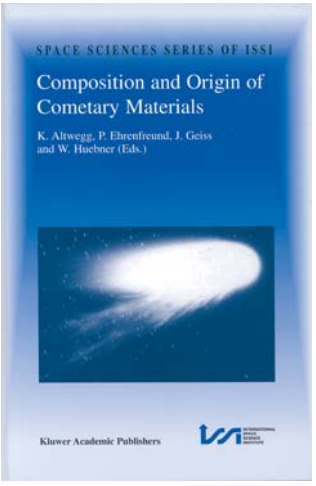

SPACE SCIENCES SERIES OF ISSI
Cosmic Rays in the Heliosphere
L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger and K.-P. Wenzel (Eds.)
Kluwer Academic Publishers

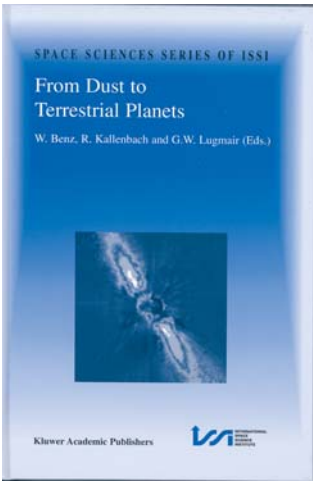

SPACE SCIENCES SERIES OF ISSI
Primordial Nuclei and Their Galactic Evolution
Nikos Prantzos, Monica Tosi and Rudolf von Steiger (Eds.)
Kluwer Academic Publishers

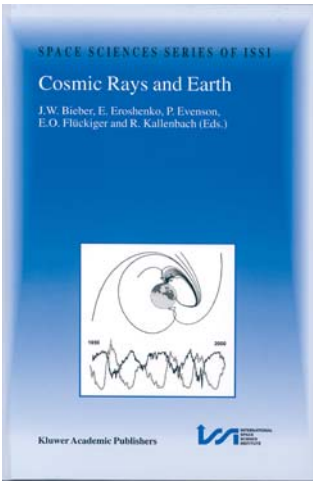

SPACE SCIENCES SERIES OF ISSI
Solar Composition and its Evolution – from Core to Corona
C. Fröhlich, M.C.E. Huber, S.K. Solanki and R. von Steiger (Eds.)
Kluwer Academic Publishers

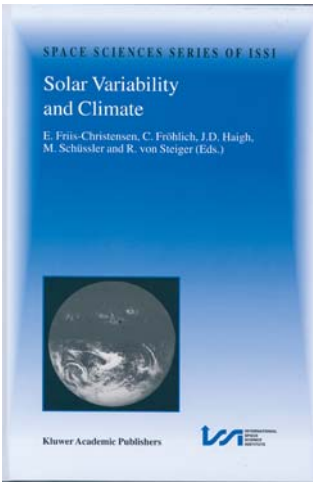

SPACE SCIENCES SERIES OF ISSI
Magnetospheric Plasma Sources and Losses
Bengt Hultqvist, Marit Øieroset, Goetz Paschmann and Rudolf Treumann (Eds.)
Kluwer Academic Publishers

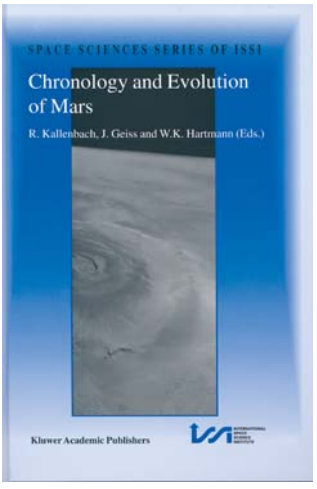

SPACE SCIENCES SERIES OF ISSI
Corotating Interaction Regions
A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach and H. Kunow (Eds.)
Kluwer Academic Publishers

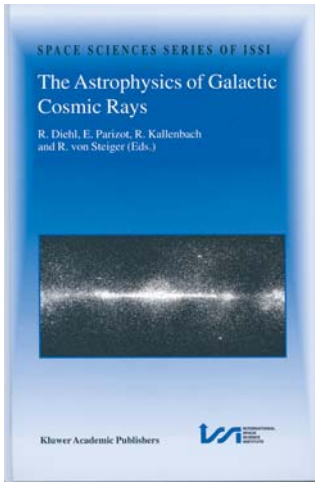

SPACE SCIENCES SERIES OF ISSI
Composition and Origin of Cometary Materials
K. Altwegg, P. Ehrenfreund, J. Geiss and W. Huebner (Eds.)
Kluwer Academic Publishers

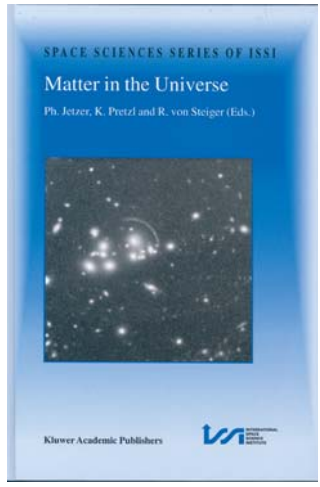

SPACE SCIENCES SERIES OF ISSI
From Dust to Terrestrial Planets
W. Benz, R. Kallenbach and G.W. Lugmair (Eds.)
Kluwer Academic Publishers

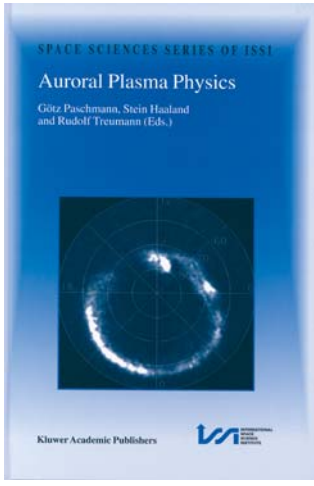

SPACE SCIENCES SERIES OF ISSI
Cosmic Rays and Earth
J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger and R. Kallenbach (Eds.)
Kluwer Academic Publishers

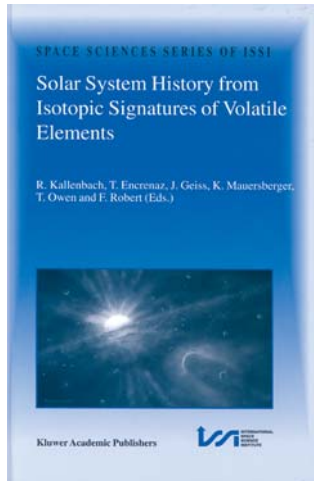

SPACE SCIENCES SERIES OF ISSI
Solar Variability and Climate
E. Friis-Christensen, C. Fröhlich, J.D. Haigh, M. Schüssler and R. von Steiger (Eds.)
Kluwer Academic Publishers

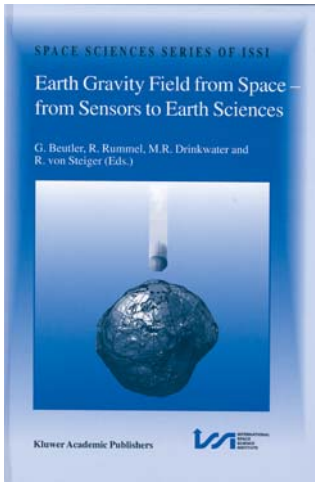

SPACE SCIENCES SERIES OF ISSI
Chronology and Evolution of Mars
R. Kallenbach, J. Geiss and W.K. Hartmann (Eds.)
Kluwer Academic Publishers

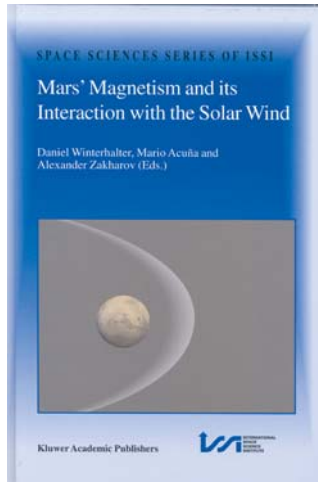

SPACE SCIENCES SERIES OF ISSI
The Astrophysics of Galactic Cosmic Rays
R. Diehl, E. Parizot, R. Kallenbach and R. von Steiger (Eds.)
Kluwer Academic Publishers

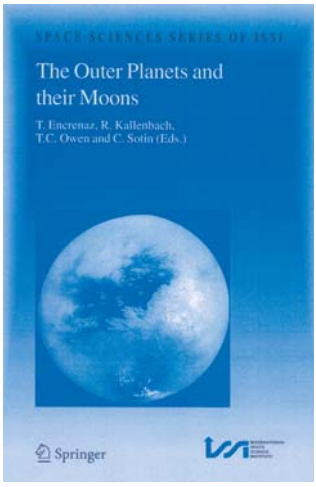

SPACE SCIENCES SERIES OF ISSI
Matter in the Universe
Ph. Jetzer, K. Pretzl and R. von Steiger (Eds.)
Kluwer Academic Publishers

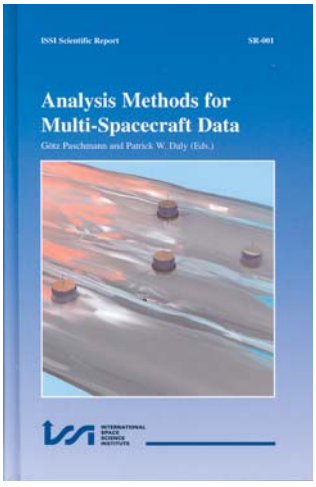

SPACE SCIENCES SERIES OF ISSI
Auroral Plasma Physics
Götz Paschmann, Stein Haaland and Rudolf Treumann (Eds.)
Kluwer Academic Publishers

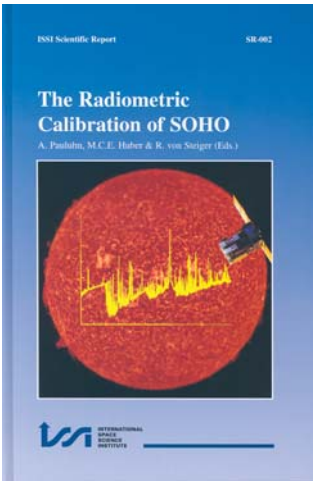

SPACE SCIENCES SERIES OF ISSI
Solar System History from Isotopic Signatures of Volatile Elements
R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T. Owen and F. Robert (Eds.)
Kluwer Academic Publishers

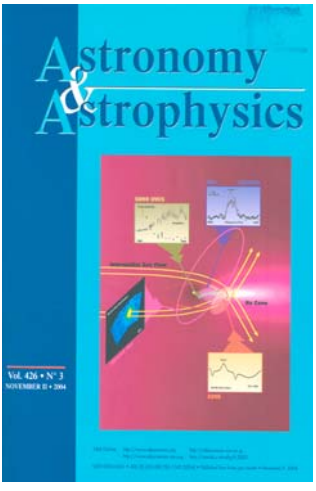

SPACE SCIENCES SERIES OF ISSI
Earth Gravity Field from Space – from Sensors to Earth Sciences
G. Beutler, R. Rummel, M.R. Drinkwater and R. von Steiger (Eds.)
Kluwer Academic Publishers

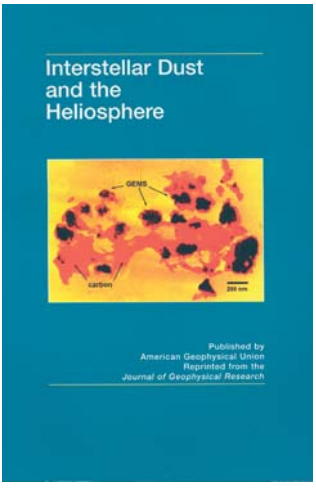

SPACE SCIENCES SERIES OF ISSI
Mars' Magnetism and its Interaction with the Solar Wind
Daniel Winterhalter, Mario Acuña and Alexander Zakharov (Eds.)
Kluwer Academic Publishers


SPACE SCIENCES SERIES OF ISSI
The Outer Planets and their Moons
T. Encrenaz, R. Kallenbach, T.C. Owen and C. Sotin (Eds.)
Springer


ISSI Scientific Report
SR-001
Analysis Methods for Multi-Spacecraft Data
Götz Paschmann and Patrick W. Daly (Eds.)


ISSI Scientific Report
SR-002
The Radiometric Calibration of SOHO
A. Pauluhn, M.C.E. Huber & R. von Steiger (Eds.)


Astronomy & Astrophysics
Vol. 426 • N° 3
NOVEMBER II • 2004


Interstellar Dust and the Heliosphere
GEMS
carbon
200 nm
Published by
American Geophysical Union
Reprinted from the
Journal of Geophysical Research

European Space Agency
Agence spatiale européenne

Contact: ESA Publications Division
c/o ESTEC, PO Box 299, 2200 AG Noordwijk, The Netherlands
Tel. (31) 71 565 3400 - Fax (31) 71 565 5433