

The background of the cover is a vibrant cosmic scene. It features a large, bright, yellowish-white spiral galaxy on the right side, with its arms swirling outwards. The galaxy is surrounded by a field of stars and nebulae in various colors, including red, blue, and purple. In the foreground, several celestial bodies are arranged in a diagonal line from the top left towards the bottom right. These include Mars (a reddish-brown sphere), Earth (a blue and white sphere showing continents and oceans), a small crescent moon, Jupiter (a large orange and white banded sphere), and a small brownish sphere. Thin, grey, curved lines represent orbital paths around these bodies. At the bottom of the image, there is a large, bright, orange and yellow sun or star, partially obscured by the text.

→ COSMIC VISION

**White Papers submitted in response to
ESA's Call for Science Themes
for the L2 and L3 Missions**

July 2013

Cover image credit: ESA – C. Carreau

This booklet is available online at: <http://sci.esa.int/white-papers-2013/>

Foreword

The ESA Scientific Programme has succeeded in putting European scientists at the forefront in many different fields, and through ESA's science missions Europe has scored a number of "firsts". These include landing the Huygens probe on the surface of Saturn's giant moon, Titan, and building and operating the Herschel infrared observatory, equipped with a monolithic 3.5 m diameter mirror, the largest ever flown in space.

More such "firsts" are in the pipeline. Gaia, the most advanced astrometric mission ever built, is scheduled to be launched a few months from now. In 2014, Rosetta will become the first spacecraft to fly alongside a comet and release a lander onto the surface of its nucleus.

The programme's success has been largely driven by two elements, namely long-term planning and a bottom-up definition of its content. Long-term planning means that the programme's largest missions are decided well in advance, allowing the scientific community and the national funding agencies to set their priorities, to develop the necessary technology and to train young researchers in the chosen fields.

The programme's first long-term planning exercise, known as Horizon 2000, took place in the early 1980's, and resulted in the commitment to implement four key "Cornerstone" missions: the combination of SOHO and Cluster (the programme's "Solar-Terrestrial Cornerstone"), Rosetta (the programme's "Planetary Cornerstone"), XMM-Newton, and Herschel.

While this decision committed the programme for almost three decades (with Herschel finally launched in 2009), it allowed the European scientific community to come of age and to eventually become a world leader in each respective field of research. Further long-term planning exercises have resulted in the development of Gaia and BepiColombo (the first now very close to launch, the second currently being built), and JUICE, Europe's mission to the icy moons of Jupiter, whose launch is planned for 2022.

The programme's long-term planning has always relied on a bottom-up process, with a broad community consultation and a peer-review process through which a group of senior scientists advises the Director about the content of the programme.

Our missions are the result of this painstaking process, and their success is a testimony to the foresight of the scientists who selected them among a treasure trove of competing proposals, and of ESA's Member States, who took their advice and agreed to implement it.

This combination of bottom-up, peer-reviewed programme definition and long-term planning is truly unique to Europe's space science programme, and has led to its becoming a clear reference for the worldwide space science community.

While the landscape of space science in Europe has changed significantly since the programme's first long-term planning effort, the benefits of the approach are still evident. With three large missions in the pipeline (Gaia, BepiColombo and JUICE), the time is now ripe for a new long-term planning cycle to define the future.

For this reason, I agreed with the Science Programme Committee (SPC) at their February 2013 meeting to start a new planning cycle by soliciting proposals from the scientific communities for the science themes that should be addressed by the two "Large" missions that will follow JUICE. On the basis of the programme's current planning, these flight opportunities (at the moment uninspiringly dubbed "L2" and "L3") are scheduled for launch in 2028 and 2034. While this may sound a long time away, 20 years is the typical planning timescale of such Large missions.

As discussed with the SPC, the Large missions are the programme's pillars, and Europe's flagships. As such, they will be Europe-led, although international participation will be welcome (as it was in the previous Cornerstone missions).

The programme's current planning foresees missions with a financial envelope of €1 billion (at today's prices), to which the value of the scientific instruments, normally funded by the Member States, will have to be added. This is comparable to the actual cost of previous cornerstone missions, once more emphasizing the programme's planning stability.

A Call for White Papers was issued in March 2013, and this booklet publishes the proposals received from the scientific community in response to the Call. The authors of the White Papers have given their consent to their publication, and these have not been edited or modified in any form. As such they represent the views from the community about which missions should be the pillars of the ESA Science Programme for the next decades. As the reader will see, the level of ambition varies among the different proposals, as does the likely maturity of the proposed "strawman missions".

I have appointed a Senior Survey Committee, chaired by Dr. Catherine Cesarsky, to advise me and to recommend, based on the White Papers published here, the science themes that should be addressed by the L2 and L3 opportunities. The Senior Survey Committee will be supported in their work by ESA's engineering expertise, in particular to ensure that the selected scientific themes can be later addressed by feasible, realistic mission concepts.

Based on the recommendation of the Senior Survey Committee, I will propose the science themes for the L2 and L3 missions to the SPC at their November 2013 meeting, and, pending a positive SPC decision, I intend to issue a Call for the L2 mission in 2014. This Call will be restricted to the science theme decided through the present process, and it will result in the selection of a specific mission concept and of the Principal Investigator teams that will provide the scientific instrumentation. A similar process for the L3 mission will take place around the end of the current decade.

The response from the community in writing the White Papers was both reassuring and staggering. Reassuring in its richness of ideas and in the level of ambition, showing a healthy community ready to extend the successes of the previous decades. Staggering in the number of solid ideas for the future, of which only two can be selected by the Senior Survey Committee, although many more would certainly deserve to be realized on the time frame of the L2 and L3 launch opportunities.

The work of the Senior Survey Committee will be challenging, but it will establish the future pillars around which the edifice of Europe's space science for the next two decades will be built.

Alvaro Gimenez
Director of Science and Robotic Exploration

Index of White Papers

The papers are listed in the order in which they were received.

Lunar Science as a Window into the Early History of the Solar System

Spokesperson: I.A. Crawford, Department of Earth and Planetary Sciences, Birkbeck College London, United Kingdom

Exploring Planetary Origins and Environments in the Infrared: A Planetary Science Infrared Observatory

Spokesperson: Leigh N. Fletcher, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, University of Oxford, United Kingdom

In Situ Exploration of the Giant Planets and an Entry Probe Concept for Saturn

Spokesperson: Olivier Mouis, Institut UTINAM, Université de Franche-Comté, France

Neptune and Triton: Essential Pieces of the Solar System Puzzle

Spokesperson: Adam Masters, Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Japan

Venus: Key to understanding the evolution of terrestrial planets

Spokesperson: Colin Wilson, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, University of Oxford, United Kingdom

INSIDER: Interior of Primordial Asteroids and the Origin of Earth's Water

Spokesperson: Pierre Vernazza, Laboratoire d'Astrophysique de Marseille, France

In situ Investigations of the Local Interstellar Medium

Spokesperson: Robert F. Wimmer-Schweingruber, Institute of Experimental and Applied Physics, University of Kiel, Germany

The Exploration of Titan with an Orbiter and a Lake-Probe

Spokesperson: Giuseppe Mitri, Istituto Nazionale di Astrofisica e Planetologia Spaziali, Istituto Nazionale di Astrofisica, Rome, Italy

Astrometry for Dynamics

Spokesperson: Erik Høg, Niels Bohr Institute, Copenhagen, Denmark

Europe returns to Venus

Spokesperson: Emmanuel Marcq, Université de Versailles Saint-Quentin / LATMOS, Versailles, France

Fundamental Processes in Solar Eruptive Events

Spokesperson: S.A. Matthews, Mullard Space Science Laboratory, University College London, United Kingdom

European Ultraviolet-Visible Observatory: Building galaxies, stars, planets and the ingredients for life among the stars

Spokesperson: Ana Inés Gomez de Castro, AEGORA-Facultad de CC Matemáticas, Universidad Complutense de Madrid, Spain

The science goals and mission concept for a future exploration of Titan and Enceladus

Spokesperson: Gabriel Tobie, Laboratoire de Planétologie et Geodynamique de Nantes, CNRS/University of Nantes, France

The Gravitational Universe

Spokesperson: Karsten Danzmann, Albert Einstein Institute Hannover, Max Planck Institute for Gravitational Physics and Leibniz Universität Hannover, Germany

SOLARIS: SOLAR sail Investigation of the Sun

Spokesperson: Thierry Appourchaux, Institut d'Astrophysique Spatiale, Orsay, France

Science from the Farside of the Moon

Spokesperson: Mark Wieczorek, Institut de Physique du Globe de Paris, Université Paris Diderot, France

Light from the Cosmic Frontier: Gamma-Ray Bursts

Spokesperson: Nial R. Tanvir, Department of Physics and Astronomy, University of Leicester, United Kingdom

Stellar Imager

Spokesperson: Thierry Appourchaux, Institut d'Astrophysique Spatiale, Orsay, France

Chronos: A NIR Spectroscopic Galaxy Formation Survey

Spokesperson: Ignacio Ferreras, Mullard Space Science Laboratory, University College London, United Kingdom

Exploring Habitable Worlds beyond our Solar System

Spokesperson: Andreas Quirrenbach, Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Germany

Venus: A Natural Planetary Laboratory

Spokesperson: Sanjay S. Limaye, University of Wisconsin, Madison, Wisconsin, USA

Space-Time Structure Explorer: Sub-microarcsecond astrometry for the 2030s

Spokesperson: Anthony Brown, Leiden Observatory, Leiden University, The Netherlands

DEX - Dark Ages eXplorer

Spokesperson: Marc Klein Wolt, Department of Astrophysics, Research Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen, The Netherlands

Solar System Debris Disk - S2D2

Spokesperson: Ralf Srama, Universität Stuttgart, Institut für Raumfahrtssysteme, Raumfahrtzentrum Baden Württemberg, Germany

PRISM: Polarized Radiation Imaging and Spectroscopy Mission

Spokesperson: Paolo de Bernardis, Dipartimento di Fisica, Università di *Roma* "La Sapienza", Italy

Sub-arcsecond far-infrared space observatory: a science imperative

Spokesperson: Marc Sauvage, CEA/DSM/Irfu/SAP, CE Saclay, France

The Case for an ESA L-Class Mission to Volatile-Rich Asteroids

Spokesperson: Geraint H. Jones, Mullard Space Science Laboratory, University College London, United Kingdom

The Science Case for an Orbital Mission to Uranus: Exploring the Origins and Evolution of Ice Giant Planets

Spokesperson: Christopher S. Arridge, Mullard Space Science Laboratory, University College London, United Kingdom

Master: A Mission to Return a Sample from Mars to Earth

Spokesperson: Monica M. Grady, The Open University, Milton Keynes, United Kingdom

Hypertelescope Optical Observatory

Spokesperson: Antoine Labeyrie, Collège de France and Observatoire de la Côte d'Azur, France

The ODINUS Mission Concept – The Scientific Case for a Mission to the Ice Giant Planets with Twin spacecraft to Unveil the History of our Solar System

Spokesperson: Diego Turrini, Istituto di Astrofisica e Planetologia Spaziali, Rome, Italy

The Hot and Energetic Universe

Spokesperson: Kirpal Nandra, Max Planck Institute for Extraterrestrial Physics, Garching, Germany

Lunar Science as a Window into the Early History of the Solar System

**A White Paper submitted in response to ESA's Call for
Proposals for Cosmic Vision L2/3 Science Themes**

I. A. Crawford, K.H. Joy, N. Bowles, R. Jaumann, et al.

(For full list of contributors and endorsers see overleaf)



Frontispiece: The Earth and Moon photographed together by the Galileo spacecraft from a distance of 6.2 million km. The binary nature of the Earth-Moon system means that the Moon's history is intimately connected with that of our own planet (NASA).

Principal Contact

Professor I.A. Crawford, Department of Earth and Planetary Sciences, Birkbeck College
London, Malet Street, London, WC1E 7HX, UK.

Telephone: +44 (0)20 3073 8026

Email: i.crawford@bbk.ac.uk

Co-Proposers

Dr Neil Bowles, University of Oxford, UK

Prof. Ralf Jaumann, DLR, Institute of Planetary Research, Germany

Dr Katherine Joy, SEAES, University of Manchester, UK

Contributors and Endorsers

Dr Mahesh Anand, The Open University, UK

Dr Sebastien Besse, ESA-ESTEC, The Netherlands

Dr Bill Bottke, Southwest Research Institute, Boulder, CO, USA

Dr Veronica Bray, Lunar and Planetary Laboratory, Tucson, USA

Prof Mark Burchell, University of Kent, UK

Dr James Carpenter, ESA-ESTEC, The Netherlands

Prof Marc Chaussidon, CRPG-CNRS, Nancy, France

Prof Julian Chela-Flores, Abdus Salam International Centre for Theoretical Physics, Italy

Prof Andrew Coates, Mullard Space Science Laboratory, UK

Prof Charles Cockell, University of Edinburgh, UK

Dr Paolo D'Arrigo, Astrium Ltd, UK

Dr Jean-Pierre de Vera, DLR, Institute of Planetary Research, Germany

Prof Heino Falcke, Radboud University, The Netherlands

Dr Vera A. Fernandes, Museum für Naturkunde, Berlin, Germany

Dr. Jörg Fritz, Museum für Naturkunde, Berlin, Germany

Prof Yang Gao, University of Surrey, UK

Dr Rebecca Ghent, University of Toronto, Canada

Dr Timothy Glotch, Stony Brook University, Stony Brook, USA

Prof Monica Grady, The Open University, UK

Prof Manuel Grande, Aberystwyth University, UK

Dr Peter Grindrod, University College London, UK

Dr Jordi Gutiérrez, Universitat Politècnica de Catalunya, Barcelona, Spain

Prof Harry Hiesinger, University of Muenster, Germany

Dr Marc Klein-Wolt, Radboud University, The Netherlands

Dr Martin Knapmeyer, DLR, Institute of Planetary Research, Germany

Dr David Kring, Lunar and Planetary Institute, Houston, TX, USA

Dr Tomas Magna, Czech Geological Survey, Czech Republic

Prof Bernard Marty, Ecole Nationale Supérieure de Géologie, Nancy, France

Dr Emanuele Monchieri, Astrium Ltd, UK

Prof Gordon Osinski, University of Western Ontario, Canada

Prof Alan Smith, Mullard Space Science Laboratory, UK

Prof Tilman Spohn, DLR, Institute of Planetary Research, Germany

Dr Nick Teanby, University of Bristol, UK

Prof Stephan van Gasselt, Freie Universität Berlin, Germany

Prof Mark Wieczorek, Institut de Physique du Globe de Paris, France

Prof Ian Wright, The Open University, UK

Dr Stephanie Werner, University of Oslo, Norway

Prof Wim van Westrenen, VU University Amsterdam, The Netherlands

Prof Lionel Wilson, Lancaster University, UK

Prof Robert F. Wimmer-Schweingruber, University of Kiel, Germany

Dr. Kai Wünnemann, Museum fuer Naturkunde Berlin, Germany

Prof Peter Wurz, University of Bern, Switzerland

EXECUTIVE SUMMARY

We propose a Cosmic Vision ‘Science Theme’ of using lunar science as a window into the early history of the Solar System. The near surface lunar environment contains a rich record of inner Solar System history. Accessing this record will directly address key elements of ESA’s Cosmic Vision themes, especially Theme 1 (‘Planets and Life’) and Theme 2 (‘How does the Solar System Work?’). In particular, the main areas of lunar science that will inform our understanding of inner Solar System evolution, and the past habitability of our own planet, are the following:

- *The bombardment history of the inner Solar System.* The lunar surface preserves a unique record of the bombardment history of the inner solar system, important for understanding the emergence of life on Earth, dating the surfaces of terrestrial planets and asteroids, and constraining the orbital evolution of the giant planets.
- *The record of lunar and extra-lunar processes recorded in the lunar regolith.* The lunar regolith is a unique witness to over 4 Ga of Solar System history and records changes in solar activity, the population of small bodies in the Solar System, and the passage of the Solar System through Galaxy. The regolith may further contain unique samples of Earth’s early surface and atmosphere not obtainable in any other way.
- *Studies of volatiles at the lunar poles.* Water and other volatiles at high lunar latitudes may reveal the nature and sources of compounds that enabled life on Earth, as well as providing a model for processes of water formation and migration on other airless bodies.

Implementation will require spacecraft to land on the lunar surface in order to make *in situ* measurements at, and/or return samples from, localities that have been carefully selected with specific scientific objectives in mind. For the Cosmic Vision L2/3 mission opportunities we propose the development of a scientific infrastructure that would enable us to address these scientific objectives. Such a lunar exploration infrastructure would also address other areas of high scientific importance, including studies of the evolution of the Moon itself as a planetary body, and geophysical, astrophysical and astrobiological investigations conducted on its surface. Here we propose two, mutually complementary, strands: (i) a mission based around multiple penetrators for the characterisation of lunar polar volatiles and (ii) a sample return mission to address the lunar impact chronology and records of the near-Earth Solar System environment preserved in regolith deposits. We consider that the development of such an ambitious lunar science architecture is worthy of consideration for the Cosmic Vision L2/3 mission opportunities. We note that many of the technical developments required for this lunar science programme are relevant for developing Mars Sample Return missions.

1 Introduction

From a planetary science perspective the primary importance of the Moon arises from the fact that it has an extremely ancient surface, mostly older than 3 billion years with some areas extending almost all the way back to the origin of the Moon 4.5 billion years ago (e.g., Hiesinger and Head, 2006; NRC 2007; Jaumann et al., 2012). Its relatively accessible near-surface environment therefore preserves a record of the early geological evolution of a terrestrial planet, which more complex planets such as Earth and Venus have long lost, and of the Earth-Moon system in particular (NRC, 2007; Jaumann et al., 2012). Moreover, the Moon's outer layers also preserve a record of the environment in the inner Solar System (e.g., meteorite flux, interplanetary dust density, solar wind flux and composition, galactic cosmic ray flux) throughout Solar System history, much of which is relevant to understanding the past habitability of our own planet (e.g., Crawford, 2006; NRC, 2007; Norman, 2009; Cockell, 2010; Crawford et al., 2012; Fernandes et al., 2013). Indeed, for the last 4.5 billion years the Earth and Moon have essentially comprised a binary planet system which is unique in the inner Solar System. During this time life has evolved and prospered on Earth, yet key aspects of our planet's early environment are poorly understood owing to active geological and meteorological cycles which have largely erased the geological record from much of Earth history. Fortunately, the binary nature of the Earth-Moon system provides a means of remedying this situation because records of the early space environment shared by the Earth-Moon system will be preserved on the ancient surface of the Moon.

Accessing this rich record of inner Solar System history will directly address key elements of ESA's Cosmic Vision themes, especially Theme 1 ('Planets and Life') and Theme 2 ('How does the Solar System Work?') and is a scientifically valuable theme for the Cosmic Vision L2/3 mission opportunities. Implementation will require spacecraft to land on the lunar surface in order to make *in situ* measurements at, and/or return samples from, localities that have been carefully selected with specific scientific objectives in mind. In what follows we describe the nature of the lunar geological record, and how it can inform our knowledge of the early history of the Earth and of the inner Solar System more generally, before going on to outline two possible, but not mutually exclusive, mission scenarios that would be well-suited to address these key scientific questions.

2 The nature of the lunar record

We here elaborate on those aspects of the lunar geological record which will provide key information concerning the evolution of the inner Solar System, including the Earth-Moon system, and the continued habitability of our own planet. Of course, the Moon is also an important object for scientific investigation in its own right, the interior structure of which records early planetary differentiation processes that will have affected all the terrestrial

planets but which the more evolved planets have long lost (see, e.g., Jaumann et al., 2012). Moreover, the Moon is a potential platform for low-frequency radio astronomy (Jester and Falcke, 2009) and for biological and astrobiological studies (Cockell et al., 2010; de Vera et al., 2012). Investigation of those other aspects of lunar science, which will largely rely on geophysical, astrophysical, and biological techniques are not directly addressed here (although some will be covered in other White Papers). That said, it is clear that strong synergies exist between the techniques and samples required to access the Moon's record of the inner Solar System environment and those required to understand the evolution of the Moon itself as a planetary body, and that lunar geophysical, astrophysical and astrobiological investigations would benefit from the development of a lunar scientific infrastructure. All these aspects of lunar science could be addressed by a suitable choice of instrument payloads on landed spacecraft.

With these caveats in mind, the main areas of lunar science that will explicitly inform our understanding of inner Solar System evolution are the following:

2.1 The Bombardment History of the Inner Solar System

The Lunar surface preserves a unique record of the bombardment history of the inner solar system, important for understanding the emergence of life on Earth, dating the surfaces of terrestrial planets and asteroids, and constraining the orbital evolution of the giant planets.

The vast majority of lunar terrains have never been directly sampled, and their inferred ages are based on the observed density of impact craters calibrated against the ages of Apollo and Luna samples (e.g., Neukum et al., 2001; Stöffler et al., 2006). However, the current calibration of the cratering rate, used to convert crater densities to absolute model ages, is neither as complete nor as reliable as it is often made out to be. For example, there are no calibration points that are older than about 3.85 Ga, and crater ages younger than about 3 Ga are also uncertain (e.g., Hiesinger et al., 2012). Improving the calibration of the cratering rate would be of great value for planetary science for the following three reasons: (i) It would provide better estimates for the ages of unsampled regions of the lunar surface; (ii) It would provide us with a more reliable estimate of the impact history of the inner Solar System, especially that of our own planet; and (iii) The lunar impact rate is used, with various other assumptions, to date the surfaces of other planets for which samples have not been obtained – to the extent that the lunar rate remains unreliable, so do the age estimates of surfaces on the other terrestrial planets.

Moreover, there is still uncertainty over whether the lunar cratering rate has declined monotonically since the formation of the Moon, or whether there was a bombardment 'cataclysm' between about 3.8 and 4.1 billion years ago characterised by an enhanced rate of impacts (Kring, 2003; Stöffler et al., 2006; Norman, 2009; Morbidelli et al., 2012). Indeed, recent studies of the ages of impact melt samples obtained by the Apollo and Luna missions

suggest a very complicated impact history for the Earth-Moon system, with a number of discrete spikes in the impact flux (Fernandes et al., 2013 and references therein). Clarifying this issue is especially important from an astrobiology perspective because it defines the impact regime under which life on Earth became established and the rate at which volatiles and organic materials were delivered to the early Earth (e.g., Maher and Stevenson, 1988; Sleep et al., 1989; Ryder, 2003). Additionally, as the inner Solar System bombardment history is thought to have been governed, at least in part, by changing tidal resonances in the asteroid belt (Gomes et al., 2005; Morbidelli et al., 2012; Bottke et al., 2012), improved constraints on the impact rate will lead to a better understanding of the orbital evolution of the early Solar System. This in turn will have implications for our understanding the habitability of planets within the 'habitable zones' of *other* planetary systems as a function of their age and the locations of any giant planets that may be present (Brock and Melosh, 2012; Johnson et al., 2012).

Obtaining an improved lunar cratering chronology requires the sampling, and radiometric dating, of surfaces having a wide range of crater densities, supplemented where possible by dating of impact melt deposits from individual craters and basins (Stöffler et al., 2006; Fernandes et al., 2013). In practice this will require robotic sample return missions to key localities from which samples have not yet been returned. Examples sites include the farside South Pole-Aitken basin (the dating of which will help determine whether or not most lunar basins formed in a single 'cataclysm'; e.g., Kring, 2003; Jolliff et al., 2010) and, at the other end of the age spectrum, young basaltic lava flows in Oceanus Procellarum on the nearside (where the dating of individual lava flows with ages in the range 1.1 to 3.5 Gyr would provide data points for the as yet uncalibrated 'recent' portion of the inner Solar System cratering rate; e.g. Stöffler et al., 2006; Crawford et al., 2007).

2.2 Treasures in the regolith

The lunar regolith is a unique witness to over 4 Ga of Solar System history and records changes in solar activity, the population of small bodies in the Solar System, and the passage of the Solar System through Galaxy. The regolith may further contain unique samples of Earth's early surface and atmosphere not obtainable in any other way.

The lunar regolith is known to contain much that is of interest for studies of Solar System history. For example, studies of Apollo samples have revealed that solar wind particles are efficiently implanted in the lunar regolith (McKay et al., 1991; Lucey et al., 2006), which therefore contains a record of the composition and evolution of the Sun throughout Solar System history (e.g., Wieler et al., 1996; Chaussidon and Robert, 1999; Hashizume et al., 2000). Recently, samples of the Earth's early atmosphere may have been retrieved from lunar regolith samples (Ozima et al., 2005; 2008), and it has been suggested that samples of Earth's early crust may also be preserved there in the form of terrestrial meteorites (Gutiérrez, 2002; Armstrong et al., 2002; Crawford et al., 2008; Armstrong, 2010). Meteorites derived from elsewhere in the Solar System may also be found on the Moon,

preserving a record of the dynamical evolution of small bodies throughout Solar System history (Joy et al., 2011; 2012). Last but not least, the lunar regolith may contain a record of galactic events, by preserving the signatures of ancient galactic cosmic ray (GCR) fluxes, and the possible accumulation of interstellar dust particles during passages of the Sun through dense interstellar clouds (Croaz et al., 1977; McKay et al., 1991; Crawford et al., 2010). Collectively, these lunar geological records would provide a window into the early evolution of the Sun and Earth, and of the changing galactic environment of the Solar System, that is unlikely to be obtained in any other way. Much of this record has clear astrobiological implications, as it relates to the conditions under which life first arose and evolved on Earth.

From the point of view of accessing ancient Solar System history it will be desirable to find layers of ancient regoliths (*palaeoregoliths*) that were formed and buried billions of years ago, and thus protected from more recent geological processes, (e.g., Spudis, 1996; Crawford et al., 2007, 2010; Fagents et al., 2010; Rumpf et al., 2013; see Figure 1 of Crawford et al., 2010 for a pictorial representation of the process). Locating and sampling such deposits will therefore be an important scientific objective of future lunar exploration activities.

2.3 Volatiles at the lunar poles

Water and other volatiles at high lunar latitudes may reveal the nature and sources of compounds that enabled life on Earth, as well as providing a model for processes of water formation and migration on other airless bodies.

The lunar poles potentially bear witness to the flux of volatiles present in the inner Solar System throughout much of Solar System history (e.g., NRC, 2007). In 1998 the *Lunar Prospector* neutron spectrometer found evidence of enhanced concentrations of hydrogen at the lunar poles (Feldman et al., 1998), which was widely interpreted as indicating the presence of water ice in the floors of permanently shadowed polar craters. This interpretation was supported by the LCROSS impact experiment, which found a water ice concentration of 5.6 ± 2.9 % by weight in the target regolith at the Cabeus crater (Colaprete et al., 2010). It seems likely that this water is ultimately derived from the impacts of comets and/or hydrated meteorites on to the lunar surface (Anand, 2010). In addition to ice in permanently shadowed craters, infra-red remote-sensing observations have found evidence for hydrated minerals, and/or adsorbed water or hydroxyl molecules, over large areas of the high latitude (but not permanently shadowed) lunar surface which may be due to oxidation of solar wind hydrogen within the regolith (Pieters et al., 2009; Liu et al., 2012).

As discussed by Anand (2010) and Smith et al. (2012), obtaining improved knowledge of the presence, composition, and abundance of water (and other volatiles) at the lunar poles is important for several reasons:

- It is probable that the ice in permanently shadowed regions is ultimately derived from comet and/or meteorite impacts. Even though the original volatiles will have been considerably reworked, it remains probable that some information concerning the composition of the original sources will remain (Zhang and Paige, 2009). Among other things, this may yield astrobiologically important knowledge on the role of comets and meteorites in delivering volatiles and pre-biotic organic materials to the terrestrial planets (Chyba and Sagan, 1992; Pierazzo and Chyba, 1999; Zhang and Paige, 2009).
- The processes involved in the creation, retention, migration, and destruction of OH and H₂O across the surface of the Moon are likely to be common on other air-less bodies, and quantifying them on the Moon will give us better insight into the volatile history and potential availability of water elsewhere in the inner Solar System.
- Lunar polar ice deposits are of considerable astrobiological interest, even if they do not retain vestigial information concerning their ultimate sources. This is because any such ices will have been continuously subject to irradiation by galactic cosmic rays and, as such, may be expected to undergo organic synthesis reactions (e.g., Lucey, 2000; Crites, et al., 2011). Analogous reactions may be important for producing organic molecules in the icy mantles of interstellar dust grains, and on the surfaces of outer Solar System satellites and comets, but the lunar poles are much more accessible than any of these other locations.
- The presence of water ice at the lunar poles, and even hydrated materials at high-latitude but non-shadowed localities, could potentially provide a very valuable resource (e.g., rocket fuel, habitation resources) in the context of future lunar exploration activities (e.g., Spudis and Lavoie, 2011).

Confirming the interpretation of the remote sensing measurements, and obtaining accurate values for the concentration of polar ice and high latitude surficial OH/H₂O will require *in situ* measurements by suitably instrumented and landed spacecraft, and we outline some possibilities below.

3 Strawman mission proposals

In this section we outline a scientific infrastructure that would enable us to address the scientific objectives described above and which we consider suitable for consideration within ESA's Cosmic Vision framework for implementation by either the L2 or L3 mission opportunities. There are two, mutually complementary, strands: (i) a mission based around multiple penetrators for the characterisation of lunar polar volatiles, and (ii) a sample return mission to address the lunar impact chronology and records of the near-Earth Solar System environment preserved in regolith deposits.

3.1 Penetrator Mission

Volatile detectors deployed on penetrators (emplaced ballistically into the lunar sub-surface), and landed within permanently shadowed craters (and/or the surrounding non-shadowed but apparently nevertheless volatile enhanced areas), would be a powerful and economical means of determining whether or not scientifically and operationally valuable deposits of volatiles exist at the lunar poles. One of the implications of the LCROSS and other recent spacecraft results is that such volatiles may be distributed very inhomogeneously in the lunar polar regions, and a mission with multiple penetrator capability would enable additional sampling of this distribution, which would be important in terms of understanding sources/sinks of polar volatiles.

Here we propose a mission that involves the delivery of a minimum of four penetrators into the lunar surface at multiple locations. Each penetrator will be ~0.5 m long and ~13 kg mass (similar to the JAXA Lunar-A mission concept; Mizutani et al., 2005). Each penetrator will consist of a supporting structure, a power system, communications system, data handling system, and payload. They will be delivered to the Moon by a spacecraft bus that will enter lunar orbit and act as a communications relay (as described by Smith et al., 2012). Provisionally, it is anticipated that two of the penetrators will be placed in permanently shadowed regions, one into a high-latitude non-permanently shadowed locality where remote sensing indicates the presence of surficial volatiles, and one penetrator at a low latitude site (either an Apollo landing site or the location of the Sample Return component discussed in Section 3.2.1) to act as a volatile-poor control.

Direct communication between these penetrator and Earth cannot be guaranteed and a lunar polar orbiting relay communications satellite (Orbiter) is therefore required. The Orbiter will carry the four penetrators and their descent modules (DMs) into lunar orbit prior to their release. This Orbiter may also act as a communications relay for the sample return component discussed in Section 3.2.1 if a farside locality is selected.

This element of the proposal closely follows that of the LunarNET proposal submitted to the Cosmic Vision M3 opportunity that has been described in detail by Smith et al. (2012). Detailed information relating to the mission profile, technological readiness, and spacecraft system requirements, for which there is insufficient space to describe here, will be found in that publication. Here we concentrate on the modified scientific payload tailored to address the scientific objectives outlined in Section 2.

The penetrator deployment is shown schematically in Figure 1, and example model payload instruments are listed in Table 3-1.

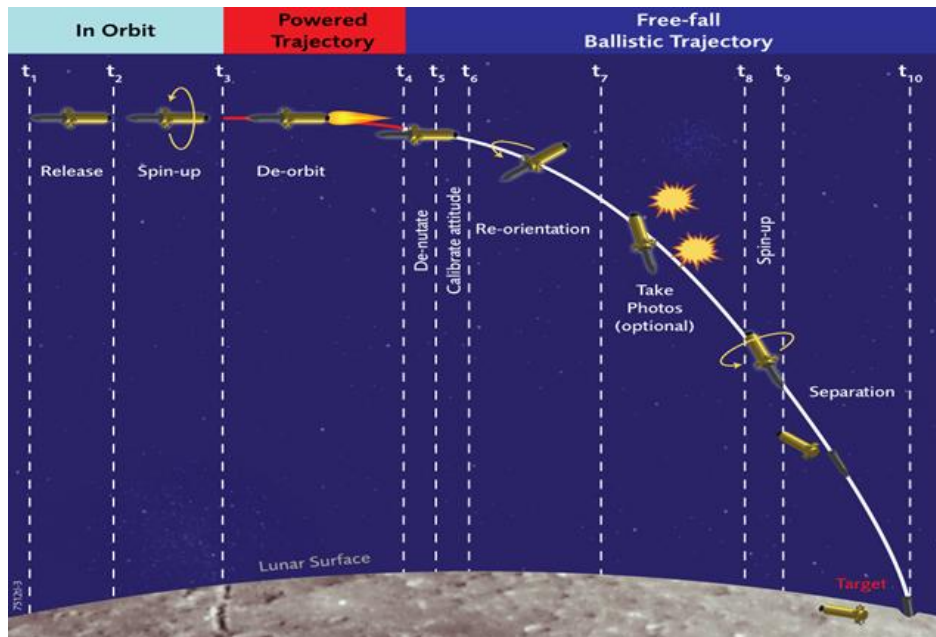


Figure 1. Provisional Descent Sequence (Courtesy Astrium)

Table 3-1: Proposed Penetrator Payload Instruments (for full details see Smith et al., 2012)

Instrument	Acro- nym	Mass [kg]	Size [cm ³]	Power [W] [W hr]	Total Data Volume [kbit]	Technical Readiness Level (TRL) Heritage *
Accelerometer (8 sensors)	ACCL	0.07	2.4	0.8 to 1.2 0.17	1 Mbit	TRL 6-8 Off-the-shelf components, Lunar A, Pendine
Descent camera	DC	0.160	27 3×3×3 cm	0.160 0.015	~ 10 Mbits after compression	TRL 7+ general camera technology TRL 2 for proposed design
Magnetometer	MAG	0.07	200 10×10 ×2	0.15 - 0.4	~1 Mbit (0.06 kbps)	TRL 5: Pendine trials
Mass Spectrometer	MSPC	0.75	1000 10×10 ×10	3-6	~0.2 Mbits	TRL 4/5: Rosetta / Beagle2
Engineering Tiltmeter	ETLT	0.010	25	0.1	1 kbit	TRL 6-8, Huygens, Mars 96
Water/Volatile Detector	BIOC	0.750	1000	3	TBD	TRL 4-8: DS-2, Huygens, ExoMars, Pendine
X-Ray Spectrometer	XRS	0.260	160	4 24	0.1 Mbits	TRL 7 : Mars 96
Microscopic Imager	MICI	TBD	TBD	TBD	TBD	
Radiation Monitor	RADM	TBD	TBD	TBD	TBD	MoonLITE

*Pendine refers to UK penetrator trials conducted in 2009 (see Smith et al., 2012).

3.2 Sample Return

In order to address the lunar chronology questions identified in Section 2, and to identify extra-lunar materials in the regolith (e.g., solar wind particles, cosmogenic nuclides produced by galactic cosmic rays, meteoritic fragments, etc.), we propose a mission element able to return of the order of 1-10 kg (TBC) of rock and soil samples to Earth. Mobility is highly desirable in order to secure a diverse set of samples, and we propose that two options be considered: (i) inclusion of a rover with a 5-10 km range, and (ii) a lander that is able to ‘hop’ to multiple (at least three) localities separated by tens or hundreds of km. A drilling capability would also be desirable to obtain samples from a vertical stratigraphic column. A possible mission architecture is described in Section 3.2.2, after we first discuss scientifically valuable landing sites for sample return.

3.2.1 Sample return sites.

We tentatively identify two sites for a sample return mission: (i) the young basaltic lava flows of Oceanus Procellarum at low latitudes on the nearside, and (ii) the farside, high southern latitude, Schrödinger Basin, which can be used to sample the South-Pole Aitkin basin as well as providing additional science opportunities. In addition to both being high priority science targets (e.g., Crawford et al., 2007; Kring and Durda, 2012 respectively), these two localities may be seen as bracketing relatively ‘easy’ and ‘difficult’ sample return locations and thereby constrain the spectrum of lunar sample return options. In the subsections below we outline the scientific advantages of each location and then summarise mission architectures required.

3.2.1.1 Oceanus Procellarum

Oceanus Procellarum consists of a patchwork of discrete lava flows with different compositions and estimated ages ranging from about 3.5 to 1.2 Gyr (Wilhelms, 1987; Hiesinger et al., 2003; Fig. 2). This is a far greater range of ages than any basalt samples collected by the Apollo missions (which occupy the narrow age range 3.8 to 3.1 Gyr).

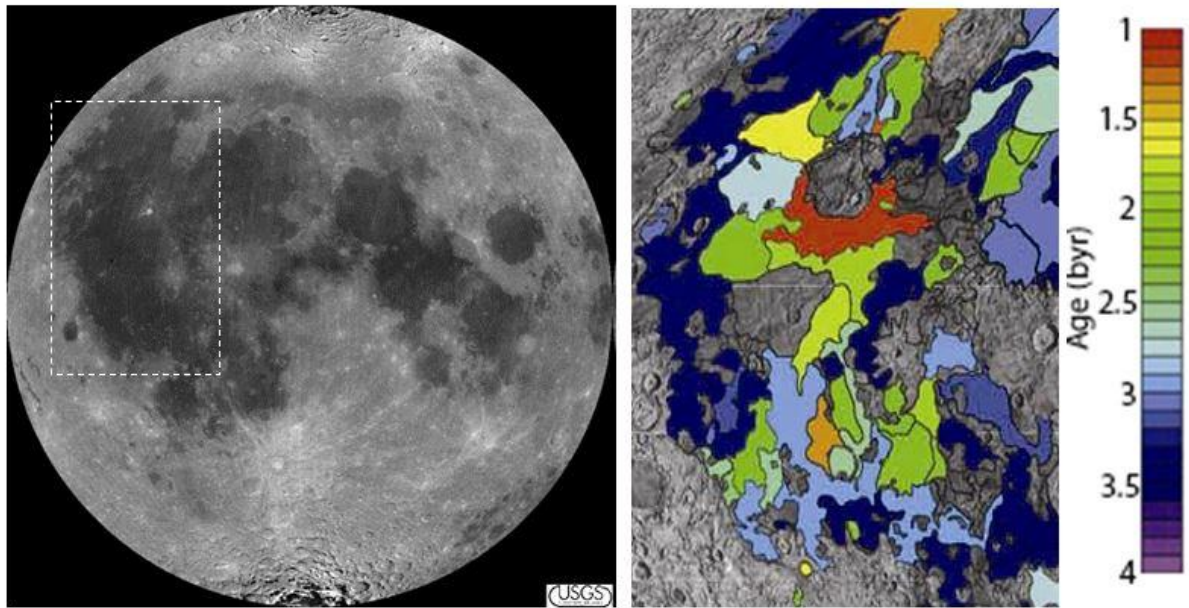


Figure 2. (left) Albedo map of the near side of the Moon. Dashed box represents region of Oceanus Procellarum mare basalts shown at right. (right) Absolute model ages of lava flows in Oceanus Procellarum, as mapped by Hiesinger et al. (2003). Sample return from one or more of these lava flows would verify these ages, with the benefits described in the text. (Image courtesy of Dr. H. Hiesinger; © AGU).

Collecting samples from one or more of these different lava flows, and returning them to Earth for radiometric dating, will directly lead to an improvement in the calibration of the lunar cratering rate for the last three billion years (see Stöffler et al., 2006). The post 3 Ga lunar cratering rate is poorly calibrated (as a result of Apollo not having visited younger surfaces), but this is the cratering rate that is used, with assumptions, to date cratered surfaces elsewhere in the Solar System (most notably the surface of Mars). Thus, better constraining the lunar cratering rate in this time interval is of importance to planetary science generally, not merely in the context of lunar geology. Moreover, extra-lunar materials collected from regoliths developed on top of lava flows of different ages (and palaeoregoliths trapped between them) will make it possible to determine how the flux and composition of solar wind particles, galactic cosmic ray particles, and meteoritic impacts have varied with time. Finally, although not directly related to the theme of this White Paper, we note that geochemical analysis of these basaltic samples would also yield information on the magmatic history of the Moon, and thus lunar mantle evolution, extending our understanding of lunar geological and thermal evolution to more recent times than is possible using the Apollo and Luna sample collections.

3.2.2.2 South Pole Aitkin Basin sample return

The South Pole Aitkin (SPA) basin is the largest (~2500 km in diameter) and oldest recognized impact basin on the Moon. Its deep structure, which may have sampled the

lunar mantle, and subsequent modification provides a unique sampling site for accessing a record of the Moon's early geological evolution and its impact history (e.g., Duke, 2003; Jolliff et al., 2010). As SPA is the Moon's oldest known impact structure, directly measuring its age would help constrain the bombardment history of the entire inner Solar System, including that of the Earth (e.g. Norman, 2009; Fernandes et al., 2013, and references therein). Moreover, dating SPA, and younger craters and basins within it, will further elucidate the extent to which the early bombardment history of the Earth-Moon system was stochastic (or 'saw-toothed', Morbidelli et al., 2012) with significant implications for the habitability of the early Earth. Last but not least, dating SPA is important because it provides temporal information for the thermal evolution of the lunar crust, and an upper age limit for the addition of a 'late veneer' to the lunar mantle (i.e. later impacts will not deliver volatiles and platinum group elements to the mantle because it will have been sealed by a thick crust).

As discussed in Section 2.1, addressing these questions will require the return of samples from SPA for analysis in laboratories on Earth. The primary mission objective is therefore to return ~1-10 kg (TBC) of lunar regolith from within SPA to determine the age of SPA itself, and ideally also younger craters and basins located within it, SPA is thought to have had a large melt sheet that forms much of the present day floor of the crater, although this melt sheet has been modified by more recent geological processes (magmatism, younger impact basins). Survival of SPA impact melt breccias in present day regolith (after mixing with ejecta from younger impacts) estimated to be ~75-80 % (e.g., Petro and Pieters, 2004), and some regions of the basin preserve this record better than others.

There are a number of suitable landing sites for SPA sample return, however, Schrödinger basin, which is a large impact basin located on the western rim of SPA (centered at 75°S, 132.5°E; Fig. 3) has been identified as particularly ideal site to both sample SPA impact products, and also address other numerous key questions in lunar and planetary science (e.g., Bunte et al., 2011; O'Sullivan et al., 2012; Kring and Durda, 2012). These include dating the age of Schrödinger itself in addition to SPA which, as one of the *youngest* lunar basins (Wilhems, 1987), would further constrain the basin-forming impact chronology. In addition, the floor of Schrödinger contains presumed young pyroclastic deposits (Fig. 3). Not only would sampling these materials provide valuable information on late-stage lunar volcanism, but palaeoregoliths covered by the pyroclastic deposit may contain information on the extra-lunar environment (e.g., solar wind, cosmic rays, meteoritic debris) from a well-defined time horizon.

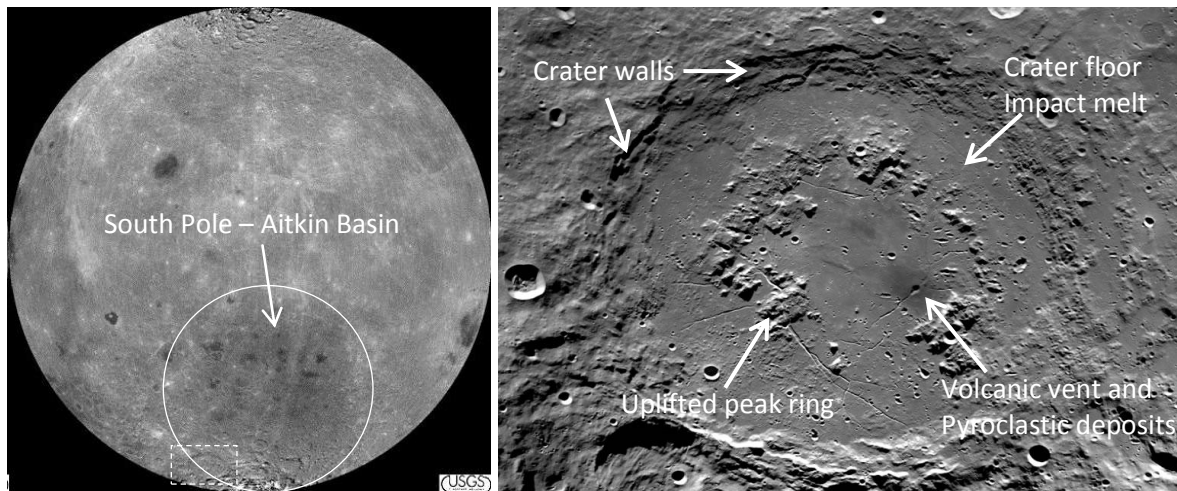


Figure 3. (left) Albedo image of the farside of the Moon. Dashed box shows location of the Schrödinger basin close to the South Pole. (right) Close up albedo map of the Schrödinger basin, which is 315 km in diameter. Some major geological features indicated. .

3.2.3 Sample return mission architecture

3.2.3.1 Moon Near Side Architecture (MNSA)

For the Moon Near Side Architecture, two main strategies can be envisaged and are reported Table 3-2.

Table 3-2: Mission Architecture Elements	
Strategy 1	Strategy 2
Carrier Spacecraft	Lunar Lander
Lunar Lander	Lunar Ascent Vehicle
Lunar Ascent Vehicle	

Both strategies nominally rely on a Soyuz-Fregat (TBC) launch to perform the injection to Geostationary Transfer Orbit (GTO), but the possibility of using a more powerful launch vehicle (e.g., Ariane 5) will be investigated in future studies with a view to enhancing the capabilities of the landed elements. The two strategies differ with regards to the mission element performing the transfer from Earth to Low Lunar Orbit (LLO, about 100-150 km altitude) as well as the return journey. In fact, for the first strategy, the transfer in both directions is performed by a Carrier Spacecraft, while in the second case the Earth to Moon transfer is performed by the Lunar Lander and the return journey by the Lunar Ascent Vehicle itself (accommodating also the Earth Re-Entry Vehicle). Both strategies will be thoroughly traded-off during future mission studies. As example, here only one will be discussed more in detail.

Strategy 1 foresees that, once delivered to GTO, the Carrier Spacecraft (CS) will carry the Lunar Lander (LL) and the Lunar Ascent Vehicle (LAV) to LLO. Therefore, the LL will separate from the CS and will descend using a dedicated chemical propulsion stage. Once on surface, lunar samples, for a total mass in the order of 1-10 kg (TBC), will be collected and stored into the LAV. The mode of sampling will be assessed during future studies, but we currently envisage a sieved and/or cored sample of regolith containing mm to cm-sized 'rocklets' suitable for dating and mineralogical and geochemical analyses; one or more core samples (depth TBD) would also provide valuable stratigraphic information about the regolith and implanted volatiles. Possible mobility requirements will also be assessed during future studies and could involve either a rover or a hopper capable of multiple landings. The final choice will depend on several factors among which: available launch vehicle, mass and power resources, and sampling site(s) location(s). During surface operations, the CS orbiting around the Moon will deliver a set of Penetrators in predefined locations to enable further scientific investigation



After completion of sample acquisition and storage, the LAV will take off from the LL leaving behind the sampling equipment, the landing stage, and some scientific instruments (Fig. 4). Once in lunar orbit, the sample container will be ejected, captured by the CS and transferred into the Earth Re-Entry Vehicle. The CS will capture the sample container and use the same propulsion system as used for the outward journey to return to Earth.

Figure 4. The LAV takes off from the lunar surface (Astrium)

A preliminary assessment of this Architecture allows for a LL of 665 kg (including 384 kg of propellant), LAV of 145 kg (including 66 kg of propellant), and a CS of 1937 kg (including 1509 kg of propellant for the two transfers, and 206 kg for four Descent Modules transporting the penetrators). More accurate mass break downs can be determined once the landing site and staging strategy are analysed during future mission studies. Trade-offs on the overall architecture will also need to be performed in the next study phases and might result in significant mass savings or enhanced mission capability. Last but not least, it has to be noted that such a mission will also directly demonstrate key technologies for Mars Sample Return (MSR), as the proposed Architecture is very similar.

3.2.3.2 Moon Far Side Architecture (MFSA)

Also for this Architecture, two possible strategies can be envisaged and they are reported in Table 3-3 below. As the lunar farside is never visible from Earth a relay element, which could be either located in LLO or in EML-2, is needed in order to support communications.

Table 3-3: Mission Architecture Elements	
Strategy 3	Strategy 4
Carrier/Orbiter Spacecraft	Deployed human spaceflight infrastructure in Earth-Moon Lagrange Point 2 (EML-2)
Lunar Lander	Lunar Lander
Lunar Ascent Vehicle	Lunar Ascent Vehicle

The strategy 3, is similar to the one presented for the MNSA, with a CS aimed at performing the transfer from Earth to Moon and return as well as delivering the set of Penetrators, but also providing communication services between the surface elements and the Ground Station(s), a LL descending on lunar surface, and a LAV hosting the collected samples and bringing them back to orbit, where they will be transferred into the Earth Re-Entry Vehicle of the CS to be transported to Earth.

Of significant interest is also the strategy currently under investigation from NASA (Alkalai et al., 2012) which foresees the exploitation of a human spaceflight infrastructure in EML-2 to perform the rendezvous with the orbiting LAV and make easier the securing of the sample container. In fact, the implementation of such an approach could be advantageous because:

- The propellant mass required to return to Earth would be saved;
- The Orbiter would be not needed, as the communications with Earth could be enabled via the EML-2 infrastructure.
- The Earth Re-Entry Vehicle would be not needed, as the Sample Container would be secured in the EML-2 infrastructure;
- Teleoperations from EML-2 to lunar surface could be performed, this increasing the mission success probability;
- Owing to the saved mass, the LAV could be bigger and accommodate a larger quantity of samples, increasing the scientific return.

As for the MNSA, both the presented strategies will be investigated during future studies (including consideration of a more powerful launch vehicle), in order to identify the associated benefits and risks.

3.2.3.3 Surface mobility requirements

In order to address the top-level science questions it is essential that samples be collected from a diverse range of localities separated by tens, or even hundreds, of km (see Figs. 2 and 3 for the scale of separation of geological units in Procellarum and Schrödinger, respectively). A rover capable of collecting rock and soil samples from a radius of a few tens of km from the landing site would be extremely valuable from this point of view. Surface mobility would also enable the sampling of ejecta from small craters that will have naturally excavated to a range of depths below the surface, thereby providing important stratigraphic information. In addition to collecting samples, and transferring them to the sample return vehicle, such a rover could be instrumented to obtain contextual information for the samples (e.g., by multi-spectral imaging, *in situ* mass spectrometry, X-Ray fluorescence/diffraction, and/or Raman-LIBS instruments). In addition, a rover could use ground penetrating radar to image shallow subsurface structure (e.g., a 1 GHz radar, easy to accommodate on a small rover, could determine internal regolith structure, to a depth of about 2 m), which would help with sample site selection/local context and regolith depth/age determination. We note that even a smaller range rover would be useful to support sample collection from outside areas contaminated by the landing, which would be especially important when considering samples containing volatiles. The Mobile Payload Element, designed in the context of ESA's proposed Lunar Lander (Haarmann et al., 2012), provides an example of a small (~14 kg) autonomous and innovative rover that could satisfy this requirement.

Despite the advantages of a rover, for some of the scientific objectives outlined above (especially in the Oceanus Procellarum mission case) a sample collection range of 50 to 100 km might be preferable. As this may be beyond the practical range of a rover that could be landed within the mass constraints, we propose that the possibility of having the lander 'hop' to multiple locations separated by tens or hundreds of km will be investigated as part of an industrial pre-phase A study of the sample return architecture.

Accessing palaeoregolith deposits, either trapped between lava flows in Oceanus Procellarum, or beneath pyroclastic deposits in Schrödinger, may require a drilling capability to be included. Determining whether or not drilling will actually be required, and if so the probable depth, will depend on whether plausible palaeoregolith outcrops can be identified in high-resolution Lunar Reconnaissance Orbiter Camera (LROC) images (see discussion by Crawford et al., 2009). The practicality of including a drilling capability will likewise be studied during an industrial pre-Phase A study should the mission concept be deemed worthy of further study.

4 Conclusions

We have proposed a Cosmic Vision 'Science Theme' of using lunar science as a window into the early history of the Solar System. The near surface lunar environment contains a rich

record of inner Solar System history. Accessing this record will directly address key elements of ESA's Cosmic Vision themes, especially Theme 1 ('Planets and Life') and Theme 2 ('How does the Solar System Work?'). Implementation will require spacecraft to land on the lunar surface in order to make *in situ* measurements at, and/or return samples from, localities that have been carefully selected with specific scientific objectives in mind.

For the Cosmic Vision L2/3 mission opportunities we propose the development of a scientific infrastructure that would enable us to address these scientific objectives. There are two, mutually complementary, strands: (i) a mission based around multiple penetrators for the characterisation of lunar polar volatiles and (ii) a sample return mission to address the lunar impact chronology and records of the near-Earth Solar System environment preserved in regolith deposits. We consider that the development of such an ambitious lunar science architecture is worthy of careful consideration for the Cosmic Vision L2/3 mission opportunities.

References

- Alkalai, L. et al., 2012. ORION/MOONRISE: Joint human-robotic lunar sample return mission concept, LEAG Meeting, GSFC.
- Anand, M., 2010. Lunar water: a brief review. *Earth Moon Planets*, 107, 65-73.
- Armstrong, J.C., 2010. Distribution of impact locations and velocities of Earth meteorites on the Moon. *Earth Moon Planets*, 107, 43-54.
- Armstrong, J.C., Wells, L.E. and Gonzales, G., 2002. Rummaging through Earth's attic for remains of ancient life. *Icarus*, 160:183-196.
- Bottke, W.F., et al., 2012. An Archaean heavy bombardment from a destabilized extension of the asteroid belt, *Nature*, 485, 78-81.
- Brock, L.S., Melosh, H. J., 2012. Impact Exchange of Material Between Planets of Gliese 581, LPSC, 43, 2467.
- Bunte, M. K., et al., 2011. A sortie mission to Schrödinger Basin as reconnaissance for future exploration. *GSA Special Papers* 483, 533-546 doi: 10.1130/2011.2483(32).
- Chyba, C.F., Sagan, C., 1992. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature*, 355, 125-132.
- Chaussidon, M., Robert, F., 1999, *Nature* 402, 270-273.
- Colaprete, A., et al., 2010. Detection of water in the LCROSS ejecta plume. *Science*, 330, 463-468.
- Cockell, C.S., 2010. Astrobiology – what can we do on the Moon? *Earth, Moon Planets* 107, 3-10.
- Crawford, I.A., 2006. The Astrobiological Case for Renewed Robotic and Human Exploration of the Moon. *Internat. J. Astrobiology*, 5, 191-197.
- Crawford I.A., Fagents S.A., and Joy K.H. 2007. Full Moon Exploration: valuable (non-polar) lunar science facilitated by a return to the Moon. *Astronomy and Geophysics*, 48: 3.18–3.21.
- Crawford, I.A., Joy, K.H., Fagents, S.A., Rumpf, M.E., 2009. The importance of lunar palaeoregolith deposits and the role of Lunar Reconnaissance Orbiter, LRO Targeting Meeting, Tempe, Arizona; Abstract # 6007.
- Crawford, I.A., Baldwin, E.C., Taylor, E.A., Bailey, J. and Tsembelis, K., 2008. On the survivability and detectability of terrestrial meteorites on the Moon. *Astrobiology*, 8, 242-252.

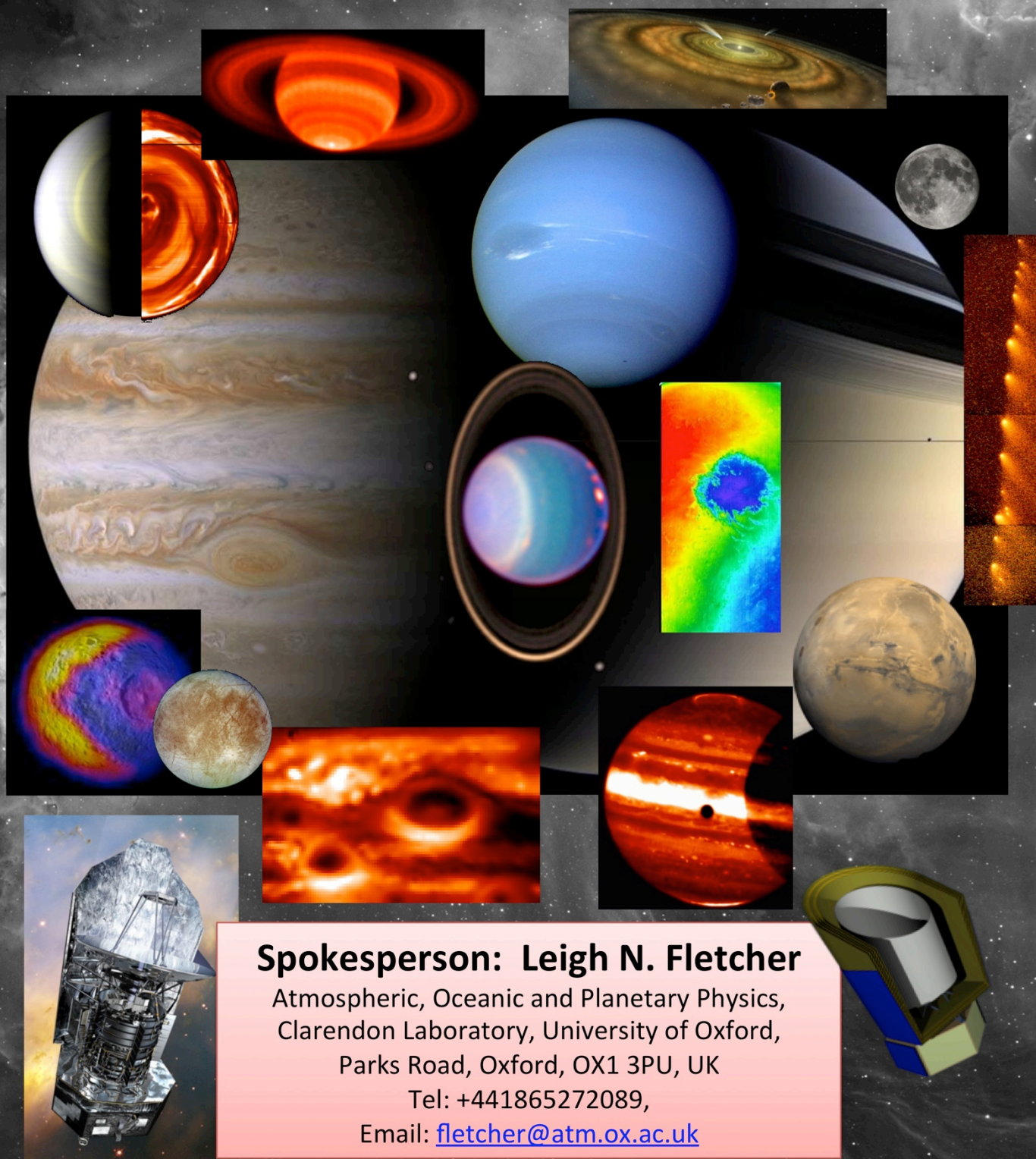
- Crawford, I.A., Fagents, S.A., Joy, K.H., Rumpf, M.E., 2010. Lunar palaeoregolith deposits as recorders of the galactic environment of the Solar System and implications for astrobiology. *Earth Moon Planets*, 107, 75-85.
- Crawford, I.A., et al., 2012. Back to the Moon: the scientific rationale for resuming lunar surface exploration. *Planet. Space Sci.*, 74, 3-14.
- Crites, S.T., et al., 2011. In-situ production of organic molecules at the poles of the Moon, AGUFM.P13D1730C.
- Gomes, R., et al., 2005. Origin of the cataclysmic late heavy bombardment period of the terrestrial planets, *Nature*, 435, 466-469.
- Crozaz, G., et al., 1977. The record of solar and galactic radiations in the ancient lunar regolith and their implications for the early history of the Sun and Moon. *Phil. Trans. R. Soc.*, A285, 587-592.
- De Vera, J.-P., et al., 2012. Supporting Mars exploration: BIOMEX in Low Earth Orbit and further astrobiological studies on the Moon using Raman and PanCam technology; *Planet. Space Sci.*, 74, 103-110.
- Duke M.B. 2003. Sample return from the lunar South Pole-Aitken Basin, *Advances in Space Research* 31, 2347-2352
- Fagents, S.A., Rumpf, M.E., Crawford, I.A. and Joy, K.H., 2010. Preservation potential of implanted solar wind volatiles in lunar paleoregolith deposits buried by lava flows. *Icarus*, 207, 595-604.
- Feldman, W.C., et al., 1998. Fluxes of fast and epithermal neutrons from Lunar Prospector: evidence for water ice at the lunar poles. *Science*, 281, 1496-1500.
- Fernandes, V.A., et al., 2013. The bombardment history of the Moon as recorded by ^{40}Ar - ^{39}Ar chronology, *Meteoritics and Planetary Science*, 48, 241-269.
- Gutiérrez, J.L., 2002. Terrene meteorites in the moon: relevance for the study of the origin of life in the Earth. ESA SP-518, 187 – 191.
- Haarmann, R., Jaumann, R., et al., 2012. Mobile Payload Element (MPE): Concept study for a sample fetching rover for the ESA Lunar Lander Mission, *Planet. and Space Sci.*, 74, 283-295.
- Hashizume, K., et al., 2000. Solar Wind Record on the Moon: Deciphering Presolar from Planetary Nitrogen, *Science*, 290, 1142-1145.
- Hiesinger, H., Head, J.W., 2006. New views of lunar geoscience: an introduction and overview. *Rev. Min. Geochem.*, 60, 1-81.
- Hiesinger, H.J. et al., 2003. Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum, *J. Geophys. Res.*, 108 (E7), 1-27.
- Hiesinger, H., van der Bogert, C.H., Pasckert, J.H., Funcke, L., Giacomini, L., Ostrach, L.R., Robinson, M.S., 2012. How old are young lunar craters? *J. Geophys. Res.*, 117, E00H10.
- Jaumann, R., et al., 2012. Geology, geochemistry and geophysics of the Moon: status of current understanding. *Planet. Space Sci.*, 74, 15-41.
- Jester, S., Falcke, H., 2009. Science with a lunar low frequency array: from the dark ages of the universe to nearby exoplanets. *New Astronomy Reviews*, 53, 1-26.
- Johnson, B.C., et al., 2012. A self-consistent model of the circumstellar debris created by a giant hypervelocity impact in the HD 172555 system *Astrophysical Journal*, 761, Issue 1, article id. 45.
- Jolliff, B. L., et al., 2010. MoonRise: Sampling South Pole-Aitken Basin as a Recorder of Solar System Events. American Geophysical Union, Fall Meeting 2010, abstract #P43A-01.
- Joy, K.H., et al., 2011. Re-examination of the formation ages of the Apollo 16 regolith breccias. *Geochimica et Cosmochimica Acta*, 75, 7208-7225.
- Joy, K.H., et al., 2013. Direct detection of projectile relics from the end of the lunar basin-forming epoch. *Science*, 336, 1426-1429.
- Kring, D.A., 2003. Environmental consequences of impact cratering events as a function of ambient conditions on Earth, *Astrobiology*, 3, 133-152.
- Kring, D.A., Durda, D.D., eds., 2012. A Global Lunar Landing Site Study to Provide the Scientific Context for Exploration of the Moon, LPI Contribution. 1694, LPI, Houston, TX, 688 pp.

- Liu, Y. et al., 2012. Direct measurement of hydroxyl in the lunar regolith and the origin of lunar surface water, *Nature Geoscience*, 5, 779-782.
- Lucey, P.G., 2000. Potential for prebiotic chemistry at the poles of the Moon. *Proc. SPIE*, 4137, 84-88.
- Lucey, P.G., et al., 2006. Understanding the lunar surface and space-Moon interaction. *Rev. Min. Geochem.* 60, 82-219.
- Maher, K.A., Stevenson, D. 1988. Impact frustration of the origin of life, *Nature*, 331, 612-614.
- McKay, D.S., et al., 1991. The lunar regolith. In: Heiken, G.H., Vaniman, D. and French, B.M. (Eds.), *The Lunar sourcebook: A user's guide to the Moon*, Cambridge University Press, pp. 285-356.
- Morbidelli, A., et al., 2012. A sawtooth-like timeline for the first billion years of lunar bombardment, *Earth Planet. Sci. Lett.*, 355, 144-151.
- Neukum, G., Ivanov, B. A., Hartmann, W. K., 2001. Cratering Records in the Inner Solar System in Relation to the Lunar Reference System. *Space Science Reviews*, 96, 55-86.
- Norman M. D. 2009 The Lunar Cataclysm: Reality or "Mythconception"? *Elements*, Vol. 5, 23-28.
- Petro, N. E., and Pieters C. M., 2004. Surviving the heavy bombardment: Ancient material at the surface of South Pole-Aitken Basin, *J. Geophys. Res.*, 109, E06004, doi:10.1029/2003JE002182.
- Mizutani, H., et al., 2005. Lunar-A mission: outline and current status., *J. Earth System Science*, 114, 763-768.
- NRC, 2007. *The Scientific Context for Exploration of the Moon*. National Research Council, National Academies Press, Washington DC.
- O'Sullivan K.M., et al., 2011. Calibrating several key lunar stratigraphic units representing 4 billion years of lunar history within Schrödinger Basin. In *Recent Advances in Lunar Stratigraphy*, D.A. Williams and W. Ambrose (eds.), pp. 117-128, *Geol.Soc.Am. Special Paper 477*, Boulder, CO.
- Ozima, M., et al., M., 2005. Terrestrial nitrogen and noble gases in lunar soils. *Nature* 436:655-659.
- Ozima, M., et al., 2008. Toward prescription for approach from terrestrial noble gas and light element records in lunar soils understanding early Earth evolution. *Proc. Nat. Acad. Sci.* 105:17654-17658.
- Pierazzo, E., Chyba, C.F., 1999. Amino acid survival in large cometary impacts. *Meteorit. Planet. Sci.*, 34, 909-918.
- Pieters, C.M. et al., 2009. Character and spatial distribution of OH/H₂O on the surface of the Moon seen by M³ on Chandrayaan-1. *Science*, 326, 568-572.
- Rumpf, M.E., Fagents, S.A., Crawford, I.A., Joy, K.H., 2013. Numerical modeling of lava-regolith heat transfer on the Moon and implications for the preservation of implanted volatiles", *Journal of Geophysical Research (Planets)*, 118, 382-397.
- Ryder G., 1990. Lunar samples, lunar accretion and the early bombardment of the Moon. *EOS* 71, 313-323 .
- Ryder, G., 2003. Bombardment of the Hadean Earth: wholesome or deleterious? *Astrobiology*, 3, 3-6.
- Sleep N.H., Zahnle K.J., Kasting K.F. and Morowitz H.J., 1989. Annihilation of ecosystems by large asteroid impacts on the early Earth, *Nature*, 342, 139-142.
- Smith, A., et al., 2012. Lunar Net—a proposal in response to an ESA M3 call in 2010 for a medium sized mission, *Experimental Astronomy*, 33, 587-644.
- Spudis, P.D., 1996. *The Once and Future Moon*. Smithsonian Institution Press, Washington D.C.
- Spudis P. D. and Lavoie A. R. 2011. Using the resources of the Moon to create a permanent, cislunar space faring system. AIAA SPACE 2011 Conference & Exposition. AIAA 2011-7185.
- Stöffler, D., et al., 2006. Cratering History and Lunar Chronology, *Rev. Min. Geochem.*, 60, 519 – 596.
- Wieler, R., Kehm, K., Meshik, A.P., Hohenberg, C.M., 1996. Secular changes in the xenon and krypton abundances in the solar wind recorded in single lunar grains. *Nature*, 384:46-49.
- Wilhelms, D.E. 1987. *The Geologic History of the Moon*, USGS Professional Paper No. 1348.
- Zhang, J.A., Paige, D.A. 2009. Cold-trapped organic compounds at the poles of the Moon and Mercury: implications for origins, *Geophys. Res. Lett.*, 36, L16203.

Exploring Planetary Origins and Environments in the Infrared

A PLANETARY SCIENCE INFRARED OBSERVATORY (PSIO)

A White Paper Response to ESA's Call for L-Class Science Themes



Contributors: Leigh N. Fletcher (University of Oxford); Patrick G.J. Irwin (University of Oxford); Bruce Swinyard (University College London); Henrik Melin (University of Leicester); Sarah Badman (University of Leicester); James O'Donoghue (University of Leicester); Sarah Casewell (University of Leicester); Neil Bowles (University of Oxford); Damien Weidman (RAL Space, STFC Rutherford Appleton Laboratory); Miriam Rengel (Max Planck Institute for Solar System Research)

List of Supporters: Richard Achterberg (University of Maryland Department of Astronomy); Matt Burleigh (University of Leicester); Thibault Cavalie (Laboratoire d'Astrophysique de Bordeaux); David S. Choi (ORAU/NASA Goddard Spaceflight Center); Stanley W H Cowley (University of Leicester); John Davies (UKATC); Michael Davis (Southwest Research Institute); Alan Fitzsimmons (Queens University Belfast); Thierry Fouchet (Observatoire de Paris); Davide Grassi (IAPS-INAF Rome); Thomas Greathouse (Southwest Research Institute); Jane Hurley, (University of Oxford); Ranah Irshad (Rutherford Appleton Laboratory); Caitriona Jackman (University College London); Remco de Kok (SRON); Stephen Lewis (The Open University); Timothy Livengood (NASA/Goddard Spaceflight Center); Adam Masters (JAXA); Raul Morales-Juberias (New Mexico Tech); Julianne Moses (Space Science Institute); Olivier Mousis (Institut UTINAM); Conor Nixon (NASA Goddard Space Flight Center); Glenn Orton (Jet Propulsion Laboratory, California Institute of Technology); Enzo Pascale (University of Cardiff); Santiago Perez-Hoyos (Universidad del Pais Vasco UPV/EHU); Kurt D. Retherford (Southwest Research Institute); Sebastien Rodriguez (Université Paris Diderot); Agustin Sanchez-Lavega (Universidad del Pais Vasco); Amy Simon-Miller, (NASA GSFC); James Sinclair (University of Oxford); Tom Stallard (University of Leicester); Nicholas Teanby, (University of Bristol); Constantine Tsang (Southwest Research Institute); Sandrine Vinatier (Observatoire de Paris-Meudon); Colin Wilson (University of Oxford)

Table of Contents

Executive Summary.....	3
1. Motivation and Background	3
Building on our Infrared Heritage	4
Enhancing our Infrared Future	5
2. Outline: A Infrared Observatory for Planetary Astronomy.....	5
3. Science Themes: Understanding Planetary Systems.....	7
Theme I: Origins.....	7
Theme II: Atmospheres.....	8
Theme III: Surfaces.....	11
Theme IV: Interaction.....	13
4. Observatory Architectures.....	15
5. Conclusion.....	18
References.....	19

Exploring Planetary Origins and Environments in the Infrared:

A Planetary Science Infrared Observatory (PSIO)

WHITE PAPER RESPONSE TO ESA CALL FOR LARGE-CLASS SCIENCE THEMES

Executive Summary

The discovery of large numbers of extrasolar planets has shown that the formation of diverse planetary systems is a common phenomenon throughout our galaxy, and yet many of the fundamental questions about the origins, evolution and environmental conditions within our own solar system remain unanswered. We propose an **observatory-class ESA mission to provide spatially resolved infrared spectroscopy of solar system and planetary objects in all their guises**, from their **origins** (remaining debris in our solar system and planet-forming discs around other stars) to their present-day appearance (**atmospheres, surfaces and interactions** with their host stars for planets in our solar system and beyond). Discovery level science will be achieved by: combining broadband spectral coverage from 3-1000 μm with high spatial resolutions from a single mirror or distributed array; long time-baseline observations for evolving planetary processes and time-domain solar system science; and selected high-spectral resolution observations to probe regimes never previously explored by observatories or visiting spacecraft. Although this white paper focuses on the infrared exploration of our Solar System to address themes at the heart of Europe's Cosmic Vision, such a facility would be of immense value to the broader astrophysical community. In particular the thermal-IR will allow probing of cooler transiting exoplanets existing within the habitable zone and low temperature brown dwarfs. The intention of this white paper is to ensure that ESA's future cornerstone missions, either as observatories or as visiting spacecraft, **retain infrared solar system observations at the core of their scientific objectives**, and to advocate for investment in the critical European technologies for high spectral and spatial resolution infrared spectroscopy.

Requirements for Planetary Science in the Infrared

- Observatory-class facility offering broadband thermal infrared spectral coverage from 3-1000 μm to address planetary science questions throughout our Solar System.
- Imaging spectroscopy at moderate resolution ($R \sim 10^3$) to map environmental characteristics on solar system targets and relate to visible-light images.
- High resolution spectroscopy in narrow selectable spectral ranges ($R > 10^6$) to probe unexplored parameter regimes for surfaces and atmospheres.
- Optimised for long duration (5+ years) to allow near-continuous monitoring of individual targets with regular observing campaigns for time-domain planetary astronomy.
- 3-5m class monolithic mirror or distributed array architectures to provide spatial resolutions 0.25-5.0" across the mid/far-IR.

Box 1 Summary of the requirements for an L-class observatory dedicated the exploring planetary systems in all their guises.

1. Motivation and Background

Infrared spectroscopy of thermal emission is the primary tool for investigating the environmental characteristics of surfaces, dust and atmospheres on objects in our solar system and beyond, and yet numerous space missions have overlooked this spectral range over the past decades in favour of reflectance spectroscopy at shorter wavelengths. This must change for future ESA cornerstone missions, particularly with the relentless improvements in the capabilities of ground-based astronomers to provide Hubble-quality visible/near-infrared imaging of our planetary neighbours. A space-based observatory is essential for thermal infrared science, to remove the obscuring influence of terrestrial atmospheric contaminants (particularly water, CO_2 , methane and

O₃), whereas short-wavelength science will be readily provided from the ground in the coming decades, limiting the necessity for missions to include reflected sunlight experiments. In this white paper we describe the scientific advances that could be provided by a dedicated thermal infrared observatory for planetary science, utilising thermal emission across a broad spectral range spanning the capabilities of ISO, JWST, Spitzer, Herschel and SPICA from a single, agile space telescope dedicated to planetary observations. This would provide **observations of multiple objects throughout a planetary system, from the smallest building blocks to the largest planets**. Such a facility would be of high value for targets both within and outside of our solar system, studying planetary formation and planetary environments in all of their diverse guises.

Why the thermal infrared? The thermal infrared encompasses many molecular (absorption and emission) and solid-state (ice absorption, dust emission/absorption) features. The strength of a feature depends upon the molecular abundances and the temperature of the line-forming region, in addition to the presence of continuum absorbers such as aerosols or dust. Deriving physical properties from these observations is inherently degenerate (for example, measuring temperatures using assumptions about the abundances of CO₂ or CH₄), and we must therefore **capture as broad a spectral range as possible** to unambiguously separate the different contributions. In the gaseous phase, middle atmospheric emission (stratosphere/mesosphere) features appear as narrow Doppler-broadened lines requiring high spectral resolution to resolve their line-shape and determine temperatures and composition. Tropospheric lines, conversely, are pressure-broadened and require lower-resolution spectroscopy. Solid-state features such as astrophysical silicate absorption, or water ice bands on satellite surfaces, can have broad features that render them difficult to identify without broadband spectroscopy. **An infrared observatory must therefore be adaptable, featuring both moderate resolution imaging spectroscopy ($R \sim 10^3$) and the capability for high spectral resolutions ($R > 10^6$)** to sample as broad a range of line-forming environments as possible.

Why dedicated to planetary observations?

Previous infrared remote sensing has either relied on filtered photometric imaging in a small number

of discrete wavelengths within the infrared windows or point spectroscopy that needs to be scanned over targets to create an image. Planetary astronomy needs: a) **long-term continuous observations of planetary phenomena, simultaneously across a broad wavelength range and capturing the spectra of all points on a planet simultaneously**, with a similar spatial resolution; and b) the need for an **agile, responsive platform able to move quickly to observe events of interest in our dynamic solar system**. Despite several decades of planetary astronomy, our knowledge of these environments comes from simple, isolated snapshots, whose frequency cannot be tuned to the timescales of interest (e.g., rapid-scale impacts or meteorological events to long-term seasonal evolution). Furthermore, comparative planetology is hampered when a single target is studied in detail at a mission rate of typically no more than once per decade with visiting spacecraft. A facility that could **observe all targets with the same instrumental capabilities** would satisfy a broader swathe of the community than a single targeted mission. Finally, the **observational requirements of the deep sky astronomical community are rather different to those of planetary observers**, who desire: large fields of view (an arc minute for Venus and Jupiter); broadband spectroscopy of varying spectral resolution; enhanced spatial resolution and the ability to view the brightest targets; and the capability to observe a target for a long period of time, whilst responding quickly to unique events. No observatory presently planned can meet these observational requirements.

Building on our Infrared Heritage

Despite the great diagnostic value of the thermal infrared, this spectral range has been utilised on relatively few planetary missions. Venera 15/16 were the last to study Venus' thermal-infrared spectrum three decades ago; and only two capable long-wavelength spectrometers have been sent to study the outer solar system, namely IRIS on the twin Voyager spacecraft and CIRS on the Cassini mission. These delivered a huge change in our understanding of atmospheric climate, circulation and chemistry on the giant planets and the properties and endogenic activity on icy satellites and rings, but such studies are not destined to be repeated in the coming decades. Jupiter, for example, has never been observed by an orbital mission with good infrared capabilities: Galileo had a very simple thermal imager which was ultimately

limited to only two spectral channels; Juno does not have capabilities longward of 5 μm ; and JUICE will not feature any thermal infrared instrumentation. Without maps of the evolving temperatures, wind shears, humidity (e.g., ammonia) and clouds, our understanding of Jupiter's meteorology and climate will be woefully incomplete. We advocate that ***thermal infrared science should be a primary component of any future ESA cornerstone mission*** in our solar system, but that a space-borne planetary observatory would provide enhanced science to all missions planned or in flight.

Today, planetary astronomy is being driven both by space observatories and ground-based facilities, although atmospheric variability, particularly in the water vapour column, prevents accurate radiometric calibration, and regions contaminated by strong telluric features are largely unusable from the ground. The Stratospheric Observatory for Infrared Astronomy (**SOFIA**, 2.5-m mirror with a suite of instruments covering the 5-240 μm range) overcomes some of these obstacles by flying at high altitude, although its observing capabilities are limited by competition, flight paths and the stability of the residual atmosphere above the aircraft. ALMA will provide unprecedented results beyond 300 μm , but will still be limited by telluric contamination. ***Hence, access to the mid- to far-IR spectral ranges and well-calibrated data requires a space-borne platform.*** ESA's Infrared Space Observatory (**ISO**, 2.3-240 μm with a 0.6-m primary mirror and a maximum $R \sim 30,000$), **Spitzer** (0.85-m primary mirror, 3-180 μm and a maximum $R \sim 600$), **AKARI** (0.67-m primary mirror, 1.8-180 μm with $R \sim 135$ at short wavelengths) and **Herschel** (3.5-m primary mirror, spanning 55-672 μm with heterodyne spectroscopy at the longest wavelengths) revolutionised our understanding of planetary conditions in our solar system, but were limited to disc-averaged snapshot observations of all targets. None of these observatories could provide heterodyne resolutions in the mid-infrared; none could provide imaging spectroscopy; none combined mid and far-IR in a single instrument chain; and none could provide the dedicated solar system observations required to address the science case below.

Enhancing our Infrared Future

In the coming decade, solar system science should form a substantial component of both the James Webb Space Telescope (**JWST**, with a 6.5-m primary mirror and a resolution up to $R \sim 3730$ with

the 5-28 μm MIRI instrument) and Space Infra-Red Telescope for Cosmology and Astrophysics (**SPICA**, 5-210 μm with a 3.2-m primary mirror). Although neither observatory is optimised for planetary studies (for example, the integral field units of JWST/MIRI's have a small FOV, ranging from 3.6" at 5 μm to 7.6" at 28 μm , and will require substantial mosaicking to image Jupiter), and neither will be able to provide long-term observations of a single target due to intense competition with deep sky phenomena, several solar system targets will be within their grasp. Mars and Venus will be beyond the reach of JWST beyond 5 μm , but several modes will be available for the giant planets (Jupiter can only be viewed at wavelengths shorter than 10 μm due to brightness limits), comets and cool rocky bodies (Lunine et al., 2010). SPICA is expected to survey a large number of Kuiper Belt Objects (KBOs) and Trans-Neptunian Objects (TNOs) in the far-IR, detect zodiacal dust clouds out to tens of parsecs, and study the roles of water and dust in planet-forming discs (Tamura et al., 2009, SPICA Workshop).

However, it is unlikely that either of these observatories could dedicate significant time to the solar system, and neither provide coverage of both the mid-IR and far-IR/sub-mm with the tunable spectral resolution from moderate (10^3) to heterodyne (10^7) as proposed here. To complement JWST and SPICA and provide a new compelling platform for the astrophysical community, this proposal seeks to add (i) long baseline observations with higher spectral resolutions in the mid-IR (3-30 μm); (ii) high spatial resolution (and wide FOV) from a distributed array and high spectral resolution in the far-IR (30-300 μm) and sub-mm (300-1000 μm); and (iii) the agility to respond quickly and track any new phenomena observed within our solar system.

2. Outline: A Infrared Observatory for Planetary Astronomy

The science case outlined below can be summarised as follows:

Understanding the processes responsible for shaping planetary systems and the diverse environments found throughout our solar system.

The science case is sub-divided into four themes – Origins, Atmospheres, Surfaces, and Interactions, and requires a program dedicated to planetary systems at all stages in their evolutionary history, from planet-forming discs, to remnant debris of the earliest accretion stages, to end products of the formation process manifested as the bewildering array of planetary objects we see today. The aim of this white paper is to make the case for long-term thermal infrared (3-1000 μm) observations of planetary environments, both within our solar system and around other stars. The basic architecture of an observatory should achieve all or a subset of the following goals:

- **Enhanced Spectral Range:** Access to the 3-1000 μm spectral range with broadband moderate resolution imaging spectroscopy ($R \sim 10^3$) and tunable heterodyne narrowband spectroscopy in narrow ranges ($R > 10^6$).
- **Time-Domain Science:** A flexible and agile platform able to respond quickly to new and interesting events, and to observe single targets with a time sampling tuned to the phenomenon of interest (e.g., tracking atmospheric features over hours, planetary impacts, volcanism/cryovolcanism, or seasonal evolution over years).
- **Superb Spatial Resolution:** Diffraction-limited system at 5 μm , providing spatial resolution in the mid-IR from a single mirror (3 to 5-m class) and in the far-IR/sub-mm from a distributed array.

Such an observatory would build on ESA's heritage from ISO and Herschel, complement flight missions that do not have capabilities beyond 5 μm , and move beyond the planetary science capabilities of JWST and SPICA. Unlike spacecraft dedicated to a single target, an observatory is capable of addressing compelling scientific questions for multiple planetary objects. And by moving beyond single snapshots, we will open up the field of time-

Target	Diameter	Target	Diameter
Venus	66"	Neptune	2.4"
Mars	25.1"	KBO/Pluto	0.11"
Ceres	0.84"	Io	1.2"
Vesta	0.64"	Ganymede	1.8"
Jupiter	50.1"	Titan	0.8"
Saturn	20.1"	Triton	0.13"
Uranus	4.1"		

Table 2 Maximum angular size of various solar system targets as viewed from Earth, measured in arcsec. For most examples, this is the maximum angular size at opposition. This should be compared to Table 1 to assess the capabilities of resolving these targets at infrared wavelengths.

domain planetary science, analogous to the enormous leap from still photography to moving pictures, and permit pioneering breakthroughs in the field analogous to today's monitoring of terrestrial atmospheric phenomena and solar variability.

Infrared observing has been historically difficult, requiring cold telescopes, cold instruments and large apertures to enhance spatial resolution, and in section 4 we describe pathways to achieving these objectives. We envisage a 3.5-5.0 m primary for this observatory and for reference, Table 1 shows how the diffraction limited spatial resolution of a range of mirror sizes varies with wavelength. This can be compared to the observed angular sizes of a range of targets in Table 2.

Although this white paper focuses on advances in our understanding of the solar system, such an **observatory must also be cross-disciplinary and highly beneficial for the study of extrasolar systems**, particularly for cooler planets orbiting within habitable zones, as their peak emission will occur at longer wavelengths than the hot Jupiters and Neptunes that are the primary target of currently-planned exoplanet characterisation

Telescope	ISO	EChO	Herschel	Palomar	JWST	Keck
Wavelength (μm)	0.6	1.5	3.5	5	6.5	10
5	2.1"	0.84"	0.36"	0.25"	0.19"	0.13"
10	4.2"	1.6"	0.72"	0.50"	0.39"	0.25"
25	10.5"	4.2"	1.8"	1.3"	0.97"	0.63"
100	41.9"	16.8"	7.2"	5.0"	3.9"	2.5"
1000	419.4"	167.8"	72.0"	50.3"	38.7"	25.2"

Table 1 Diffraction limit for a range of observatory primary diameters (in metres) in the 5-1000 μm range (numbers are in arcseconds, for comparison with object angular diameters). Note that any Earth-based observatory (such as Keck) would suffer from seeing fluctuations in the range of 0.4-1.0 arcsec. The angular resolution of the 3.5-5m class observatories should be compared to the targets listed in Table 2 to give an idea of the proposed capabilities of PSIO for each target.

missions. However, exoplanet science places enormous constraints on telescope stability and detector sensitivity, and these science targets are the subject of other white papers. Our purpose is not to define a single observatory that is capable of addressing each of these science goals equally, but rather to demonstrate the unique science that can be achieved in this spectral range, and why it should form a crucial part of any future observatory or spacecraft mission.

3. Science Themes: Understanding Planetary Systems

In the following sections, we identify the key science questions that spatially-resolved infrared spectroscopy seeks to address, related to ESA's cosmic vision objectives - *what are the conditions for planet formation and the emergence of life; and how does the solar system work?*

Theme I: Origins

Key Question: What processes govern planetary formation, the architecture of our planetary system and the evolution of habitable environments?

The present-day architecture of our solar system is the end product of a variety of processes that we are only beginning to understand. Theory suggests that early accretion of planetary building blocks, composed of varying amounts of rock and ice depending on the conditions within the early

nebula, determined the chemical makeup of the planets that we see today (e.g., Mizuno 1980). Subsequent migration of the planets, especially Jupiter and Saturn, redistributed the giant planets and leftover debris (comets, asteroids, KBOs, TNOs and the Oort cloud) to their present locations, and created the stable, habitable conditions we observe in our inner solar system today. Confirmation of these theories requires observational evidence, both from our own solar system and in planetary nurseries around other stars. The proposed observatory would **employ remote sensing of atomic, molecular, isotopic and solid state signatures to understand the evolution of our solar system** (e.g., Figure 1).

Giants: The giant planets were the final repositories for nebula gases hydrogen and helium, in addition to icy planetesimals trapping volatile species. Their bulk composition bears witness to the ratios of elements and isotopes in the source reservoirs, and detailed comparisons between the four giants would reveal shared reservoirs and common formational processes (e.g., Atreya et al., 1999). Although some species are locked away in deep condensation clouds (e.g., O/H cannot be measured due to the condensation of deep H₂O clouds), and others have no spectral signatures (e.g., noble gases), high spectral-resolution remote sensing will provide comparisons of elemental enrichments of cosmologically abundant species (C, N, P, S) and isotopic ratios (D/H, ¹³C/¹²C, ¹⁵N/¹⁴N, ¹⁸O/¹⁶O). C/H appears to increase with radial distance from the Sun, but the deuterium

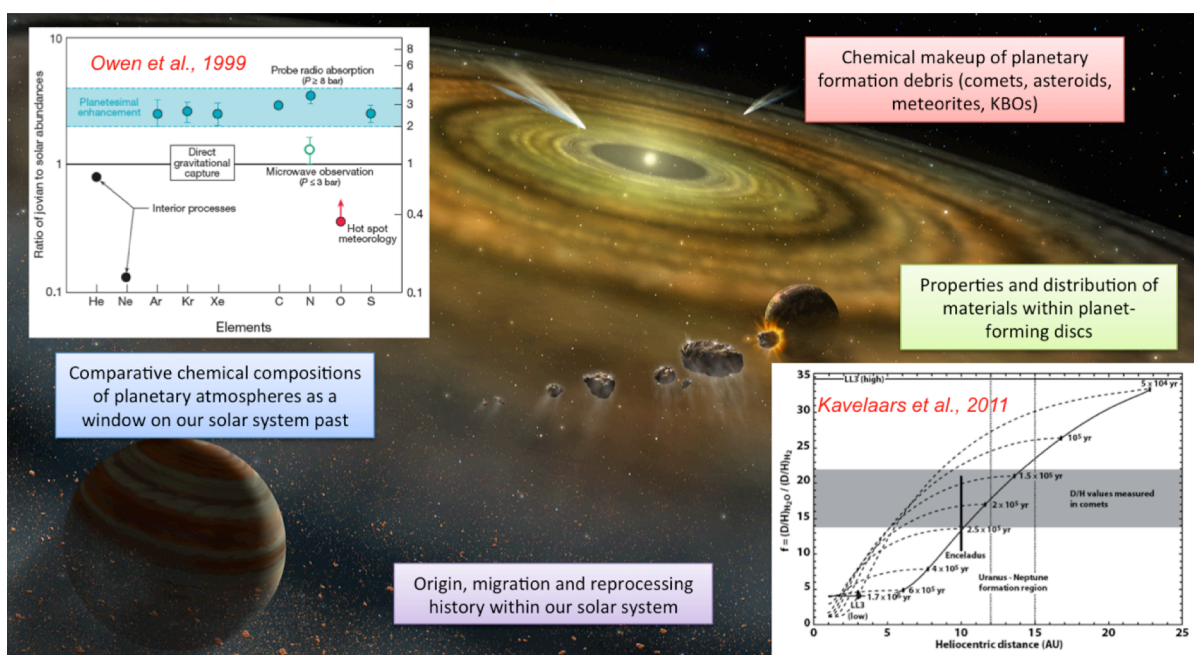


Figure 2 Thermal infrared spectroscopy from PSIO will be used to assess the elemental and isotopic inventories of a variety of solar system targets, from planetary atmospheres to icy and rocky material left over from the birth of our solar system.

enrichment groups the giants into two categories, the ice giants being enhanced in deuterium compared with the gas giants (e.g., Feuchtgruber et al., 2013 from Herschel and ISO studies). Far-IR sounding of the hydrogen-helium continuum will provide an accurate measure of the helium inventory on each planet (e.g., Conrath and Gautier, 2000), crucial to understanding the cooling history of the giants. Furthermore, far-IR sounding is sensitive to the peak of the Planck emission from the cool atmospheres of Uranus and Neptune, allowing a re-determination of their intrinsic heat flux for the first time since Voyager (e.g., Pearl et al., 1991). Furthermore, these same elemental enrichments, if detected on extrasolar giant planets, will allow us to identify common formation scenarios in different planetary systems (for example, the C/O ratio determined from transit spectroscopy could reveal their carbon-rich or oxygen-rich origins, (Madhusudhan 2012). By **comparing the elemental and isotopic enrichments of multiple giant planet atmospheres**, this observatory will place constraints on the chemical inventories of different planetary systems.

Debris: The composition of the giants will be **compared with the chemical make-up of rock-ice remnants of planetary formation**, from comets and near-Earth objects, to asteroids, TNOs and KBOs. The volatile inventories of such objects, especially the fraction of water and deuterated species, will be used to understand the distribution of icy material in the early solar system. The D/H ratio in cometary H₂, H₂O and CH₄, from the Jupiter family, to long period and the newly-recognised families of main belt comets, will shed light on the delivery mechanism for volatiles to the early Earth, especially the origins of our world's oceans. For example, higher D/H ratios with radial distance from the Sun (e.g., Kavelaars et al., 2011, Figure 1) will support a cold, distant origin for Earth's water content. The properties of the surface ices, dust and minerals of planetary satellites, KBOs and TNOs will be revealed by a sensitive search for broadband spectral features, providing clues to **their origins, subsequent migration and the nature of reprocessing over their history**. Finally, the sensitive constraints on the D/H ratio observed in terrestrial planet atmospheres (HDO on Venus and Mars) reveals insights into the loss processes for water from these worlds, and the potential limits of the habitable zone.

Discs: These compositional signatures provide a window onto the chemistry, radial mixing, disc-clearing and migrational processes at work in the early solar system, but to place these results in a broader context the infrared results must be compared with the spectral signatures of ices and dust in planet-forming nebulae, debris discs and protostellar discs. This would establish the validity of well-defined 'snow lines' and reveal the radial distributions of different source materials (carbon, deuterium, oxygen, nitrogen), while sampling the range of possible system architectures resulting from changes in stellar type and metallicity. The **distribution of materials within our own solar system, revealed by infrared remote sensing, will be compared with planetary systems of various ages**, from the youngest planet-forming discs to the continuum of planetary types in established systems. Taken together, these observations will reveal the cycling of planetary building blocks within these systems.

Theme II: Atmospheres

Key Question: What powers the circulation, chemistry and dynamics of planetary atmospheres, from the deep troposphere to the thermosphere?

The combination of unprecedented infrared spatial resolution, tunable spectral resolutions and long-baseline observations will permit **pioneering new studies of the atmospheres of Mars, Jupiter and Saturn, along with disc-averaged studies of rotational variability of the ice giants, Titan and planets around other stars** (e.g., Figure 2). Solar system atmospheres are natural laboratories for testing our understanding of fluid processes and chemistry under extreme conditions, and provide a template for our understanding of planets around other stars. The infrared has long been the primary tool for studying the circulation, meteorology, chemistry and cloud formation in planetary atmospheres, using well-mixed species (H₂, CO₂, CH₄) as thermometers to map thermal structures in three dimensions and the plethora of emission and absorption features of minor species to trace the motion to understand compositional variability (e.g., Hanel et al., 2003). For those planets that can be spatially resolved, this observatory will address a long-standing problem by allowing us to match changes in albedo, winds and colouration observed in reflected sunlight (to be provided by

simultaneous ground-based observations) to environmental parameters such as temperature, wind shear, gaseous composition (i.e., humidity) and cloud properties (e.g., Fletcher et al., 2010). By providing broadband spectroscopy with spatial resolutions approaching that of reflected-sunlight images, we will measure for the first time how these meteorological variations relate to the atmospheric plumes, waves, vortices, instabilities and the general banded structures that are commonplace in planetary atmospheres. This comprehensive meteorological study will not be achieved by future outer planet missions (NASA/Juno and ESA/JUICE for Jupiter) because of an absence of infrared spectral coverage. The observatory will **reveal how energy, momentum and material are transported both horizontally and vertically from the troposphere to the thermosphere, allowing us to place the Earth's habitable atmosphere into its broader astrophysical context.**

Composition: High spectral resolutions in the mid and far-IR are a unique element of this proposal, allowing us to resolve narrow Doppler-broadened emission features in **planetary stratospheres and mesospheres to probe atmospheric regimes typically ignored** by studies sensitive only to tropospheres. UV photolysis, a close connection with planetary aurora, and an external influx of particles from comets, asteroids and dust produces a 'zoo' of chemical species in planetary upper

atmospheres, which are then redistributed by the general circulation (for example, HCN in the atmospheres of Titan and Jupiter trace the atmospheric flow). High spectral resolutions will allow us to trace those species, determine their vertical profiles and to discover new ones to **plug the gaps in our knowledge of atmospheric chemistry.** Venus' stratospheric composition may be influenced by injections of sulphur-bearing materials due to geologic processes on the surface (e.g., volcanism), and these species could be mapped (by an IR observatory capable of viewing the inner solar system) to relate them to the UV-absorbent dark patches observed above the Venusian H_2SO_4 clouds, building on the near-IR discoveries of Venus Express. Mars' atmospheric water cycle and spatial distribution (e.g., the relation to topography and dynamics, Fouchet et al., 2007), the distribution of CO, temperature and winds will be monitored over seasonal timescales. Furthermore, Mars' atmospheric composition may be perturbed by episodic injections of methane (Formisano et al 2004; Mumma et al 2009), either from geologic or possibly astrobiological sources, and mapping the sources and sinks of Martian methane in the mid-IR is a crucial goal for understanding the atmospheric cycles on our closest neighbour. Trace species such as phosphine, arsine and germane, along with the far-IR signatures of the spin state of H_2 (ortho/para ratio, Conrath et al., 1998), will be used to understand disequilibrium processes and vertical

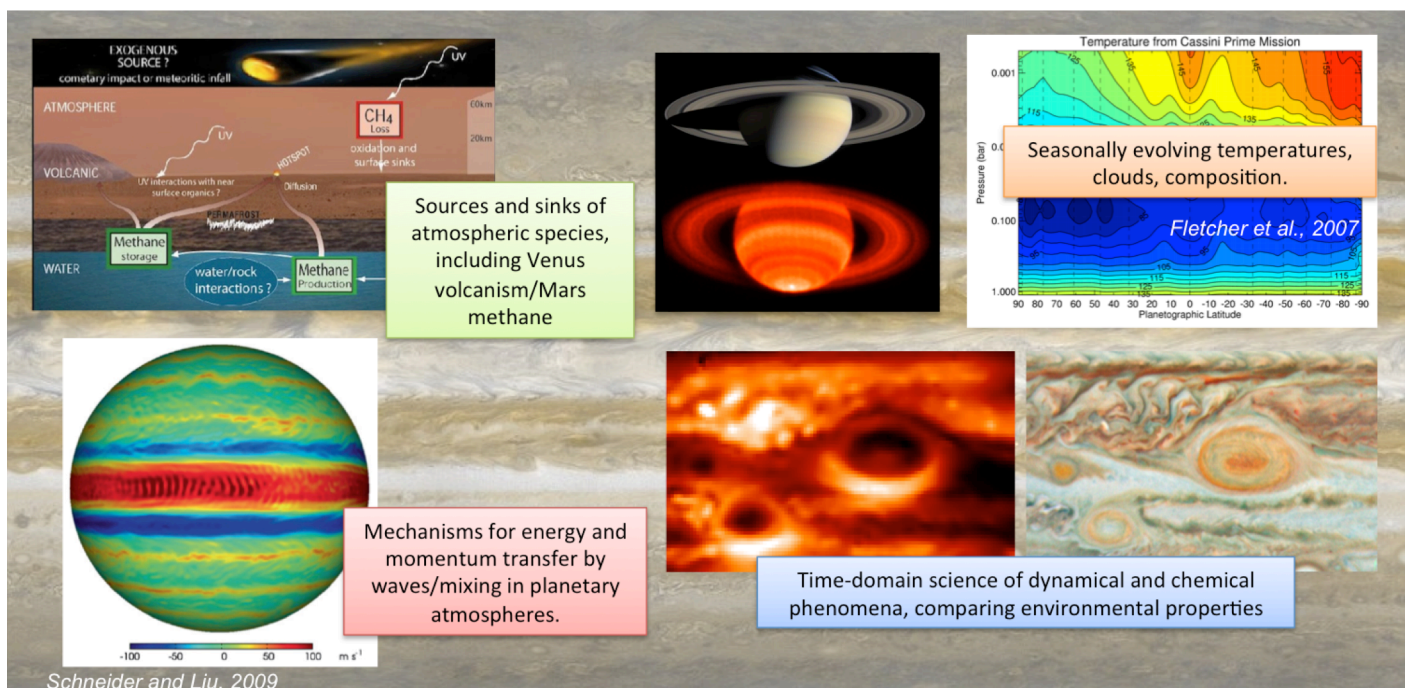


Figure 3 Imaging spectroscopy across the thermal infrared reveals the environmental variability associated with visible changes (atmospheric plumes, planetary impacts, volcanism, winds and waves, slow seasonal evolution), and near-continuous monitoring of rapidly-evolving targets will open up the field of time-domain planetary astronomy.

mixing within giant planet tropospheres. Finally, the chemistry of the thick N₂ atmosphere of Titan could well be representative of the type of environments observed on the pre-biotic Earth, and the cycle of methane between geological sources, seasonal lakes, seas, atmosphere and clouds will provide insights into this unique atmosphere long after the end of the Cassini mission. Regular disc-averaged observations of Titan at heterodyne resolutions in the mid-IR will be used to search for new species, determine vertical distributions and track compositional variability associated with methane-cloud activity to better understand the meteorological variability of this unique world.

Middle Atmosphere: Where heterodyne-resolution spectroscopy with $R \sim 10^7$ really provides advances is in measuring the Doppler shift and profiles of stratospheric lines, allowing us *to map the three-dimensional wind structure above the cloud tops of Venus, Mars, Jupiter and Saturn for the first time*. Recent research has begun to reveal the dynamic nature of these stratospheres, with a multitude of wave phenomena (e.g., tropical oscillations similar to Earth's quasi-biennial oscillation, and horizontal thermal waves forced by tropospheric storms) and polar vortex phenomena (e.g., seasonally-variable polar collars on Saturn, Titan, Uranus and Neptune) with direct relevance to our own planet. Well-separated lines of CH₄, CO, and H₂O permit wind measurements in regions where there are no visible cloud tracers (e.g., Lellouch et al., 1991; Lellouch et al., 2008). By providing heterodyne resolutions in the mid-IR, we are able to provide wind speed measurements at spatial resolutions far exceeding the capabilities of Herschel. For example, Kostiuik et al. (2005) utilized ethane emission at 12 μm to measure Titan's middle atmospheric winds from the Subaru observatory with a 0.4" field of view. The vertical variability of zonal and meridional winds, particularly in association with vertically propagating waves, has far reaching implications for *how planetary atmospheres move energy, momentum and material from place to place*. Furthermore, by opening up this 'middle atmosphere' on all of these planets, we will reveal the connectivity between the convective weather layers and the upper atmosphere, ionosphere and thermosphere. Indeed, energy transport by waves from below *could* be important in heating of the high atmospheres of the giant planets to temperatures far exceeding that expected from solar heating alone (e.g., Yelle et al., 2004), a

situation termed the 'energy crisis.' However, the heating and cooling effects of gravity waves in particular have been demonstrated to be insufficient, (Hickey et al., 2000), so that ionospheric drag processes and Joule heating are sometimes invoked to explain the discrepancy (Muller-Wodarg et al., 2006). By mapping atmospheric winds and waves and their variability over time, this observatory will directly address this longstanding mystery.

Broad Coverage: Simultaneous spectral coverage from the mid- to the far-IR is another requirement for this observatory, providing access to temperature, aerosol and compositional measurements at the same moment in time. Separation of these measurements causes immense difficulties in the interpretation of these rapidly evolving atmospheres. For example, observations have never been able to simultaneously trace giant planet clouds in the 5- μm window (probing 1-4 bar depths) and the 10- μm and 100- μm regions (0.1-1 bar), whilst also relating them to the visible patterns observed at the cloud tops. Furthermore, dust and ice spectral features are broad and difficult to identify when narrow spectral ranges are used. Modelling a wide spectrum allows us to separate the signal of these ices from thermal and compositional signatures, breaking the degeneracy and yielding significant advances in our understanding of the global cloud morphology and composition. In particular, the observatory will study the cycling of dust in Mars' atmosphere, the triggers of global dust storms and the radiative influence of the particulates. Provided the platform has the necessary stability for *observations of exoplanetary transits*, the broad spectral coverage beyond 5- μm offers: (i) significant potential to detect NH₃, CH₄, H₂O, CO, CO₂, associated hydrocarbons, nitriles and other exotic species in their atmospheres; (ii) the capability to separate thermal and compositional signatures that would be impossible with narrow spectral ranges (e.g., using the CO₂ band at 15 μm for super-Earths); (iii) access to cooler transiting planets around M stars, whose black body emission peaks at mid-IR wavelengths; and (iv) access to bands of astrobiological significance on super-Earths (e.g., O₃ at 9.6 μm). The spectral coverage also offers the possibility to *observe brown dwarf atmospheres in the mid-IR*, something that was only previously possible with AKARI (e.g. Sorohana et al., 2012). These results highlighted the lack of knowledge of molecules in the at wavelengths exceeding 3 μm ,

with NH_3 being weaker than expected, and CO being present in all low-temperature, methane dominated brown dwarfs. This ***broad spectral coverage and tunable spectral resolution offers significant potential for new discoveries for both the planetary and exoplanetary communities.***

Time Domain Science: Planetary atmospheres evolve over a range of timescales, from short-term meteorological variations that can be discerned by cloud tracking, to year-on-year variations (e.g., the life cycles of Jupiter's belts and zones, enormous storms on Saturn) and seasonal variability (Martian hydrological cycle, thermal and compositional asymmetries on the giant planets). The underlying sources of these variations (e.g., latent heat release, injection of material from rising volcanic or atmospheric plumes) remain a mystery due to inadequate time sampling and uncertain causal connections. The extreme conditions found on the ice giants (the large axial tilt of Uranus, and the unusual balance of solar forcing and internal heat emission on Neptune) are of particular importance to our understanding of seasonal processes. Atmospheric features have never been tracked at these wavelengths in continuous movies, and a year-by-year comparison of seasonal changes has never been performed systematically. Furthermore, tracking thermal emission variations as an object rotates (e.g., Titan, Uranus and Neptune) can reveal longitudinal contrasts and their causes (unusual cloud activity, waves, storms,

vortices, etc.). Irradiated exoplanets and brown dwarfs both have atmospheres that vary with time, either due to irradiation from their host star, or shifting molecular bands, particularly for objects on the L-T transition (Buenzli et al., 2012). All of these ***time-domain studies require a dedicated platform*** able to revisit targets on multiple occasions over a long mission lifetime, to provide the next stage in our understanding of atmospheric processes across a broad range of astrophysical objects.

Theme III: Surfaces

Key Question: What are the thermophysical properties of primitive bodies (comets, asteroids, TNOs, KBOs), planetary surfaces, rings and satellites in our solar system?

The thermal infrared provides important diagnostics for the conditions to be found on rock-ice objects with tenuous 'atmospheres', such as the icy satellites of the giant planets, or the comae of cometary objects (e.g., Figure 3). The peak of their spectral energy distributions occurs at the longest wavelengths in the far-IR, and the characterization of spectral intensity and slope reveals the temperature of these objects, and their physical sizes from energy balance if the albedo is known. ***Infrared imaging also reveals the physical state of the surface ices and dust*** (e.g., amorphous and crystalline forms provide different spectral

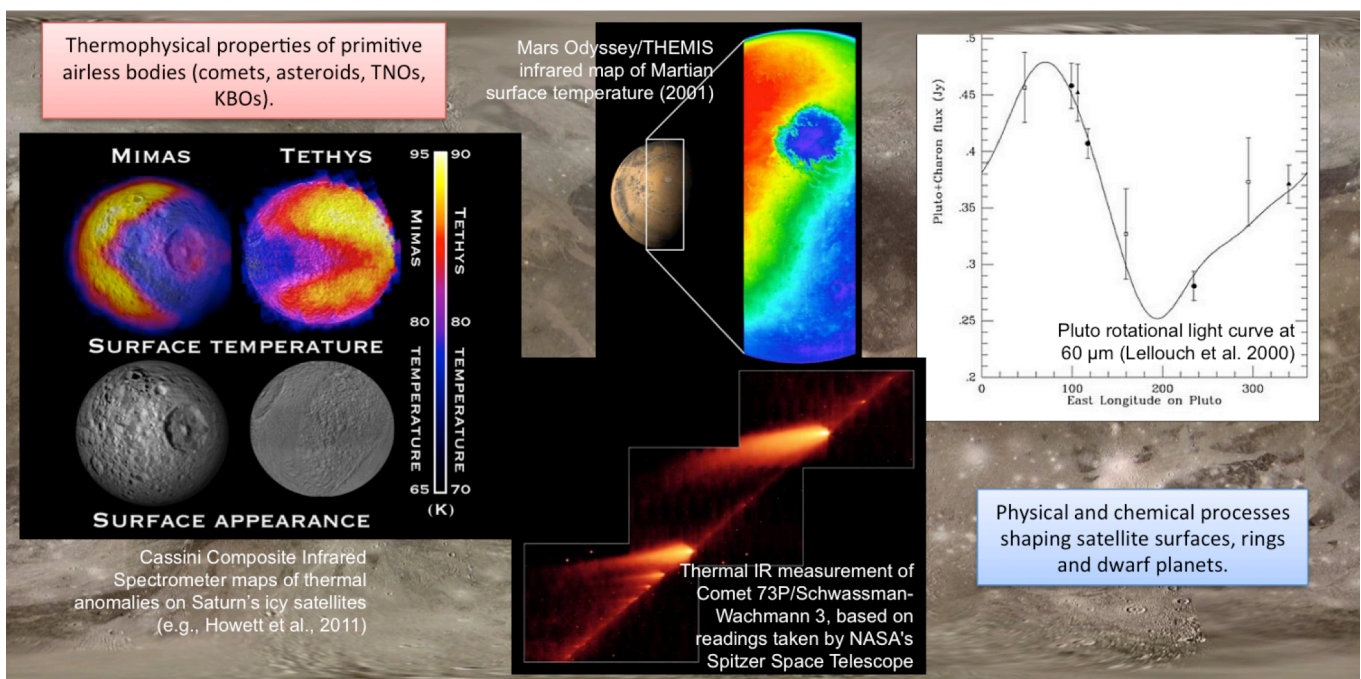


Figure 4 Thermal infrared mapping of comets, asteroids, planetary satellites and dwarf worlds will reveal the thermophysical properties and composition of their surfaces, as well as constraining their energy budgets and identifying the presence of emission variations related to endogenic activity.

signatures in the infrared), and broad signatures can reveal the presence of exotic materials such as organics. Unlike the larger planets that can be spatially resolved, the majority of planetary satellites, asteroids, KBOs and TNOs will be measured as disc-averages. However, measurement of rotational variability can be used to determine spatial contrasts in thermal inertia properties as heterogeneous regions rotate into sunlight (e.g., Lellouch et al., 2000), and place constraints on the physical size of small, unresolved objects. **Comparing compositional differences between these objects provides insights into their origins** and the processes responsible for shaping our planetary system.

Comets and Asteroids: Emission from cometary nuclei and their expelled gas and dust tails will be monitored as these objects approach and recede from perihelion, revealing the properties and origins of this primitive solar nebula material and identifying the ices, volatiles (H₂O, CO, CO₂, CH₃OH, etc.), organics and dust production rates and properties. For example, olivine and pyroxene signatures in the dust between 9-35 μ m would reveal the fraction of crystalline silicates from the hotter regions of the solar nebula, versus amorphous silicates of a cooler, more distant origin. The importance of the D/H ratio was described in Theme 1, but the spin-state of hydrogen in H₂O also reveals the processes at work during their formation. Furthermore, distinguishing contrasts in the crystallinity of the different dust jets could reveal the inhomogeneity of the nucleus. The cometary inventory of organics - spectrally detected as polycyclic aromatic hydrocarbons (PAHs) between 6-12 μ m - will allow us to investigate the **delivery of organics into the forming terrestrial planets**. All of these parameters will be compared with similar measurements of asteroids (particularly Vesta), main belt comets, and recently identified objects in the asteroid belt that exhibit volatile degassing upon their perihelion passages. This discovery has turned our naïve impression of volatile-rich comets and desiccated asteroids on its head, and could have implications for volatile delivery mechanisms in the early solar system.

Dwarf Planets: There are more than 1500 objects beyond the orbit of Neptune, and although Herschel focused on around 130 targets, a long-lived observatory such as PSIO has the **potential to discover and determine thermal properties of**

new objects whilst refining our understanding of those previously observed (e.g., the Herschel program for cool TNOs and Plutinos, Müller et al., 2009; Mommert et al., 2012). Thermophysical conditions, shapes, sizes and rotational properties of dwarf planets and KBOs, such as Pluto, Ceres and Eris, could be determined and refined by monitoring their rotational light curves at a variety of infrared wavelengths. Pluto, for example, is believed to have a tenuous CO₂ atmosphere that collapses as surface frosts at aphelion, with a rudimentary atmospheric circulation between summer and winter poles. The atmospheric pressures and densities could be determined from occultation studies, while the heterogeneity of surface thermal inertia and emissivity might be deduced from light curve measurements. The same could be provided for other KBOs and TNOs in the distant solar system to **assemble reliable statistics on the population, size distributions and environmental characteristics of these worlds**. Their sizes and albedos will be determined from thermal measurements (assuming radiative balance), and potentially their densities and porosities if their gravitational fields can be determined via mutual interactions. The energy balance and thermal inertia will allow us to investigate heat transport mechanisms within asteroidal regoliths of varying densities. Finally, the surface mineralogy (balance of crystalline and amorphous silicates) could be determined via spectral features in the infrared, particularly the ice water hydration band near 3.1 μ m.

Satellites: The satellites of the giant planets will be studied in a disc-averaged sense, monitoring their infrared lightcurves as they rotate around their gas giants. With a spectral range covering the peak and tail of the Planck emission, the observatory will use spectral emissivity variations with wavelength to constrain surface composition and water ice grain sizes. Recent thermal-IR studies of the Saturnian satellites have revealed unique **near-surface heat sources** (e.g., the tiger stripes of Enceladus) and **thermal contrasts associated with surface-magnetosphere coupling** (e.g., the 'Pacman' of Mimas and Tethys, Howett et al. 2011) and **leading-trailing hemisphere asymmetries** as these tidally-locked satellites move along their orbits (e.g., Europa, Spencer et al., 1999). Of particular interest is Io, the most volcanically active object in our solar system, and the infrared tracking of gaseous SO₂ lines emitted either from sublimating surface frosts or active volcanism to study the surface

distribution. Voyager observed SO₂ bands on Io at 7.4 μm to be enhanced over one of the volcanoes (Pearl et al. 1979), and additional subtle features between 10-20 μm were observed and studied from the ground (e.g., Spencer et al., 2005). Repeated long baseline observations of Io and the other satellites can be used to monitor ***variations in thermal emission caused by endogenic activity and discrete hotspots*** on these satellite surfaces, and to ***constrain the energy budget and extent of tidal heating***, providing the crucial energy source necessary for the habitability of these satellites. Even closer to home, such an observatory could provide global coverage of our Moon, investigating the trapping of volatiles within cold, shaded regions of the lunar surface (e.g., Paige et al., 2010) and the thermal inertia as solar illumination varies with time.

Dust: Finally, thermal-IR imaging could be used to study the variable temperatures with Saturn's broad, dense rings as a function of radial distance and season, and to identify mineralogical signatures of silicate dust, water ice and the various contaminants responsible for giving the rings their colours. Broadband thermal emission from dusty environments could be studied both in our solar system (e.g., zodiacal dust) and in planetary debris discs around other stars, searching for common compositional signatures of dust material throughout our solar system.

Theme IV: Interaction

Key Question: How do solar system objects interact with one another, with the parent star and how do they vary seasonally?

The final theme to be addressed by the observatory is that of time-variable interactions between all of the various bodies within our solar system (Figure 4), allowing us to investigate planetary processes in using the expected long operational baseline (5+ years) of the observatory.

Planet-Star Interactions: Planets interact with their host stars in two ways – via stellar irradiation, causing atmosphere/surface heating and sparking photochemistry; and via solar wind plasma interacting with their extensive magnetospheres. The latter is most evident in the infrared at the poles, where charged particles accelerated into the upper atmosphere generate dramatic auroral light shows and can lead to unusual chemistry at high latitudes. Planetary aurorae interact with many factors, including the IMF direction (interstellar magnetic field), the presence of internal plasma sources and the solar wind pressure, so auroral monitoring can be used as a proxy for studying the balance between solar wind and internal plasma source activity at work within planetary magnetospheres. Planetary aurorae and ionospheric winds will be mapped in H₃⁺ emission near 3.4-3.6 μm (and possibly fluorescence of other molecular species), showing how the morphologies of auroral ovals shift in shape and intensity in

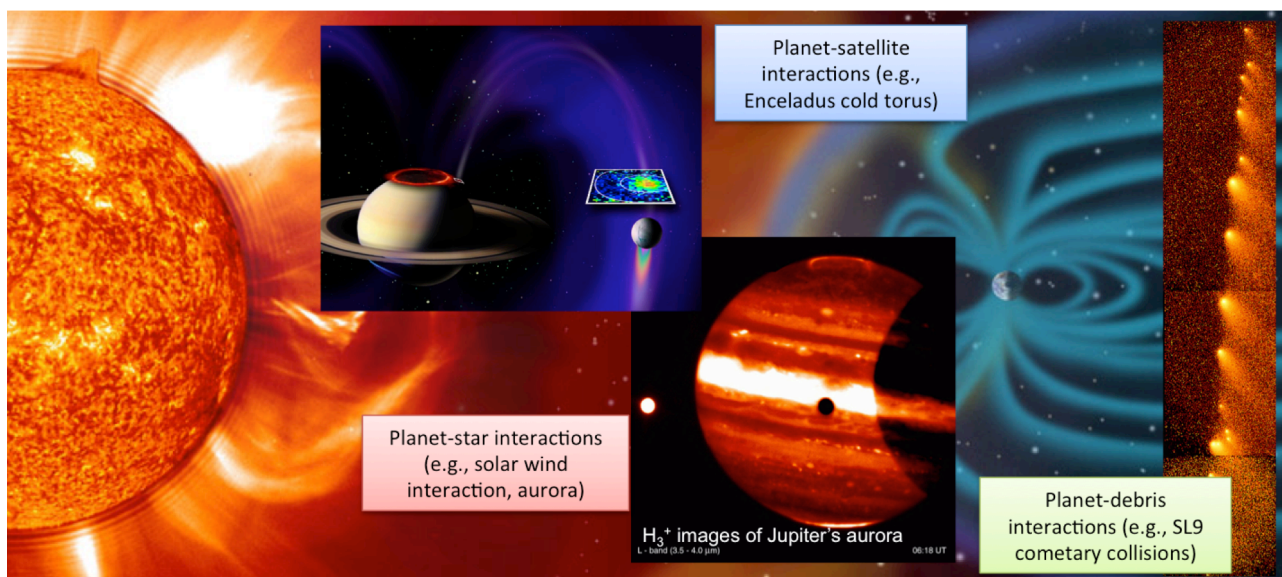


Figure 5 Long-baseline monitoring of solar system targets can be used to understand their complex interactions, both with the Sun and within each planet-satellite-ring system. Cometary impacts, planetary upper atmospheres and aurora will be of particular interest to PSIO.

response to the varying solar wind pressure and other factors. Such H_3^+ mapping of the giant planets allows studies of thermospheric temperatures, gaining insights into the high-temperature 'energy crisis' described in Theme 2 as well as probing seasonal variability, upper atmospheric winds/circulation and the strength of the magnetic field. If coupled with near-simultaneous observations of H_2 emission near $2\ \mu m$ (extending below the present arbitrary $3\text{-}\mu m$ cut off for PSIO), this would also allow us to estimate the auroral electron energy in addition to the temperature to enable us to study the coupling between the solar wind, magnetosphere and ionosphere (e.g., Tao et al., 2012). Furthermore, the high thermospheric temperatures determined at wavelengths near $4\ \mu m$ will be directly correlated with stratospheric 'hot spots' observed in the $7\text{-}14\ \mu m$ range in hydrocarbon emission. The high sensitivity of an infrared observatory might also allow the first detection of H_3^+ emission from Neptune (e.g., Melin et al., 2011), and will certainly enhance studies of all the giant planets by **performing invaluable remote magnetospheric and ionospheric science via continuous 'real-time' auroral monitoring over several planetary rotations.**

The effects of the seasonal variability of sunlight will be investigated on all planetary surfaces and atmospheres over a long mission baseline. For example, the near-IR reflectance of Neptune shows unexplained correlations with the degree of solar activity (e.g., Lockwood and Jerzykiewicz, 2006), possibly due to albedo effects in the clouds, so monitoring how planetary atmospheres respond to solar cycle variations could reveal new unforeseen interaction mechanisms. Thermal and compositional mapping of Saturn and Titan will **track the evolution of hemispheric asymmetries driven by seasonal heating and photochemical patterns**, extending our understanding of seasonal processes in the Saturn system long after the conclusion of the Cassini mission.

Planet-Satellite Interactions: Visiting spacecraft have revealed that giant planet satellites are depositing significant quantities of material into the

magnetospheric plasma environment. Torii of plasma are deposited in the wake of these satellites, notably Io and Enceladus, and these torii can interact with the planetary atmosphere, producing 'footprints' in the auroral ovals and transferring dust and ice to 'rain down' on the atmosphere at locations governed by the magnetic fields. The water ice materials being actively vented by fissures on Enceladus create a cold torus orbiting Saturn that absorbs the infrared emission from Saturn's stratosphere (discovered in the far-IR by Herschel, Hartogh et al., 2011). Furthermore, the tidal forces keeping the satellite interiors in a fluid state vary as a function of radial distance from the gas giant, and the thermal balance on these satellites will provide estimates of the efficiency of tidal energy deposition. These **interactions between planets and their satellites serve as laboratories for planetary system processes on a variety of scales.**

Planet-Debris Interactions: Finally, planets continue to interact with the debris (comets, asteroids and interplanetary dust) left over from the epoch of planetary formation. Planetary impacts, such as Shoemaker Levy-9 in 1994 on Jupiter (Harrington et al., 2004), create debris fields that persist for many weeks post impact, and chemical perturbations that can last for years (e.g., HCN and CO on Jupiter). Fast et al., (2002) studied the temporal behaviour of ammonia lofted into Jupiter's stratosphere by the SL9 collision using high-resolution mid-IR spectroscopy. The thermal properties, chemistry and mineralogy of those debris fields have direct relevance to airburst explosions in Earth's atmosphere, and can only be investigated in the thermal infrared with an agile and flexible platform. Interplanetary dust also rains down on these planets, contributing oxygenated species (CO, CO₂, H₂O) into the upper atmospheres of planets that can be detected via stratospheric emission lines (e.g., Moses et al., 2000). Thus **the planets are fundamentally linked to the host star, the satellite system and the wider interplanetary space** via a variety of processes that could be investigated by this observatory.

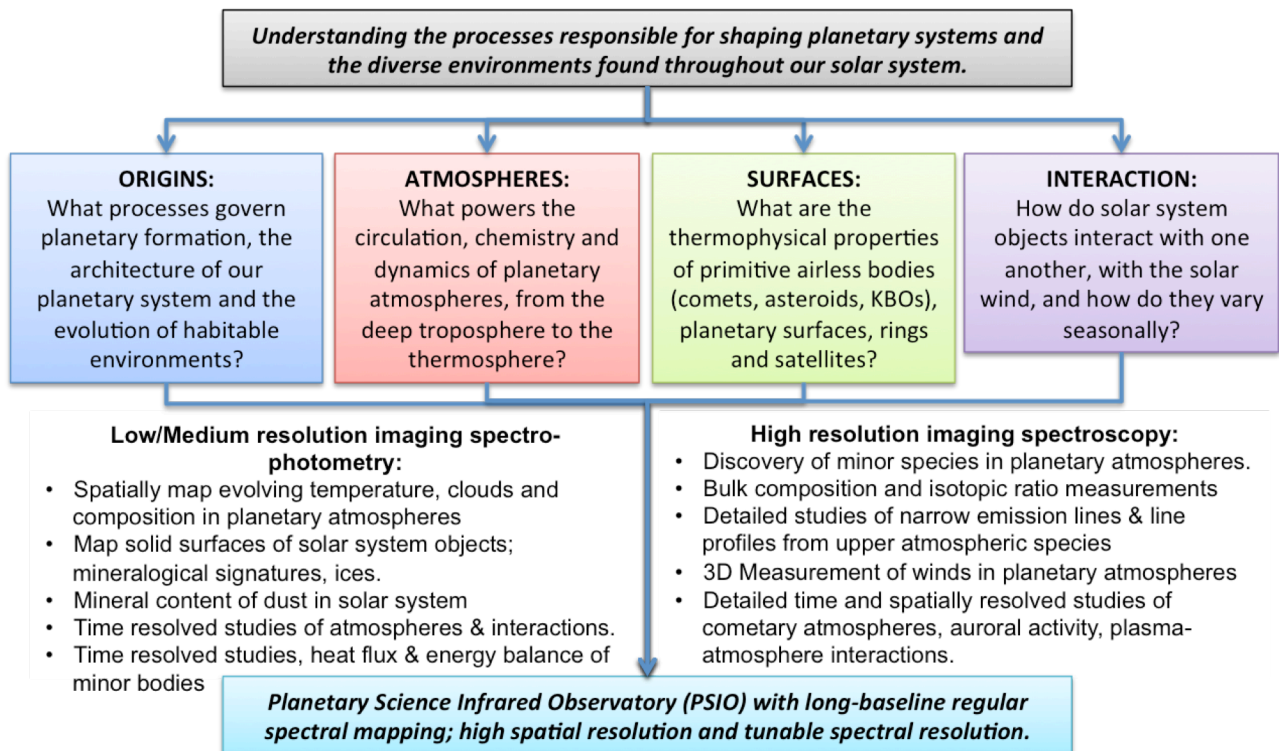


Figure 6 Outline of the four themes to be addressed by a dedicated planetary science observatory, showing the flow from our top-level science question, to the four themes and a subset of specific objectives for PSIO.

4. Observatory Architectures

We summarise the four science themes for understanding the processes shaping our planetary system in Figure 5. To address these objectives, we advocate an observatory class mission, providing imaging spectroscopy at moderate ($R \sim 10^3$) resolution over an arcmin FOV; and heterodyne-level resolution ($R > 10^6$) in selected mid-IR and far-IR bands to probe planetary regimes not previously explored. Spatial resolution must be sufficient to resolve contrasts across the discs of Venus, Mars, Jupiter and Saturn, suggesting a 3.5-5 m monolithic mirror or a distributed array for longer wavelengths. The observatory must be flexible and agile to respond quickly to new events, and must be capable of devoting significant time to single targets to advance time-domain science within our solar system. If possible, the observatory should be stable enough to address ancillary science of exoplanets, brown dwarfs and planetary discs, although we recognise that no single mission architecture will be capable of addressing the four science themes equally. The science requirements for an infrared planetary observatory break into two distinct categories as shown in Figure 5: low to medium resolution spectrophotometry and very high-resolution spectroscopy.

Possible mission architectures

Two basic mission architectures present themselves: a single monolithic mirror passively cooled to space and a distributed array of smaller dishes structurally connected to synthesise a large aperture through interferometry. The **monolithic mirror concept** could either be a Herschel or a non-actively cooled SPICA type on-axis Cassegrain telescope; or a Planck style off axis telescope. While there is much design heritage with on-axis telescopes, their size is limited by the need to fit them into a launch vehicle fairing – 3.5 m is currently the largest that can be accommodated. With an off-axis design the physical area can be made larger using an elliptical primary as illustrated by the JPL SAFIR concept shown in figure 6. A 10-m design requiring cooling to <10 K is shown, but a planetary observatory would not require this degree of cooling and both the dish and the Sun shields would be considerably smaller. A 5-m off-axis design would provide spatial resolutions such as those listed in Table 1.

At longer wavelengths where the resolution offered by a single dish is insufficient, the idea for a **structurally connected small dish array** stems from heterodyne spatial interferometers successfully operating on Earth. As well as the very large sub-mm ALMA instrument, there are two other instruments of interest: the Sub-Millimeter Array (SMA) on Hawaii operating at ~ 100 s GHz and

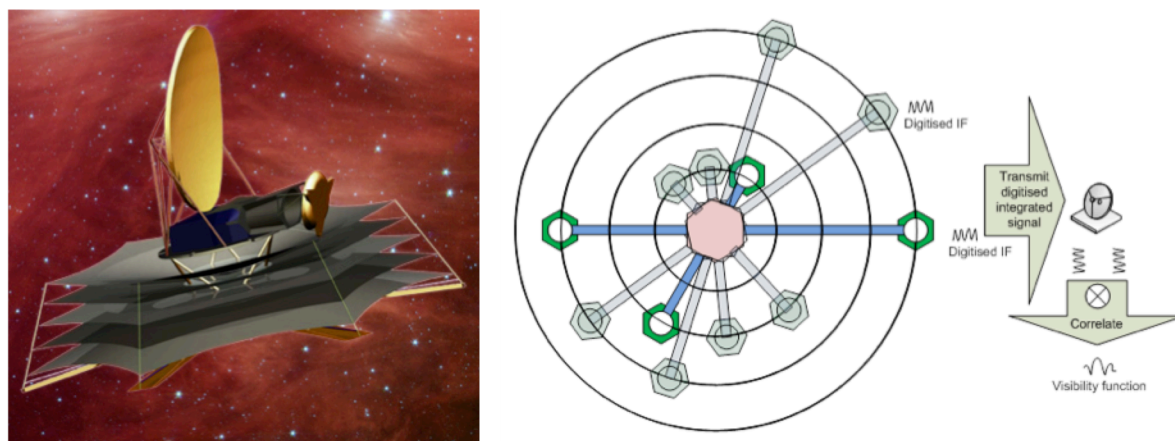


Figure 6: Left: The JPL SAFIR concept for a cooled off axis 10 m dish, taken for the SAFIR study website. Right: A possible concept for a distributed spatial interferometer array. This is shown in plan view with four small receivers deployed on booms from a central hub. The deployed elements have a ~ 1 m antenna with heterodyne mixer technology receivers centrally phase locked. The whole satellite is rotated sweeping out a partially populated uv plane as indicated by the light grey and the solid circles. The spatial correlation could be achieved either through a correlator placed on the central spacecraft or by digitising the spectrum from each receiver, time stamping it to high precision and telemetering the result to the ground – similar to a VLBI radio system.

the Infrared Spatial Interferometer (ISI) on Mount Wilson operating in the $10\ \mu\text{m}$ atmospheric window (Hale et al., 2000). Both instruments use heterodyne receivers with distributed phase locking systems for the local oscillators. A similar arrangement could be achieved in a structurally connected array of small dishes operating in space. An outline concept is illustrated in figure 6 where ~ 1 m dishes are deployed on extendable booms around a central hub spacecraft. The whole satellite is then rotated and a large aperture is synthesised. Deployable booms up to 5 m have been demonstrated on various missions and the small nature of the dishes, combined with the use of heterodyne receivers, reduces the necessity for low temperatures at the deployed receivers. The advantage of heterodyne spatial interferometers is that phase knowledge is a natural consequence of the detection technique provided the local oscillators are phase locked. Various methods of phase locking can be contemplated, the easiest being the distributed phase lock signal already used in ISI, SMA and ALMA. In the MIR, this concept also obviates the difficult task of optical correlation and use well known RF techniques instead. For the local oscillators themselves, new technology is available or being actively developed for low power low mass LOs working at 100s-GHz (photonic mixers) to supra-THz (Quantum Cascade Lasers, QCLs) which have demonstrated performance from ~ 2.5 THz up to 30 THz. No moving parts are required apart from the deployment booms and the technique of phase closure could be used to eliminate variations in path differences due to any motion of the booms during observations. While a multiple dish

interferometer working at THz represents a challenging mission but would give the chance to provide far-IR spatial resolutions to match the mid-IR (Table 1) and most, if not all, of the technology required is under active development.

Instrument options

For the imaging spectro-photometer several instrument options are possible ranging from an Integral Field Unit (IFU) grating spectrometer such as the PACS instrument on Herschel, an imaging Fourier Transform Spectrometer (iFTS) like the SPIRE instrument on Herschel or a broad band imaging radiometer using dichroic and filter chains to define fixed bands. For the high-resolution system the only extant options in the sub-mm and FIR are heterodyne receivers. For the MIR one could envisage an immersion grating type spectrometer or a heterodyne system. In the NIR an immersion grating might also be used. For all wavebands an alternative might be to use Spatial Heterodyne Spectrometers (SHS), which are a form of high-resolution static FTS utilising gratings in a beam splitter arrangement.

Spectro-Photometer: Our preferred option for the spectro-photometer is to use an iFTS. With an FTS, the beam splitter and choice of detector limit the spectral range and the spectral resolution is programmable based on the maximum achievable path difference in the instrument. With flight development programmes for beam splitters made of extremely broadband materials such as diamond now well advanced (e.g. the OTES instrument currently in development for NASA's OSIRIS-REx

asteroid sample return mission) a single instrument can provide spectral coverage from the near-IR (e.g. 1 μm) to greater than 100 μm at moderate resolution (e.g., $>2\text{ cm}^{-1}$) using broadband bolometer-type detectors. Depending on the detector configuration these instruments can also be used to provide a basic imaging capability. Spectral range for higher resolution instruments is typically limited by the requirement for cooled detectors, usually based on Mercury Cadmium Telluride (MCT). By using cooled (e.g. $<80\text{ K}$) detectors spectral resolutions of $<0.1\text{ cm}^{-1}$ are routinely achieved by Earth Observing instruments. For example, the MIPAS instrument on ESA's highly successful ENVISAT was a 320 Kg instrument with an unapodised spectral resolution 0.035 cm^{-1} and spectral coverage from 4.15 – 14.6 μm . A notable example of an FTS used in planetary science is the Cassini Composite Infrared Spectrometer (CIRS), which has operated continuously in Saturn orbit for almost a decade covering 7-1000 μm .

High-Resolution: For the high-resolution instrument we would propose that heterodyne technology is used for the FIR and MIR wavebands. The most challenging channel for this is the MIR. In this band there is on-going development of laser heterodyne spectroscopy that has been widely used from the ground for both Earth's atmosphere (Weidmann et al., 2011) and planetary atmosphere observations (Mumma et al., 1981; Kostiuk et al., 2005). The advent of MIR QCLs has enabled a far

greater flexibility in terms of LO. QCLs are readily available to cover from 4 to 12 μm and current MIR QCL technology is applicable to devices up to 20 μm for high temperature continuous wave operation. If cooled to 80-100 K, QCLs can operate down to $\sim 2\text{ THz}$. QCLs have been space qualified by NASA JPL (e.g., MSL and CLARREO).

In the future, MIR and FIR heterodyne spectrometers would clearly benefit from **advances in local oscillator (LO) and mixer technology**. On the LO side, improving the frequency agility while keeping laser spectral purity and phase locking combined monolithic design would be desirable to enhance flexibility. Several options include Continuous Wave optical parametric oscillators, high power difference frequency generation, and broadly tuneable QCLs with coverage up to 400 cm^{-1} have been demonstrated. The latter would require monolithic tuning mechanism through microelectromechanical systems devices or MIR acousto-optical tunable filters. Monolithic frequency comb generators in the NIR are also under development, which have utility in phase locking QCLs over a wide waveband when combined with photonic mixing elements. Development of QCLs at wavelengths greater than 12 μm would also be desirable and, for THz applications the development of higher operating temperature devices is also required.

Mixer technology represents another development

Parameter	Spectro-photometer	High Resolution Spectrometer
Spectral Resolution ($\lambda/\Delta\lambda$)	Low res (LRS) $\sim 5\text{-}10$ Med res (MRS) $\sim 100\text{-}1000$	HRS $\sim 10^6$ ($\equiv 15\text{ m/s}$)
Spectral coverage	LRS at least to 200 μm MRS 3 – 30 μm	Three bands at: NIR (3-5 μm) CO, H ₃ ⁺ , H ₂ O, NH ₃ , PH ₃ and other minor species. MIR ($\sim 10\text{ }\mu\text{m}$) : NH ₃ , PH ₃ , CH ₄ and higher hydrocarbons, HCN and nitriles, oxygenated species. FIR/sub-mm ($\sim\text{THz}$) – H ₂ O, CH ₄ , OH, CO + isotopes
Instantaneous Spatial Coverage	40-50" (Jupiter)	Single pixel only required – array receiver desirable
Spatial Resolution - equivalent aperture	3.5 m required for MIR 5 m desirable for MIR For $>100\text{ }\mu\text{m}$ larger baseline required (See Table 1)	At NIR and MIR same as spectrophotometer For THz longer baselines ($>10\text{ m}$) may require distributed array
Sensitivity limiting case	KBOs – require 10's μJy in MIR and 0.5-1 mJy at $>50\text{ }\mu\text{m}$ Planets do not require very high sensitivity	As long as system is stable NETDs $\sim 10\text{ s K}$
Time Resolution	Typically hours to days and weeks to track variation over all timescales. Rarely some events occur on timescales of minutes	
Mission Duration	Minimum 5 years; 10+ years desirable to track seasons in outer Solar system	

Table 3 Instrument options for PSIO.

challenge. The most efficient MIR mixers have been based on MCT photodiodes. Specific technological development towards high heterodyne efficiency, higher speed to extend the spectral multiplexed coverage, and higher saturation level to benefit from high power LO would **greatly enhance the capabilities of MIR heterodyne spectrometers**. The availability of high power LOs would also open the path to heterodyne imaging through the development of mixers' arrays. Current MCT photodiodes cut off after 11-12 μm , so specific development for longer wavelengths is required. Alternate technologies for MIR mixers are more far reaching but may provide solutions. MIR hot electron bolometer (HEB, graphene-based), transition edge bolometer, Metal Insulator Metal or Metal oxy Metal diodes using plasmonics have started to be developed. HEBs and SIS mixers are commonplace in the sub-mm band (the Herschel HIFI instrument for instance) but require liquid helium type cooling for their operation. These devices also do not work at supra-THz frequencies where Schottky diodes represent the only viable option to date. Whilst a mixture of these technologies may be sufficient for the needs of planetary observation, it is clear that such an observatory would benefit greatly from enhanced performance and/or higher operating temperatures for mixers across the THz band.

Orbital mission configurations

For both the monolithic and array architectures, the most obvious choice of orbit is to place the observatory at L2 like Herschel, Planck and JWST. This will provide an extremely stable thermal environment allowing the telescope to cool passively to below 40 K. In the case of the PSIO this will be especially stable, as the observatory will only be required to point within about ± 10 degrees from the ecliptic plane. This will allow observations of all the planets except Mercury, Venus and the Earth. Whilst only providing a limited view of comets as they approach the Earth, this should be sufficient for most purposes. An alternative, and more radical, mission would be to place the satellite into Solar orbit towards either L4 or L5. The telescope could be pointed at an oblique angle with respect to the Sun allowing observation of Venus, the Earth, near Earth objects and comets as they approach the Sun. Whilst this would be a less favourable thermal environment, with possibly a warmer telescope, it would provide a unique mission concept allowing ground-breaking observations of the inner Solar system in the infrared band.

Required Technology Roadmap

Whether or not one of the proposed mission architectures is selected, there are a number of technology developments required that would benefit not only a remote sensing observatory but also future planetary missions in general. We list the most significant ones here:

Imaging spectrophotometers: High temperature (>30 K) NIR/MIR detectors; Static iFTS designs for NIR/MIR; and broad-band immersion gratings for NIR/MIR
THz and supra-THz heterodyne receivers: Low power compact LOs such as photonic mixers and QCLs; Receiver arrays; high temperature mixers; and low power fast digital spectrometers.
Mission architecture: Lightweight large area passively cooled telescopes; spatial interferometer systems design; systems design for long duration (10+ year) missions; and backup low power highly compact mechanical coolers.

5. Conclusion

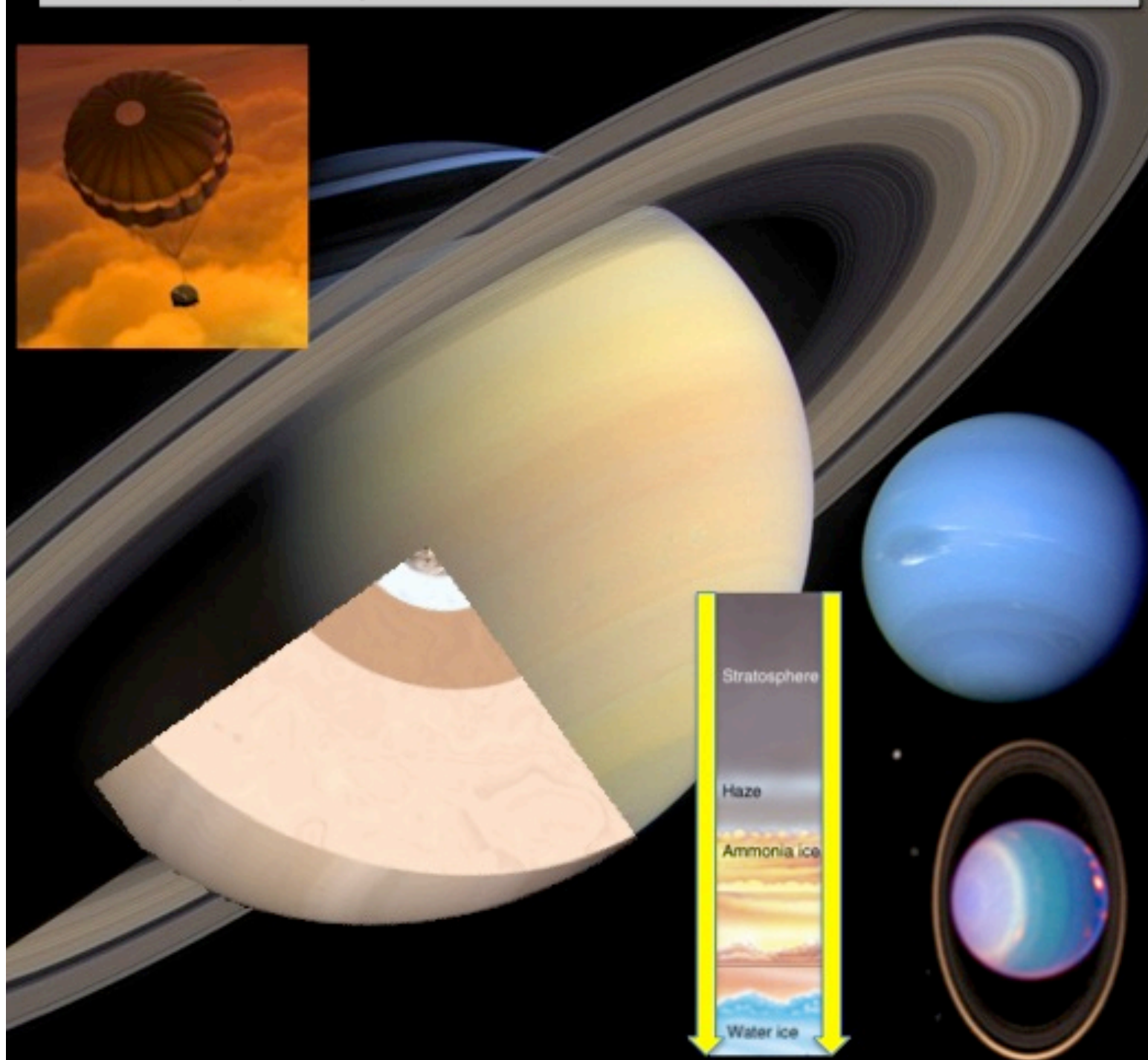
A step change in our understanding of planetary origins and the evolving environments of our solar system **requires a dedicated observatory to advance the field of time-domain planetary science** in a way that snapshot observations (e.g., JWST and SPICA) cannot. Uninterrupted infrared monitoring without the need to remove the contamination of our own atmosphere will reveal the processes at work on solar system objects in all their guises, from the smallest planetary building blocks to the largest planets of our solar system. The combination of broad spectral coverage; tunable high-resolution spectroscopy in selected channels; a spatial resolution approaching that of JWST and an **observatory optimised for longevity** would make PSIO a unique platform for solar system science. The observatory configurations and technology roadmap outlined above **represent significant advances in infrared astronomy beyond the themes covered by this paper** (e.g., exoplanet science in the habitable zone; brown dwarfs; discs and nebulae, etc.). While no single mission can achieve all of the science goals outlined here, we advocate the inclusion of infrared observations of solar system targets in any of ESA's future cornerstone missions, and development of the critical detector, heterodyne receiver and telescope technologies required to meet the aims of PSIO.

References

- Atreya et al., 1999, *Planetary and Space Science* 47, p1243-1262
- Buenzli et al., 2012, *ApJ*, 760L, 31
- Conrath et al., 1998, *Icarus* 135, p501-571
- Conrath and Gautier, 2000, *Icarus* 144, p124-134
- Fast et al., 2002, *Icarus* 156, p485-497
- Feuchtgruber et al., 2013, *Astronomy & Astrophysics* 551, A126.
- Fletcher et al., 2010, *Icarus* 208, p306-328
- Fouchet et al., 2007, *Icarus* 190, p32-49.
- Hale et al., 2000, *Astro. Journ.*, 537, p998-1012
- Hanel et al., 2003, *Exploration of the Solar System in the Thermal Infrared*, Cambridge University Press
- Harrington et al., 2004, Chapter 8 in *Jupiter – The Planet, Satellites and Magnetosphere*
- Hartogh et al., 2011, *Astronomy & Astrophysics* 532, L2
- Hickey, M. P., Walterscheid, R. L., & Schubert, G. 2000, *Icarus*, 148, 266
- Howett et al., 2011, *Icarus* 216, p221-226.
- Kavelaars et al., 2011, *Astrophysical Journal Letters* 734, L30.
- Kostiuk et al., 2005, *Geophysical Research Letters*, 32, L22205
- Lellouch et al., 1991, *Astrophysical Journal* 383, p401-406.
- Lellouch et al., 2000, *Icarus* 147, p220-250
- Lellouch et al., 2008, *Planetary and Space Science* 56, p1355-1367
- Lockwood and Jerzykiewicz, 2006, *Icarus* p442-452.
- Lunine et al., 2010, *JWST Planetary Observations within the Solar System*, White Paper
- Madhusudhan 2012, *Astrophysical Journal* 758, p36
- Melin et al., 2011, *MNRAS* 410, p641-644.
- Mizuno 1980, *Progress in Theoretical Physics* 64, p544-557
- Mommert et al. 2012, *A&A*, 541, A93
- Moses et al., 2000, *Icarus* 145, p166-202
- Müller et al 2009. *Earth, Moon, and Planets*, 105, 209-219.
- Muller-Wodarg et al., 2006, *Icarus* 180, p147-160.
- Mumma et al., 1981, *Science* 212, p45-49
- Owen et al., 1999, *Nature* 402, p269-270
- Paige et al., 2010, *Science* 330, p479
- Pearl et al., 1979, *Nature* 280, p755-758.
- Pearl et al., 1991, *Journal of Geophysical Research* 96, p18921
- Schneider, T., and J. Liu, 2009, *Journal of the Atmospheric Sciences*, 66, 579-601.
- Sorahana & Yamamura, 2012, *ApJ*, 760, 151
- Spencer et al., 1999, *Science* 284, p1514.
- Spencer et al., 2005, *Icarus* 176, p283-304.
- Tamura et al., 2009, *SPICA joint European/Japanese Workshop, Key Sciences of SPICA Mission: Planetary Formation, Exoplanets, and our Solar System*
- Tao et al., 2012, *Icarus* 221, p236-247.
- Yelle and Miller, 2004, in *Jupiter: The Planet, Satellites and Magnetosphere*

IN SITU EXPLORATION OF THE GIANT PLANETS and an Entry Probe Concept for Saturn

A White Paper Response to ESA's Call for L-Class Science Themes



Spokesperson: Olivier Mousis

Institut UTINAM, Université de Franche-Comté, France

Email: olivier.mousis@obs-besancon.fr

Tel: +33 381 666 921

Coordinators: Olivier Mousis (University of Franche-Comté, France), Leigh N. Fletcher (Oxford University, UK)

Core Team: Kathrin Altwegg (U. of Bern, Switzerland), Nicolas André (IRAP, France), Michel Blanc (IRAP, France), Athena Coustenis (LESIA, France), Leigh N. Fletcher (Oxford University, UK), Daniel Gautier (LESIA, France), Wolf Dietrich Geppert (Stockholm University, Sweden), Tristan Guillot (OCA, France), Patrick Irwin (Oxford University, UK), Jean-Pierre Lebreton (LPC2E, France), Bernard Marty (CRPG, France), Andrew Morse (Open University, UK), Olivier Mousis (University of Franche-Comté, France), Carl Murray (Queen Mary, U. of London, UK), Agustin Sanchez-Lavega (Universidad del País Vasco, Spain), Jean-Marc Petit (University of Franche-Comté, France), François-Xavier Schmider (OCA, France), J. Hunter Waite (SWRI, USA), Peter Wurz (U. of Bern, Switzerland)

Contributors: Thibault Cavalié (LAB, France), Ioannis Daggis (University of Athens, Greece), Georg Fischer (Space Research Institute, Austria), Ravit Helled (Tel Aviv U., Israel), Ricardo Hueso (Universidad del País Vasco, Spain), Theodor Kostiuk (NASA Goddard, USA), Laurent Lamy (LESIA, France), Emmanuel Lellouch (LESIA, France), Raphael Moreno (LESIA, France), Matthew McKay Hedman (Cornell U., USA), Nadine Nettelmann (Universität Rostock, Germany), James O'Donoghue (U. of Leicester, UK), Miriam Rengel (Max Planck Institut, Germany), Aymeric Spiga (LMD, France), Thomas Spilker (Solar System Science and Exploration, USA), Mark Marley (NASA Ames, USA), Glenn Orton (JPL, USA), Matthew Tiscareno (Cornell U., USA)

List of Supporters: Mate Adamkovics (USA), M. Ali Dib (France), Sushil K. Atreya (USA), David Atkinson (USA), Kevin Baines (USA), George Bampasidis (Greece), Erika Barth (USA), Bruno Bézard (France), Anil Bhardwaj (India), Laurent Boireau (France), Scott Bolton (USA), Vincent Boudon (France), Cristelle Briois (France), Julie Castillo (USA), Sébastien Celestin (France), James Cho (UK), Régis Courtin (France), Jun Cui (China), Iannis Dandouras (France), Imke de Pater (USA), Michel Dobrijevic (France), Yves Ellinger (France), Scott Eddington (USA), François Forget (France), Thierry Fouchet (France), Robert Frampton (USA), Yang Gao (UK), Patrick Gaulme (USA), Davide Grassi (UK), Denis Grodent (Belgium), Sandrine Guerlet (France), Leonid Gurvits (The Netherlands), Jason Jackiewicz (USA), Franck Hersant (France), Sarah Hörst (USA), Hauke Hussmann (Germany), Guy Israel (France), Caitriona Jackman (UK), Ralf Kaiser (USA), Ozgur Karatekin (Belgium), Philippe Lamy (France), Yves Langevin (France), Denis Lebleu (France), Jérémy Leconte (France), Mark Leese (UK), Anny-Chantal Levasseur-Regourd (France), Jianping Li (China), Sanjay Limaye (USA), Timothy Livengood (USA), Jonathan I. Lunine (USA), Anni Määtänen (France), Paul Mahaffy (USA), Jean-Pierre Maillard (France), Javier Martin-Torres (Spain), Stéphane Mathis (France), Henrik Melin (UK), David Mimoun (France), Karl L. Mitchell (USA), Giuseppe Mitri (Italy), Manuel Moreira (France), Neil Murphy (USA), Richard Nelson (UK), Conor Nixon (USA), Toby Owen (USA), Manish Patel (UK), Françoise Pauzat (France), Joel Poncy (France), François Poulet (France), François Raulin (France), Peter Read (UK), Henri Rème (France), Kim Rey (USA), Frank Robb (USA), Francis Rocard (France), Rafael Rodrigo (Spain), Sébastien Rodriguez (France), Philippe Rousselot (France), Heikki Salo (Finland), Sho Sasaki (Japan), Joachim Saur (Germany), Ella Sciamma O'Brien (USA), Juergen Schmidt (Finland), Anezina Solomonidou (Greece), Linda Spilker (USA), Larry Sromovsky (USA), Tom Stallard (UK), Daphne Stam (The Netherlands), Kathrin Stephan (Germany), Nick Teanby (UK), Gabriel Tobie (France), Elisabeth Turtle (USA), Pierre Vernazza (France), Paul Withers (USA), Ian Wright (UK), Roger Yelle (USA)

In Situ Exploration of the Giant Planets and an Entry Probe Concept for Saturn

WHITE PAPER RESPONSE TO ESA CALL FOR LARGE-CLASS SCIENCE THEMES

Executive summary

Comparative studies of the elemental enrichments and isotopic abundances measured on the four giant planets would provide unique insights into the processes at work within our planetary system at the time of giant planet formation, providing an invaluable *window onto the earliest evolutionary stages of our diverse solar system*. In situ measurements via entry probes remain the only reliable, unambiguous method for determining the atmospheric composition from the thermosphere to the cloud-forming regions of their complex weather layers. Furthermore, in situ experiments can reveal the properties of planetary atmospheres (temperatures, densities, composition, clouds, winds, waves and mechanisms responsible for energy, momentum and material transfer across different atmospheric layers) to *provide ‘ground truth’ for orbital remote sensing*. Following the orbital reconnaissance of the Galileo and Cassini spacecraft, and the single-point in situ measurement of the Galileo probe to Jupiter, we believe that in situ measurement of a second giant planet (either Saturn or an ice giant) should be an *essential element of ESA’s future cornerstone missions*, providing the much-needed comparative planetology to reveal the origins of our outer planets. This *quest to reveal the origins of our solar system and the nature of planetary atmospheres* drives to the heart of ESA’s Cosmic Vision, and has vast implications for the origins of planetary systems around other stars. Furthermore, such a mission would *build on ESA’s successful heritage* following the Huygens lander on Titan, and kick-start technology development for the in situ exploration of other planetary bodies (e.g., Venus). This White Paper presents the top-level science questions for in situ exploration of a giant planet, and in the last section we focus on *an entry probe for Saturn as the logical next step* beyond the Galileo probe to Jupiter and the Cassini orbital exploration of Saturn. The development of giant planet entry probe technology within Europe will pave the way for future collaborative efforts with NASA, who are pursuing their own studies of

thermal protection requirements at a variety of planetary targets, and will also build upon ESA’s Concurrent Design Facility reports into entry probe mission scenarios and previous Cosmic Vision concepts (KRONOS, Marty et al., 2009) for the in situ exploration of Saturn.

1. Motivation and background

Giant planets contain most of the mass and the angular momentum of our planetary system and consequently they must have played a significant role in shaping the architecture of our planetary system and the evolution of the smaller, inner worlds. Furthermore, the formation of the giant planets affected the timing and efficiency of volatile delivery to the Earth and other terrestrial planets (Chambers and Wetherill 2001). Therefore, understanding giant planet formation is essential to understand the origins and evolution of the Earth and other terrestrial planets capable of maintaining life in the form of complex organisms. The origins of the giant planets, their influence on planetary system architectures, and the plethora of physical and chemical processes at work within their atmospheres, make these destinations crucial for future exploration. *This White Paper advocates in situ exploration of giant planets to address questions at the heart of ESA’s Cosmic Vision*. There is a fundamental difference between the interiors of Jupiter and Saturn and those of Uranus and Neptune. Because Jupiter and Saturn have massive envelopes essentially composed of hydrogen and helium and a relatively small core, they are called *gas giants*. Meanwhile, data indicate that Uranus and Neptune also contain hydrogen and helium atmospheres but unlike Jupiter and Saturn, their H₂ and He mass fractions are smaller (5 to 20%). They are called *ice giants* because their density is consistent with the presence of a significant fraction of ices/rocks in their interiors. Despite this apparent grouping into two classes of giant planets, the four giants likely exist on a continuum, each a product of the particular characteristics of their formation environment. Comparative planetology of the four giants is therefore essential to reveal the formational, migrational and evolutionary processes

at work in the early solar nebula. ***This White Paper describes the science themes to be addressed by the future probe exploration of the giant planets of our Solar System. In this context, we outline the importance of in situ exploration of Saturn*** in the spirit of the previous ESA Cosmic Vision concept KRONOS (Marty et al. 2009) as a counterpoint to Galileo's in situ exploration of Jupiter, in addition to the science themes to be addressed ***by future probe exploration of an ice giant.***

Remote-sensing observations have always been the favoured approach of astronomers for studying the giants of our Solar System. However, the efficiency of this technique has some limitations when used to study the bulk atmospheric composition crucial to the understanding of planetary origins, namely due to degeneracies between temperatures, clouds and abundances in shaping the emergent spectra. A remarkable example of these restrictions is illustrated by the exploration of Jupiter, where key measurements such as the determination of the noble gases and helium abundances have only been made in situ by the Galileo probe. These measurements revealed unexpected results concerning the Ar, Kr and Xe enrichments with respect to their solar abundances, which suggest that the planet accreted icy planetesimals formed at temperatures possibly as low as 20-30 K to allow the trapping of these noble gases. Another remarkable result was the determination of the Jovian helium abundance obtained by a dedicated instrument aboard the Galileo probe (von Zahn et al. 1998) with an accuracy of 2%. Such accuracy on the He/H₂ ratio is impossible to derive from remote sensing, irrespective of the giant planet being considered, and yet precise knowledge of this ratio is crucial for the modelling of giant planet interiors and thermal evolution. The Voyager mission has already shown that these ratios are far from being identical, which presumably result from their evolution. An important result also obtained by the mass spectrometer aboard the Galileo probe was the determination of the ¹⁴N/¹⁵N ratio, which suggested that nitrogen present in Jupiter today originated from the solar nebula essentially in the form of N₂ (Owen et al. 2001). The mass spectrometer aboard Galileo unfortunately did not make measurements at levels deeper than 22 bars, precluding us from determining the H₂O abundance at levels representative of the bulk oxygen enrichment of the planet. Furthermore, the probe descended into a region depleted in certain volatiles and gases by unusual 'hot spot' meteorology (Orton et al. 1998; Wong et al. 2004), so may not be representative of the planet as a whole.

Nevertheless, the Galileo probe provided a giant step forward our understanding of Jupiter, but one can wonder if these measurements are really representative or not of the whole set of giant planets of the solar system. ***In situ exploration of more than one giant is the only way to address this crucial question for planetary science.*** In situ exploration of giant planet atmospheres addresses two broad themes, each of which we will explore in turn:

- A. Formation history of our solar system investigated by comparing bulk elemental enrichments and isotopic ratios;***
- B. Atmospheric processes (dynamics, waves, circulation, chemistry and clouds) from the upper atmosphere to below the cloud tops.***

Both themes have relevance far beyond the leap in understanding gained about an individual giant planet – the stochastic and positional variances produced within the solar nebula, the depth of the zonal winds, the propagation of atmospheric waves, the formation of clouds and hazes, disequilibrium processes of photochemistry and vertical mixing are common to all planetary atmospheres, from terrestrial planets to gas and ice giants, to brown dwarves and hot extrasolar planets. ***To progress beyond the Galileo probe findings, we now require a point of comparison.*** This could either be a second gas giant, to test theories of the joint origins of the two largest planets, or one of the ice giants, to assess their compositional differences compared to Jupiter. Despite the wealth of remote sensing data returned by the Cassini spacecraft for Saturn, several key questions still require an in situ probe to answer. ***This White Paper advocates any giant planet mission that incorporates elements of in situ exploration, whether for an ice or a gas giant. This paper also focuses on the description of a Saturn probe scenario because such a mission will appear the next natural step beyond Galileo's in situ exploration of Jupiter, and the Cassini spacecraft's orbital reconnaissance of Saturn.*** Moreover, many of the scientific questions outlined below remain valid additions to the cases for Uranus and Neptune orbital exploration (see White Papers by Arridge et al., Masters et al.). In the following, we detail why in situ exploration is vital to understand giant planet formation and atmospheric processes from the thermosphere to deep below the clouds, and we state the case specifically for a Saturn probe as a vital

comparison to the Galileo results. We also provide examples of approach phase science that could enhance the scientific return of any in situ mission.

2. Primary Science Themes

Theme A: Planet Formation and the Origin of the Solar System

Giant planets formed 4.55 Gyr ago, from the same disk of gas and solids that formed the Sun and eventually the entire Solar System. A significant fraction of their mass is composed of hydrogen and helium, the two lightest and most abundant elements in the Universe. Disks dominated by hydrogen and helium are almost ubiquitous when stars appear, but their lifetimes do not exceed a few million years. This implies that the gas giants Jupiter and Saturn had to form rapidly in order to capture their hydrogen and helium envelopes, more rapidly than the terrestrial planets which took tens of millions of years to attain their present masses, and retained only negligible amounts of the primordial gases as part of their final composition. Due to formation at fairly large radial distances, where the solid surface density is low, the ice giants Uranus and Neptune had longer formation timescales (slow growth rates) and did not manage to capture large amounts of hydrogen and helium before the disk gas dissipated. As a result, the masses of their gaseous envelopes are small compared to their ice/rock cores. *A comparative study of the properties of these giant planets thus gives information on spatial gradients in the physical/chemical properties of the solar nebula as well as on stochastic effects that led to the formation of the Solar System.*

Data on the composition and structure of the giant planets, which hold more than 95% of the non-solar mass of the Solar System (Marty et al. 2009), remain scarce, despite the importance of such knowledge. The formation of giant planets is now largely thought to have taken place via the core accretion model in which a dense core is first formed by accretion and the hydrogen-helium envelope is captured after a critical mass is reached (e.g. Mizuno 1980; Pollack et al. 1996). Once accounting for planet migration (e.g. Lin and Papaloizou 1986; Ward 1997), such a model can explain the orbital properties of exoplanets, although lots of unresolved issues remain (e.g. Ida and Lin 2004; Mordasini et al. 2012). However, an alternative scenario for the formation of giant planets remains the disk instability model (e.g. Boss 1997,

2001), in which they form from the direct contraction of a gas clump resulting from local gravitational instability in the disk. In principle, measurements of bulk elemental enrichments and isotopic ratios would help us to *distinguish between these competing formation scenarios*.

Formation and evolution models indicate that the total mass of heavy elements present in Jupiter may be as high as 42 *Earth-masses* (hereafter EM) whereas the mass of the core is estimated to range between 0 and 13 EM (Saumon and Guillot 2004). In the case of Saturn, the mass of heavy elements can increase up to 35 EM with the mass in the envelope varying between 0 and 10 EM and the core mass ranging between 0 and 20 EM (Helled and Guillot 2013). The masses of heavy elements are found to be in the 10.9-12.8 and 12.9-15.2 EM ranges for Uranus and Neptune, respectively (Helled et al. 2011). Direct access to heavy materials within giant planet cores to constrain these models is impossible, so we must use the composition of the well-mixed troposphere to infer the properties of the deep interiors. *Remote sounding cannot provide the necessary information* because of a lack of sensitivity to the atmosphere beneath the cloudy, turbulent and chaotic weather layer. These questions must be addressed by in situ exploration.

The availability of planetary building blocks (metals, oxides, silicates, ices) is expected to vary with position within the original nebula, from refractories in the warm inner nebula to a variety of ices of water, CH₄, CO, NH₃, N₂ and other simple molecules in the cold outer nebula. Turbulent radial mixing, and the evolution of the pressure-temperature gradient in the disk could have led to distinct regions where some species dominated over others (e.g., the water-ice snow line or N₂ over NH₃). Furthermore, both inward and outward migration of the giants during their evolution could have provided access to different material reservoirs at different epochs. A giant planet's bulk composition therefore depends on the timing and location of planet formation, subsequent migration and the delivery mechanisms for the heavier elements. By measuring a giant planet's chemical inventory, and contrasting it with measurements of (i) other giant planets, (ii) primitive materials found in comets and asteroids, and (iii) the abundance of our parent star and the local interstellar medium, can reveal much about the *conditions at work during the formation of our planetary system*. Furthermore, comparison to the compositions of the larger ensemble of extrasolar giant planets would place our own planetary origins in a broader context.

Galileo at Jupiter: To date, the Galileo probe at Jupiter (1995) remains our only data point for interpreting the bulk composition of the giant planets. Galileo found that Jupiter exhibited an enrichment in carbon, nitrogen, sulfur, argon, krypton and xenon compared to the solar photospheric abundances, with some notable exceptions – water was found depleted, may be due to meteorological processes at the probe entry site; and neon was depleted, possibly due to rain-out to deeper levels (Niemann et al. 1998, Wong et al. 2004). In any case the oxygen abundance in Jupiter remains an enigma. The Juno mission, which will arrive at Jupiter in 2016, may provide an estimate of the tropospheric O/H ratio. Interestingly, the nitrogen isotope composition of Jupiter is similar within errors to the protosolar nebula value (Marty et al. 2011) whereas the N isotope composition of comets is very different (enriched in ^{15}N by a factor of two). Explaining the high abundance of noble gases requires either condensing these elements directly at low temperature in the form of amorphous ices (Owen et al. 1999), trapping them as clathrates in ices (Gautier et al. 2001; Hersant et al. 2008; Mousis et al. 2009, 2012) or photoevaporating the hydrogen and helium in the protoplanetary disk during the planet's formation (Guillot and Hueso 2006). The Galileo measurements at Jupiter also include a highly precise determination of the planet's helium abundance, crucial for calculations of the structure and evolution of the planet. Figure 1 represents fits of the volatile enrichments measured

at Jupiter in the context of two different formation models, both being based on the hypothesis that Jupiter's building blocks formed from a mixture of rocks and crystalline ices but postulating a different oxygen abundance in the formation zone of Jupiter in the primordial nebula. While the quality of the matching of the volatile abundances is fairly similar, these two scenarios provide different predictions of the oxygen abundance in Jupiter. These calculations illustrate the strong connection between the formation circumstances of the planet and its bulk composition, and similar measurements for Saturn or an ice giant would enable comparison of their formation mechanisms to Jupiter.

Saturn Probe: Because of the absence of in situ measurements, the noble gas abundances are unknown in Saturn. However there is some indication for a non-uniform enrichment in C, N and S. Hersant et al (2008) suggest that ground-based and space-based (Cassini) observations are well fitted if the atmospheric carbon and nitrogen of the planet were initially mainly in reduced forms at 10 AU in the solar nebula. Alternatively, Mousis et al. (2009) find that it is possible to account for the volatile enrichments in Saturn in a way that is consistent with those measured at Jupiter if the building blocks of the two planets shared a common origin. *A determination of the oxygen abundance on Saturn via in situ exploration would distinguish between these scenarios.* Furthermore, *a determination of*

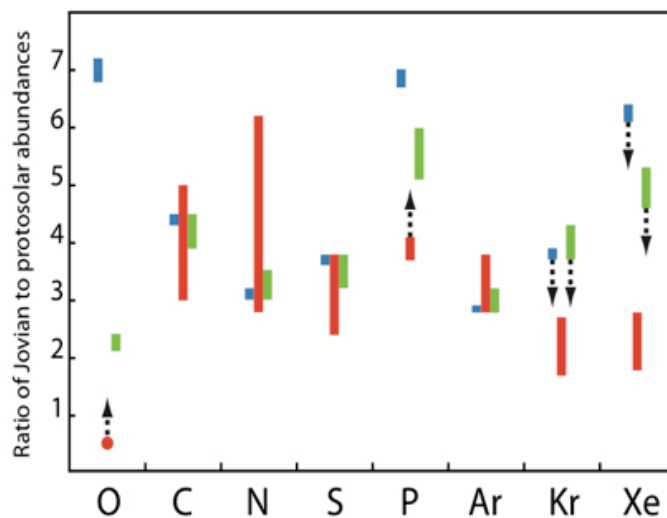


Figure 1. Ratio of Jovian to protosolar abundances (Mousis et al. 2012). Red bars and the red dot correspond to observations made by the Galileo probe. Green and blue bars correspond to calculations based on an oxygen abundance that is 0.5 and 1 times the protosolar value in the feeding zone of Jupiter, respectively. The corresponding oxygen abundances are predicted to be about 2 and 7 times protosolar in the Jovian atmosphere. Arrows pointing up correspond to the possibility that the measured oxygen and phosphorus abundances are lower than their bulk abundances, and arrows pointing down to the possibility that planetesimals could be impoverished in krypton and xenon (Pauzat et al. 2013).

noble gases is indispensable to understand the formation conditions of Saturn. On one hand, Hersant et al. (2008) predict that Ar and Kr should be solar in Saturn while Xe could be supersolar, whereas Mousis et al. (2009) find that all these species should be significantly supersolar. In addition, a determination of the volatile enrichments in Saturn could also provide a constraint on its rotation period, which will help to better infer its internal structure (Helled and Guillot 2013; Nettelmann et al. 2013). Moreover, as Saturn's atmosphere is believed to be depleted in helium as a result of H/He phase separation and subsequent helium rain, a precisely measured He/H₂ value of Saturn's atmosphere is crucial for probing the theoretical H/He demixing phase diagram, which is impossible with current laboratory technology for high-pressure physics. Helium-rain has long been predicted to occur in Saturn as an explanation for its high luminosity. Therefore, an entry probe measurement of the helium abundance is required to resolve this riddle.

Ice Giant Probes: Compared to the two gas giants, only a small amount of the solar nebula gas appears to have been accreted by Uranus and Neptune (about 3.6 EM for Uranus and 4.2 EM for Neptune; Helled et al. 2011), implying that elemental enrichments and isotopic ratios could vary significantly from the gas to the ice giants, indicative of their different formational mechanisms. Hersant et al. (2004) have calculated that Ar and Kr should be solar in the hydrogen envelope, while the Xe enrichment would be oversolar by a large factor. Therefore comparative measurements of noble gases in Saturn and the ice giants, for comparison to the Galileo probe results, should provide a firm representation of the theory of

volatile enrichments with respect to the Sun. A very interesting peculiarity of the composition of Uranus and Neptune is that microwave observations have revealed a very large oversolar S/N ratio (Gulkis et al. 1978). This has been interpreted as resulting from the formation of NH₄SH clouds in the troposphere from the combination of NH₃ in H₂S, but it should be noted that models fitting the microwave spectra leave room for significant ambiguity, which will only be resolved by in situ sampling. For Saturn and the ice giants, a precise determination of the He/H₂ is necessary for constraining the models of interiors of Uranus and Neptune and their cooling history since formation (i.e., to explain their intrinsic luminosities). Following the in situ exploration of Saturn, we propose that the second priority for giant planet exploration features **future bulk compositional measurements of an ice giant**, in tandem with orbital exploration, to reveal why their evolution diverged so substantially from that of the gas giants. Table 1 shows the measured abundances of the heavy elements (relative to solar) and several key isotopes at Jupiter, Saturn, Uranus and Neptune. On the three latter planets, reliable data concerning elemental abundances are much more limited, due to the absence of situ measurements and substantial ambiguities in the remote sensing data.

Element	Jupiter/Sun	Saturn/Sun	Uranus/Sun	Neptune/Sun
He	0.8 ^(a)	0.6–0.9 ^(a)	0.92–1 ^(a)	0.9–1.0 ^(a)
Ne	0.59 ^(a)	?	?	?
O	0.3–0.7 ^(b) (*)	?	?	?
C	3–5 ^(b)	8.8–9.6 ^(c)	20–30 ^(a)	30–50 ^(a)
N	2.8–6.2 ^(b)	?	?	?
S	2.4–3.8 ^(b)	?	?	?
P	3.7–4.1 ^(b)	8.9–13.5 ^(c)	?	?
Ar	2.8–3.8 ^(b)	?	?	?
Kr	1.7–2.7 ^(b)	?	?	?
Xe	1.8–2.8 ^(b)	?	?	?
Isotope	Jupiter	Saturn	Uranus	Neptune
D/H	2.6 +/-0.7E-5 ^(a)	2.3 +/-0.4E-5 ^(a)	4.4 +/-0.4 E-5 ^(d)	4.1 +/-0.4E-5 ^(d)
³ He/ ⁴ He	1.7 +/-0.0E-5 ^(a)	?	?	?
¹⁵ N/ ¹⁴ N	2.3 +/-0.3E-3 ^(a)	?	?	?

Table 1. Elemental abundances (relative to solar) and isotopic ratios measured in Jupiter, Saturn, Uranus and Neptune. ^(a) Atreya (2007), ^(b) Mousis et al. (2012), ^(c) Mousis et al. (2009), ^(d) Feuchtgruber et al. (2013), (*) this value is probably a lower limit.

Measurement Requirements for Constraining the Origins of Saturn and Ice Giants

The “ideal” measurement requirements are:

- The atmospheric fraction of He/H₂ with a 2% precision on the measurement;
- The elemental enrichments in cosmogenically abundant species C, H, O, N and S C/H, N/H, S/H and O/H should be sampled with a precision better than +/- 10%.
- The isotopic ratios in hydrogen (D/H), oxygen (¹⁸O, ¹⁷O and ¹⁶O), carbon (¹³C/¹²C) and nitrogen (¹⁵N/¹⁴N), to determine the key reservoirs for these species (e.g., delivery as N₂ or NH₃ vastly alters the ¹⁵N/¹⁴N ratio). ¹³C/¹²C, ¹⁸O/¹⁶O and ¹⁷O/¹⁶O should be sampled with a precision better than +/- 1%. D/H, ¹⁵N/¹⁴N should be analyzed in the main host molecules with a precision order of +/- 5%.
- The abundances and isotopic ratios for the chemically inert noble gases He, Ne, Xe, Kr and Ar, provide excellent tracers for the materials in the sub-reservoirs existing in the proto-solar nebula. The isotopic ratios for He, Ne, Xe, Kr and Ar should be measured with a precision better than +/- 1%.
- The elemental enrichments in minor species delivered by vertical mixing (e.g., P, As, Ge) from the deeper troposphere. P/H, As/H and Ge/H should be sampled with a precision better than +/- 10%.

The depth of probe penetration will determine whether it can access the well-mixed regions for key condensable volatiles. In the case of Saturn, a shallow probe penetrating to 5-10 bar would sample ammonia and H₂S both within and below their cloud bases, in the well-mixed regions of the atmosphere to determine the N/H and S/H ratios, in addition to noble gases and isotopic ratios. This would not be the case for Uranus and Neptune, where a shallow entry probe must focus on noble gas enrichments and

isotopic ratios. Such a probe would be able to sample condensable volatiles such as methane (condensing in the 1-2 bar region) and ammonia/H₂S above or possibly within their cloud bases (condensing in the 5-10 bar region), but would not reach the well-mixed tropospheres of the ice giants. Shallow entry probes would be unlikely reach the deep hypothesised water clouds on any giant planet, so that the deep O/H ratio would remain elusive unless accompanied by remote sensing experiments on an carrier spacecraft capable of probing these depths (e.g., the Juno microwave radiometer, currently en route to Jupiter). Of all these targets, a deeper entry probe on Saturn, penetrating to the 20-30 bar level, might reach the water cloud base to permit a precise measurement of Saturn’s O/H ratio, providing a strong *argument for targeting Saturn as the next destination for in situ exploration beyond the Galileo and Juno investigations of Jupiter*. Nevertheless, measuring elemental abundances (in particular He, noble gases and other cosmogenically-common species) and isotopic ratios using a shallow entry probe on Saturn or an ice giant would provide a vital comparison to Galileo’s measurements of Jupiter, and crucial ‘ground-truth’ for the remote sensing investigations of the Cassini spacecraft.

Theme B: Planetary Atmospheric Processes

Planetary atmospheres constitute our only accessible gateway to the processes at work within the deep interiors of the giant planets, and yet we must extrapolate from this thin, dynamic region over many orders of magnitude in pressure, temperature and density to infer the planetary properties deep below the clouds. Remote sensing provides insights into the complexity of the transitional zone between the external environment and the fluid interior, but there is much that we still do not understand. In situ measurements are the only method providing ground-truth to connect the remote-sensing inferences with the environmental conditions below the clouds, and yet this has only been achieved twice in the history of outer solar system exploration, via the Galileo probe for Jupiter and the Huygens probe for Titan. In situ studies provide *access to atmospheric regions that are beyond the reach of remote sensing*, enabling us to study the dynamical, chemical and aerosol-forming processes at work from the thermosphere to the troposphere below the cloud decks. The scientific objectives are summarised in Figure 2 and have relevance to both Saturn and the ice giants, but any of these targets would provide a fascinating comparison to the results of the Galileo probe at Jupiter. In this

theme, we demonstrate how in situ sampling addresses two crucial questions:

- *What processes are at work in planetary atmospheres, shaping the dynamics and circulation from the thermosphere to the deep troposphere?*
- *What are the properties and conditions for cloud formation as a function of depth and temperature in planetary atmospheres?*

Meteorology and Dynamics

Giant planets are natural planetary-scale laboratories for the study of geophysical fluid dynamics without the complicating influences of terrestrial topography, yet remote sensing only provides access to limited altitude ranges, principally via visible and infrared observations in the upper troposphere just above the condensate clouds and within the tropospheric hazes; and secondly in the mid-stratosphere near the 1-mbar level via mid-infrared emissions. Regions below the top-most clouds and in the middle/upper atmosphere are inaccessible, limiting our knowledge of the vertical variations of temperatures, densities, horizontal and vertical winds and waves, compositional profiles and cloud/haze properties. A

probe would be able to measure continuous profiles of these parameters during descent. Temperatures and densities in the upper atmosphere could be determined via the deceleration caused by atmospheric drag, connecting the high temperature thermosphere at nanobar pressures to the middle atmosphere at microbar and millibar pressures (e.g., Yelle et al. 2004). An atmospheric structure instrument, (e.g., Galileo/ASI, Seiff et al. 1998) would measure temperatures, pressure and densities throughout the descent to the clouds, sampling both the radiatively-cooled upper atmosphere and the convectively-cooled troposphere, precisely constraining the static stability, radiative-convective boundary (i.e., how far down does sunlight penetrate?) and the levels of the tropopause, stratopause, mesopause and turbopause. Thermal structure measurements of Saturn or an ice giant could be directly compared to those on Jupiter to ***understand the energetic balance between solar heating, thermal cooling, latent heat release, wave heating and internal energy*** for driving the complex dynamics of all the different atmospheric layers on the giants, and how this balance differs as a function of distance from the Sun.

Perturbations of the temperature structure due to vertical propagation of gravity waves are expected to be common features of the stably stratified middle atmospheres. Wave activity is thought to be a key

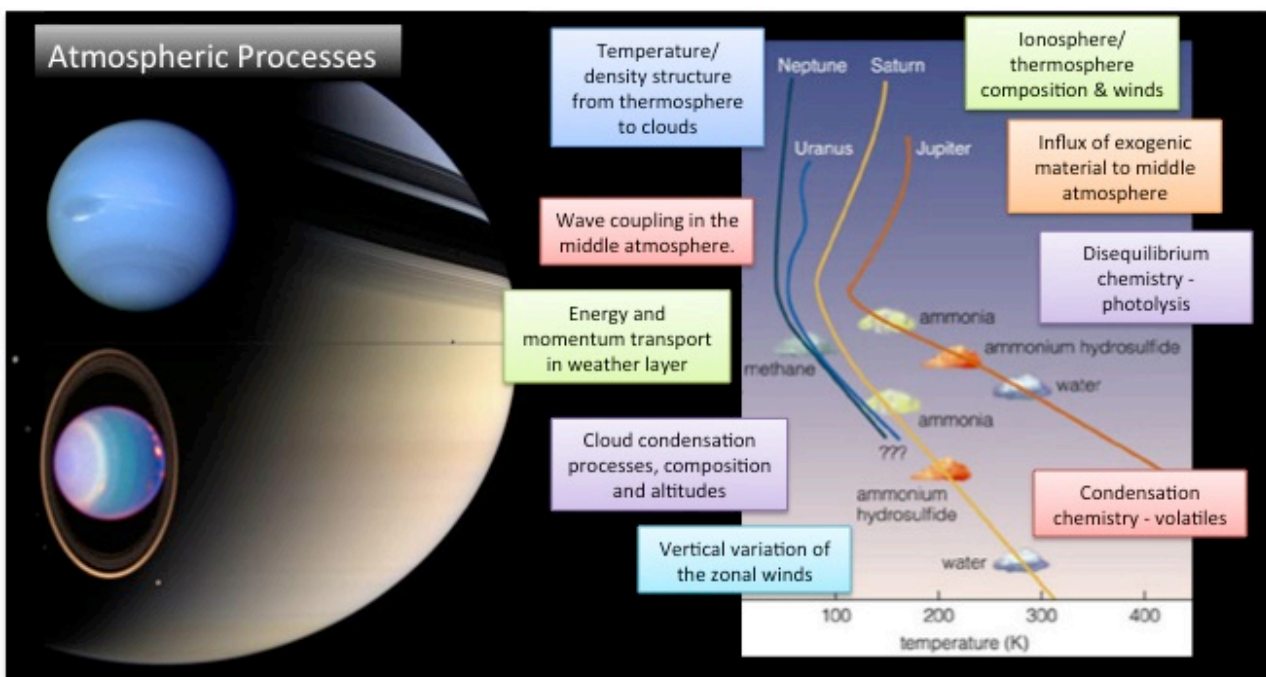


Figure 2. Examples of the vertical temperature structures of the giant planets, highlighting the scientific themes to be addressed via in situ remote sensing. Vertical profiles of temperature, density, radiant flux, chemical and aerosol composition will all be acquired during the descent of an entry probe from the upper atmosphere to depths within and below the cloud-forming regions.

coupling mechanism between the convective troposphere (e.g., gravity waves and Rossby/planetary waves radiated by rising plumes and vortices) and the stable middle/upper atmosphere, being responsible for transporting energy and momentum vertically through the atmosphere. Energy and momentum transfer via waves serves as a source of both heating and cooling for the hot thermospheres, whose temperatures far exceed the expectations from solar heating alone, although the precise origins of the heating source has never been satisfactorily identified (e.g., Hickey et al., 2000; Yelle et al. 2009). The periodicity of gravity waves measured by the Galileo probe on Jupiter has allowed us to reconstruct the zonal wind profile from the lower thermosphere to the upper troposphere (Watkins and Cho 2013), permitting identification of the turbopause (where molecular and eddy diffusion become comparable and gravity waves break to deposit their energy), above which the atmosphere separates into layers of different molecular species. ***Understanding the propagation, periodicity and sources of wave activity on the giant planets will reveal the properties of the background medium and the coupling of the ‘weather layer’ to the middle atmosphere,*** and facilitate direct comparison with Jupiter.

In situ exploration would also allow us to tackle one of the most enduring mysteries for the giant planets – ***what powers and maintains the zonal winds responsible for the planetary banding, how deep do those winds penetrate into the troposphere, and what are the wind strengths in the middle atmosphere?*** Remote sensing of temperature contrasts (and hence wind shears via thermal wind relationships), or inferences from the properties of atmospheric plumes at the cloud-tops (e.g., Sanchez-Lavega et al. 2008) cannot directly address this question, but in situ measurements of the vertical variation of winds on Saturn or an ice giant may help distinguish between shallow and deep models for this zonal organisation of the atmospheric flows. The Galileo probe’s Doppler Wind Experiment (DWE, Atkinson et al. 1998) reported that jovian winds were at a minimum at the cloud tops (where most of our understanding of zonal winds and eddy-momentum fluxes originates from), and increased both above and below this level. In the deep atmosphere, DWE demonstrated that the winds increased to a depth of around 5 bars, and then remained constant to the maximum probe depth of around 20 bars. Similar measurements on Saturn could ***sample the transition region between two different circulation regimes*** – an upper tropospheric region where eddies cause

friction to decelerate the zonal jets and air rises in cloudy zones, and a deeper tropospheric region where the circulation is reversed, eddy pumping is essential to maintain the jets and air rises in the warmer belts (e.g., del Genio et al. 2009; Fletcher et al. 2011). A single entry probe to either a mid-latitude belt or a zone would sample both regimes, albeit at different altitudes. Multiple entry probes would be highly desirable for any of the four giant planets, but is beyond the scope of the present white paper. Reconciling these two views of tropospheric circulation on Saturn would have implications for all of the giants, and provide crucial new information to solve a mystery left by the Cassini spacecraft. Alternatively, the measurement of the vertical variation of winds on an ice giant would ***establish the fundamental similarity or difference between circulation patterns on gas and ice giants***. Finally, direct measurements of winds in the middle atmospheres of any of these targets would establish the reliability of extrapolations from the jets in the cloud tops to the stratosphere in determining *the general circulations of planetary stratospheres*.

Composition and Clouds

In Theme A we demonstrated the need for reliable measurements of bulk elemental enrichments and isotopic ratios to study the formation and evolution of the giant planets. However, vertical profiles of atmospheric composition (both molecular and particulate) are essential to ***understanding the chemical, condensation and disequilibrium processes at work***. Jupiter’s atmospheric composition (e.g., Atreya et al. 1999) was measured by the Galileo probe’s mass spectrometer (Niemann et al. 1998), optical interferometer (specifically for a refractive index measurement to determine the helium abundance, von Zahn et al. 1998), net flux radiometer (NFR) and nephelometer (NEP), between pressures of 0.5 and 20 bar. The attenuation of the probe signal as it moved deeper into the atmosphere also revealed the density and composition. These instruments revealed an unexpectedly dry region of the jovian troposphere, depleted in clouds and volatiles, which was consistent with ground-based observations of the probe entry into a warm cyclonic vortex (e.g., Orton et al. 1998). For this reason, the compositional profiles measured by Galileo are not thought to be globally representative, leading to a desire for ***multiple entry probes for different atmospheric regions in future, more ambitious missions***. Nevertheless, a single atmospheric entry probe for Saturn or an ice giant would provide an intriguing counterpoint to Galileo’s sampling of Jupiter’s unusual meteorology.

A poor understanding of cloud and haze formation in planetary atmospheres of our solar system may be the key parameter limiting our ability to interpret spectra of extrasolar planets and brown dwarfs (e.g., Marley 2013). Although equilibrium cloud condensation models, combined with the sedimentation of condensates to form layers, have proven successful in explaining the broad characteristics of the planets (methane ice clouds on ice giants, ammonia ice clouds on gas giants), they apparently fail rather spectacularly in predicting the location, extent and microphysics of the observed cloud decks. The Galileo probe results defied expectations of equilibrium condensation by revealing clouds bases at 0.5, 1.3 and 1.6 bar, plus tenuous structure from 2.4-3.6 bar and no evidence for a deep water cloud (West et al. 2004). Ammonia ice on Jupiter has only been spectroscopically identified in small, localised regions of powerful convective updrafts (e.g., Baines et al. 2002). The spectral signature of pure ammonia ice is likely obscured by a coating or mixing with other products, such as photolytically produced hydrocarbons, hydrazine or diphosphine (e.g., Sromovsky et al., 2010; West et al., 2004). The spectral properties of these mixtures are poorly known, rendering cloud remote sensing highly ambiguous. The only way to resolve these questions is by *in situ sampling of the clouds and hazes formed in a planet's atmosphere*, using instruments designed to measure the particle size distributions, radii, number densities, mass densities, optical depth and size properties. Combined with the vertical profiles of condensable volatiles (e.g., CH₄ on the ice giants; NH₃, H₂S and H₂O on all the giants) and photochemically-produced species (hydrocarbons, hydrazine, diphosphene), this would give an estimate of the composition of giant planet condensation clouds and upper atmospheric hazes for the first time. Precise determination of bulk atmospheric composition is made difficult since all species in the upper troposphere are affected by fractionation and mixing. However, Saturn's atmosphere provides the most accessible cloud decks for this study (condensates of NH₃ and H₂O are locked away at considerably higher pressures on the ice giants), the most useful comparison to remote sensing data (e.g., from Cassini) and the most similar composition to Jupiter for a full understanding of gas giant clouds.

Volatiles are removed from the gaseous phase both by condensation, and by photolytic destruction. Indeed, all the giant planets exhibit a rich chemistry due to the UV photolysis of key atmospheric species. Their stratospheres are dominated by the

hydrocarbon products of methane photolysis (e.g., Moses et al. 2005), and their tropospheres by the photolysis of residual saturated ammonia and phosphine dredged from their deeper interiors by vigorous atmospheric mixing (e.g., Fletcher et al. 2009). Additional trace species in the troposphere (GeH₄, AsH₃, CO) provide constraints on the *strength of atmospheric mixing* from deeper, warmer levels below the clouds; unusual chemical products (HCN, HCP, CS) *reveal coupled chemistry* due to lightning activity or shock chemistry due to planetary impacts; and oxygenated species in the high stratosphere (CO, CO₂, H₂O) reveal the *strength of exogenic influx* of materials (comets, interplanetary dust, e.g., Harrington et al., 2004) into the upper atmospheres. Sensitive mass spectrometry of these species, combined with probe measurements of atmospheric temperatures and haze properties, could *reveal the processes governing the soup of atmospheric constituents on the giant planets*. Once again, Saturn's trace species are expected to be the most accessible, as volatiles and disequilibrium species (e.g., PH₃ and NH₃) have so far eluded detection on the ice giants.

This rich chemistry extends into the thermospheres of the giants planets, where gaseous constituents are ionised by solar EUV radiation at mid-to-high latitudes and by charged-particle bombardment in the circumpolar regions. At lower altitudes where solar UV rays and charged particles do not penetrate, ionization by cosmic rays may become important (Capone et al. 1979). Photoionisation of CH₄ can lead to CH⁺, CH₂⁺ and CH₃⁺ ions that react efficiently with molecular hydrogen to form CH₃⁺ and CH₅⁺ ions. At lower altitudes CH₃⁺ ions can react with acetylene and methane to produce hydrocarbon ions containing more carbon atoms. The dissociative recombination of these species then leads to neutral hydrocarbons (Kim and Fox 1994) and thus starts a very complex hydrocarbon chemistry leading to e.g., benzene. Isotopic ratios measured in these ionic species could reveal their formational pathways. In situ measurements of this ion-neutral chemistry using mass spectrometers capable of distinguishing ions and isotopes of very similar masses would *revolutionize our understanding of upper atmospheric chemistry*, with capabilities far beyond that of the Galileo probe. *The possibility to identify ions and neutrals with high masses and to determine their abundances will allow a vastly improved understanding of important processes like the formation of aerosols and, subsequently, generation of haze and clouds.*

Measurement Requirements for the atmospheres of Saturn and ice giants

Although the science case differs slightly for Saturn and the ice giants due to the accessibility of certain species (e.g., methane condensation on the ice giants versus ammonia on the gas giants, and the improved detectability of tropospheric trace species on Saturn), we summarise the requirements for atmospheric science as follows:

- Determine the thermal and density profile from thermosphere to troposphere, and the balance between different energy sources controlling atmospheric dynamics and structure;
- Measure the strength of the winds as a function of altitude and the importance of wave perturbations on atmospheric structure;
- Sample and determine the properties of cloud and haze layers as a function of depth (e.g., methane and hazes on ice giants; NH_3 , NH_4SH and N- and P-bearing hazes on Saturn);
- Measure the vertical profiles of chemical products, disequilibrium species and ions to understand vertical mixing and atmospheric chemistry.

3. Ancillary Science Themes: Approach Phase & Carrier

In Section 2 we demonstrated the specific scientific case for in situ measurements by an atmospheric probe for Saturn or an ice giant. The measurement requirements specified above should be considered as essential for any giant planet entry probe mission. However, a mission flying to one of these targets presents a substantial opportunity for secondary science during the approach and flyby phases. In the following sections, we describe a subset of science themes related to the in situ atmospheric experiments: Doppler seismology to *probe the existence of a planetary core* and supplement Theme A (Origins); *atmospheric electricity to support the in situ exploration of the weather layer* in Theme B (Atmosphere); and suggestions for *in situ sampling of material associated with the rings and inner magnetosphere of Saturn* to study the connectivity between the planet and its immediate magnetospheric environment. A detailed trade study would be conducted for any future mission proposal to assess their feasibility, but this provides a taste of the opportunity that the giant planet probe mission represents.

Doppler Seismology Approach

As mentioned in Theme A, the present internal structure of giant planets is the result of their formation and evolution. Unveiling this internal structure would give unprecedented constraints both on the formation scenario and on the physics controlling its evolution. After decades of fruitless attempts, the seismology of giant planets is finally providing very exciting results. Ground-based observations demonstrated the existence of acoustic modes of $\sim 30\text{cm/s}$ amplitude and frequencies between 1 and 3 mHz on Jupiter (Gaulme et al.

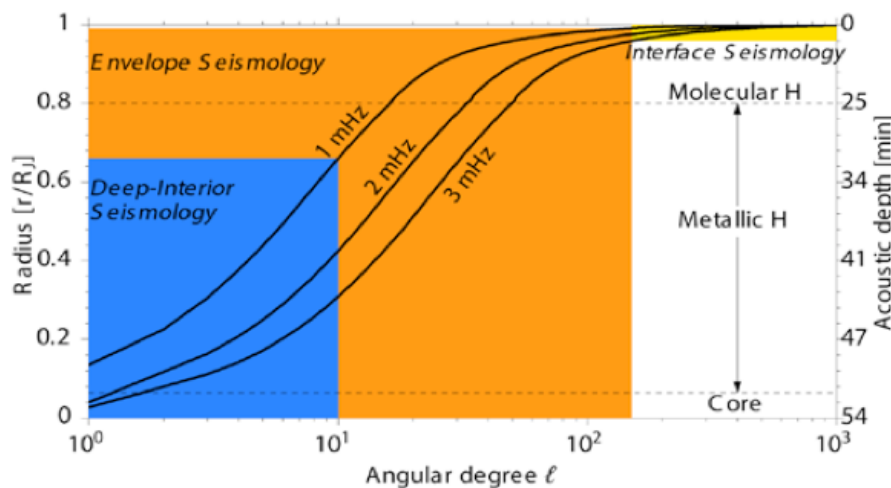


Figure 3. Internal turning point of acoustic oscillations as a function of the degree, for different frequencies in the expected, and observed range of excited modes (here calculated for Jupiter). It shows the zone inside the planet that can be explored following the angular resolution: $R/10$ in blue, $R/150$ in orange, higher resolution in yellow (local seismology).

2011), consistent with expectations (e.g., Vorontsov et al. 1976; Bercovici and Schubert 1987). Observations of stellar occultation by Saturn's rings, made by the Cassini spacecraft, exhibited density waves structures that can only be explained as a consequence of some of Saturn's oscillations modes resonating in the rings (Hedman and Nicholson 2013). This result was also predicted (Marley and Porco 1993). These discoveries are extremely exciting because they announce a revolution in the study of the interiors of these planets similar to that experienced for the Sun with solar seismology (e.g., Christensen-Dalsgaard 1996). Unfortunately, the data that can be acquired now, or even in the future, from the ground remains insufficient to get the full reward of the techniques.

Seismology requires *continuous observations of spatially resolved global modes for period of weeks, or months*. A *spacecraft approaching a planet would provide a long enough observing period of several months, with resolution sufficient to detect all modes with degree $l > 10$, sounding precisely the central part of the planet, particularly the core*. As the probe get closer, the observing time decreases, but the number of observed modes increases with the square of the inverse of the distance, giving access to the whole internal structure and rotation. At close distance, time-travel analysis (local seismology) will give estimation of the temperature and motions just below the surface that will help to understand the conditions found by the entry probe. At the same time, the same instrument would also provide a monitoring of the surface in the visible domain as well as instantaneous Doppler velocity maps, allowing the study of the global and detailed dynamics of the atmosphere, even in absence of clouds for tracking, and separation between waves and winds. In conclusion, the addition of a Doppler seismometer to the planetary probe carrier would provide complementary information on both origins and atmospheres for the in situ measurements of the probe.

Lightning and Atmospheric Electricity

The in situ exploration of a giant planet weather layer will provide new insights into the cloud-forming processes below the levels normally visible to remote sensing (Theme B). Lightning flashes most likely exist in the atmospheres of all gas planets (Yair et al. 2008), and the Galileo Probe lightning and radio emission detector (LRD) used a magnetic antenna to detect signals of lightning from Jovian clouds with an electric dipole moment change about 100 times that of terrestrial lightning (Rinnert et al. 1998). The

existence of lightning in Saturn's atmosphere has been proven by Voyager and Cassini measurements of radio emissions (Fischer et al. 2008) and direct optical flash observations (Dyudina et al. 2010). The thunderstorms tend to appear infrequently at the equator and in the "storm alleys" at the latitudes of 35° north and south. The flashes originate from a depth of 125-250 km below the 1-bar level, most likely in the water clouds. So far, Saturn lightning radio emissions have only been measured above the ionospheric cutoff frequency (~ 1 MHz). Measurements in the VLF region (3-30 kHz) can reveal the unknown spectrum at lower frequencies, where lightning radio emissions are expected to be strongest and to be able to propagate over thousands of kilometers below the ionosphere. Another unique and new measurement for gas planets concerns Schumann resonances in the TLF (< 3 Hz) and ELF regions (3-300 Hz), which should be excited by lightning in their gaseous envelopes (e.g. Sentman 1990). It has been suggested that such a measurement could even constrain the water abundance on giant planets (Simões et al. 2012), and it would be very useful in conjunction with conductivity measurements throughout the descent of the probe.

Ring Science

The ring systems of Jupiter, Saturn, Uranus and Neptune show diverse physical and dynamical properties. The rings of Jupiter and Saturn have vertical corrugations likely due to recent impact events (Showalter et al. 2011; Hedman et al. 2011) while direct observations of impacts on Saturn's rings (Tiscareno et al. 2013) have constrained the impact flux in the outer solar system. For all the ring systems, timescale problems suggest either a relatively recent origin or a continuous source of material from nearby satellites. Results from the Cassini-Huygens mission have revolutionised our understanding of Saturn's rings, having provided the opportunity to study in detail the closest example of an astrophysical disk as it evolves (Cuzzi et al. 2010). Key discoveries include (i) observations of "propellers" in Saturn's A ring (Tiscareno et al. 2006) and their orbital migration (Tiscareno et al. 2010), (ii) the detection of self-gravity wakes (Colwell et al. 2006) and (iii) observations of objects forming in the F ring (Beurle et al. 2010) and colliding with it (Attree et al. 2012). Despite its success, the Cassini mission cannot address the question of the physical appearance and composition of ring particles and the exact nature of their localised clustering in self-gravity wakes, which are crucial to determining the origin and lifetime of the rings.

Voyager images and recent Keck observations reveal that Saturn has a major interaction between its atmosphere and ring-dominated inner magnetosphere. Charged water particles bound to magnetic field lines “rain” down on to the ionosphere of Saturn from the rings, causing the appearance of dark bands in Saturn's upper atmosphere (Connerney 1986). In the dark regions, the influx of charged water products reacts with the ionosphere (Jontof-Hutter and Hamilton 2012a, 2012b; O’Donoghue et al. 2013). This “ring-rain” must act to alter the chemistry and temperature of Saturn's atmosphere, but current and future telescopes cannot gather the required light to study this unique interaction much further. *In situ measurements of this material flux are absolutely essential in order to understand how the atmosphere responds to its planetary ring system raining into it.* Our understanding of the evolution and lifetime of the rings as well as the influence of the rings on the upper atmosphere will be enhanced by both the carrier and in situ phase of this mission.

Magnetic dynamo, magnetosphere and radiation environment

The relative orientations of the rotation and the magnetic dipole axis and the direction of the solar wind flow lead to important differences in all the planetary magnetospheres in our Solar System (Russell and Dougherty 2010). Jupiter and Saturn have giant magnetospheres shaped by a broad diversity of internal plasma sources and the fast planetary rotation. Uranus and Neptune (see the

white papers by Arridge et al. and Masters et al.) have unusual asymmetric magnetospheres that result from the significant tilt between the planetary magnetic dipole and the rotation axis, and the existence of significant multipolar magnetic fields.

Saturn's rich magnetospheric environment is unique in the solar system. It strongly interacts with all other components of the saturnian system: the planet, its rings, numerous satellites (icy moons and Titan) and various dust, neutral and plasma populations. In the innermost regions of this System, the main rings are a very strong absorber of energetic particles so that Saturn's main radiation belts stop exactly at the outer edge of the main rings. A second radiation belt has however been discovered planetward of the D ring (Krimigis et al. 2004). The saturnian magnetosphere is shaped by a nearly axisymmetric intrinsic planetary magnetic field, the rotational and magnetic axis of Saturn being almost aligned, but a strong modulation related to a longitudinal asymmetry of yet unknown nature was detected in SKR emissions (Carrery and Mitchell 2013). Very close magnetic field measurements are needed to construct a more complete intrinsic magnetic field model, with more high degree moments being better resolved, and decipher the unique dynamo that may operate at Saturn (Cao et al. 2012). Accurate measurements of the magnetic field very close to Saturn's atmosphere and thus unfiltered by the ring is a key to both the composition and conductivity of Saturn's interior.

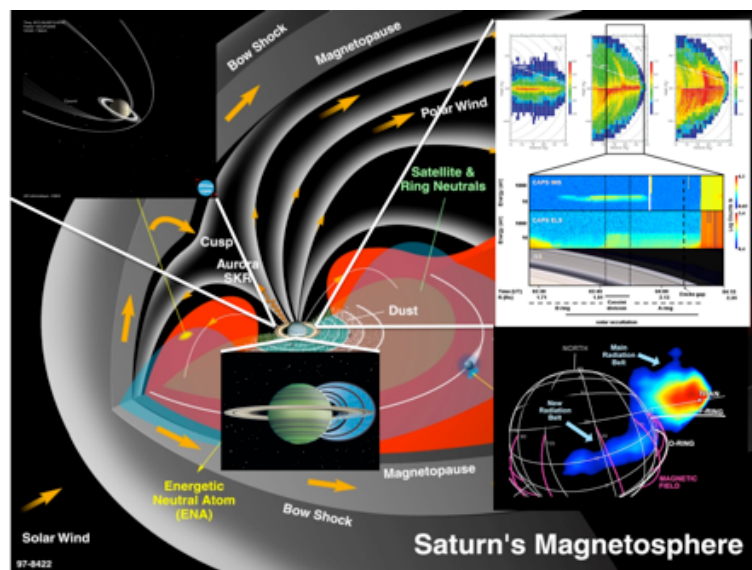


Figure 4. Saturn's magnetospheric regions and processes. Bottom left inset ring ‘rain’ onto Saturn's low-latitude ionosphere (O’Donoghue et al. 2013); Bottom right inset new radiation belt observed in ENA (Krimigis et al. 2005); Top left inset Cassini proximal orbits (3DView/CDPP); Top right inset ring ionosphere and exosphere (André et al. 2008).

The Cassini-Huygens mission has been expanding our understanding of the Saturn system since 2004. Juno-like inclined proximal orbits at the end of the Cassini Solstice mission in 2016-2017 will provide the opportunity to study in-depth the interaction between the rings, ionosphere and magnetosphere, and the formation of the ring atmosphere/ionosphere and its coupling to Saturn's ionosphere (Spilker et al. 2009). On the other hand, taking advantage of an instrumented carrier for a Saturn probe would enable us to *obtain in situ magnetospheric observations in the innermost unexplored regions of Saturn's magnetosphere*. These unique observations would complement the ones obtained by Cassini at the end of its mission (including the very close magnetic field measurements) and, by analogy, refine our understanding of magnetized protoplanetary disks.

4. Mission Architectures

The primary science objectives described in Section 2 can be addressed by an atmospheric entry probe that would descend under parachute, and start to perform *in situ* measurements in the stratosphere as soon as feasible after parachute deployment and continue in the troposphere down to a minimum of 10 bars. The 10 bar value is just a representative value. Future proposals would conduct a careful study of the trade-offs between the science return and the added complexity of a probe than could operate down to higher pressures to determine the deep oxygen abundances. Accelerometry measurements would also be performed during the entry phase to probe the upper layers of the atmosphere prior to starting *in situ* measurements under parachute. A spacecraft carrier would be required in order to bring the probe to its target planet. The range of possible carriers include:

- A carrier (carrier option 1) that would detach prior to probe entry, follow the probe path and destroy itself when entering the atmosphere;
- A fly-by carrier (carrier option 2) that would release the probe several months prior to probe entry and deflect its trajectory to be used for both probe data relay and for performing flyby science;
- An orbiter, which would provide a similar configuration to that of the Galileo orbiter/probe mission. The carrier would be placed in orbit after the probe relay is over and would perform

The mission architecture would depend on the chosen target, taking into account the orbital science that has already been performed by Cassini, but it remains to be done at Uranus or Neptune.

Saturn mission reference case:

Taking the reference case of a future mission to Saturn to describe a mission architecture concept, in addition to the entry probe, we do not consider an orbiter/probe configuration and briefly address the different mission architectures with the two carrier options. A mission architecture with the carrier option 1, if the carrier is instrumented properly, would allow performing approach science and *in situ* pre-entry science. In this architecture, the probe data transmission would solely rely on a Direct-to-Earth link. As an alternative, if the carrier would be separated early enough prior to probe entry and slowed enough to be delayed by a couple of hours, it could also be used as a probe radio relay. A mission architecture with the carrier option 2, would provide the capability to perform approach science (for months) and flyby science (for a few days) in addition to being used as the probe radio relay. It would also allow many retransmissions of the probe data to reduce risk. Alternative technologies, such as the solar photonic sail or solar wind electric sail (Janhunen et al. 2010, 2013), could be potentially used for the propulsion of the spacecraft.

Entry atmospheric probe: An entry probe to Saturn, Uranus or Neptune, would in many respects resemble the Jupiter Galileo probe. The concept was further developed for Saturn in the KRONOS mission proposal (Marty et al. 2009). Concept probe studies to the giant planet have been studied by ESA in

2010¹. As an example, the KRONOS probes had a mass of ~330kg, with a 220kg deceleration module (aeroshell, thermal protection system, parachutes and separation hardware) and a 117kg descent module, including the science instruments and subsystems). A representative payload for the Saturn probe that would allow addressing the measurement requirements identified for themes A and B are shown in Table 2.

In situ pre-entry science: In carrier option 1, some capability would exist to carry pre-entry science, and possibly approach science. Following probe release, any surplus ΔV could be used to alter the carrier trajectory towards the equatorial plane. A close approach to the rings increases the risk of destruction of the essentially now redundant carrier, but has the advantage of improving measurements concerning Ring science. In this scenario instruments are selected towards Science objectives that cannot be met by the probe (see Table 3).

Carrier/Flyby science: Carrier option 2 provides the opportunity to carry both approach science and flyby science. The flyby science observations would be aimed at addressing the science objectives addressed in Section 3. A list of instruments is given in Table 4.

Table 2. Measurement requirements identified for themes A and B.

Instrument	Measurement	Science Theme
Helium Abundance Detector	Accurate He/H ₂ ratio	A
Mass Spectrometer	Elemental & chemical composition	A
	Isotopic composition	A
	High molecular mass organics	B
Atmospheric Structure Instrument	Pressure, Temperature, Density, molecular weight profile	B
Highly sensitive accelerometer	High altitude Atmospheric structure (during entry phase)	B
Doppler Wind Experiment	Measure winds, speed and direction	B
Nephelometer	Cloud structure	B
	Solid/liquid particles	B
Net flux	Thermal/solar energy	B

¹ <http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=47568>

radiometer	profiles	
Lightning detector	Detect lightning, measure energetic particles	B

Table 3. Measurement requirements identified for *in situ* pre-entry science.

Instrument	Measurement	Science Theme
Camera	Probe entry	Context
Camera	Ring particles, close structure and size	Ring
Lightning detector	Low Frequency emissions (VLF, possibly TLF and ELF) water abundance	B
Fourier Spectro Imager	Doppler seismology	A/B
Spectrometers (IR/ UV)	Temperatures, Clouds and Chemical composition	B/Rings

Table 4. Measurement requirements identified for carrier/flyby science.

Instrument	Measurement	Science Theme
Camera	Probe entry	Context
Lightning detector	Low Frequency emissions (VLF, possibly TLF and ELF) water abundance	B
Fourier Spectro Imager	Doppler seismology	A/B
Spectrometers (IR/UVIS)	Temperatures, Clouds and Chemical composition	B/Rings
Magnetometer	Magnetic dynamo, magnetosphere and radiation environment	Ancillary

Carrier: The carrier (either for option 1 or 2) may benefit from subsystems developed for JUICE. In addition to studying specific carrier architectures, it is suggested to look into an approach similar to the one that allowed to develop MEX and VEX based on Rosetta by studying how the carrier design would benefit from the re-use of JUICE elements.

Power generation: It is believed that all mission architectures proposed can be solely designed on batteries and solar power, pending LILT qualification extension to 9 AU conditions. The probe will be powered with primary batteries as were the Galileo and Huygens probes. The carrier, in both options 1 and 2, would be equipped with a combination of solar panels, rechargeable batteries (option 1) and possibly a set of primary batteries for the phase that will require a high power demand. Nuclear power would be considered for the carrier only if available solar power technology would not work.

Interplanetary trajectory and entry zone of the probe: Many trajectory options have been identified, using both direct and gravity-assisted transfers to Saturn, and more would be identified in subsequent studies. Trajectory selection would consider all the trajectory options, the selected carrier option, and the launch vehicle capabilities available at the time. Different trajectories may be envisaged for carrier options 1 and 2. Concerning the entry zone, mid latitudes should be a safer destination for the probe. Volatile-depleted regions are probably located at the cyclones in both poles and may be at the storm-alley (region of low static stability able to develop updrafts and downdrafts). More generally, the peaks of westward jets can be unstable based on the stability of the wind system and eastward jets (particularly the anticyclonic branch of eastward jets) and might be good locations to retrieve the deep values of volatiles at higher levels in the atmosphere (Read et al. 2009).

International collaboration: One of the key probe technologies, which would be new for Europe industry, is the heat shield material for an entry probe into a giant planet. Careful trade-offs would have to be made for either development of this new technology within Europe or establishing collaboration with an international partner that may have this technology readily available. In particular, recent NASA studies have been made concerning the thermal protection requirements for a Saturn entry probe. International collaboration may also be looked at for other mission elements, including the ground segment.

history of our Solar System and the processes at work in the atmospheres of giants. We advocate that any giant planet mission incorporating elements of in situ exploration, whether for the gas or ice giants, should form an essential element of ESA's future cornerstone missions. *We describe the concept of a Saturn probe as the next natural step beyond Galileo's in situ exploration of Jupiter, and the Cassini spacecraft's orbital reconnaissance of Saturn.* Two missions designs are envisaged and derived from the KRONOS concept previously proposed to ESA. Both scenarios envisage the *transport of the probe to Saturn via a spacecraft carrier* that would detach either a few months before or just prior the probe entry. The first scenario would enable some in situ pre-entry science whereas the second scenario would allow both pre-entry and flyby science (ring science, Doppler seismology, magnetosphere science). In situ exploration builds on ESA's successful heritage with Cassini-Huygens, and paves the way for ESA leadership in future international collaboration in outer solar system exploration.

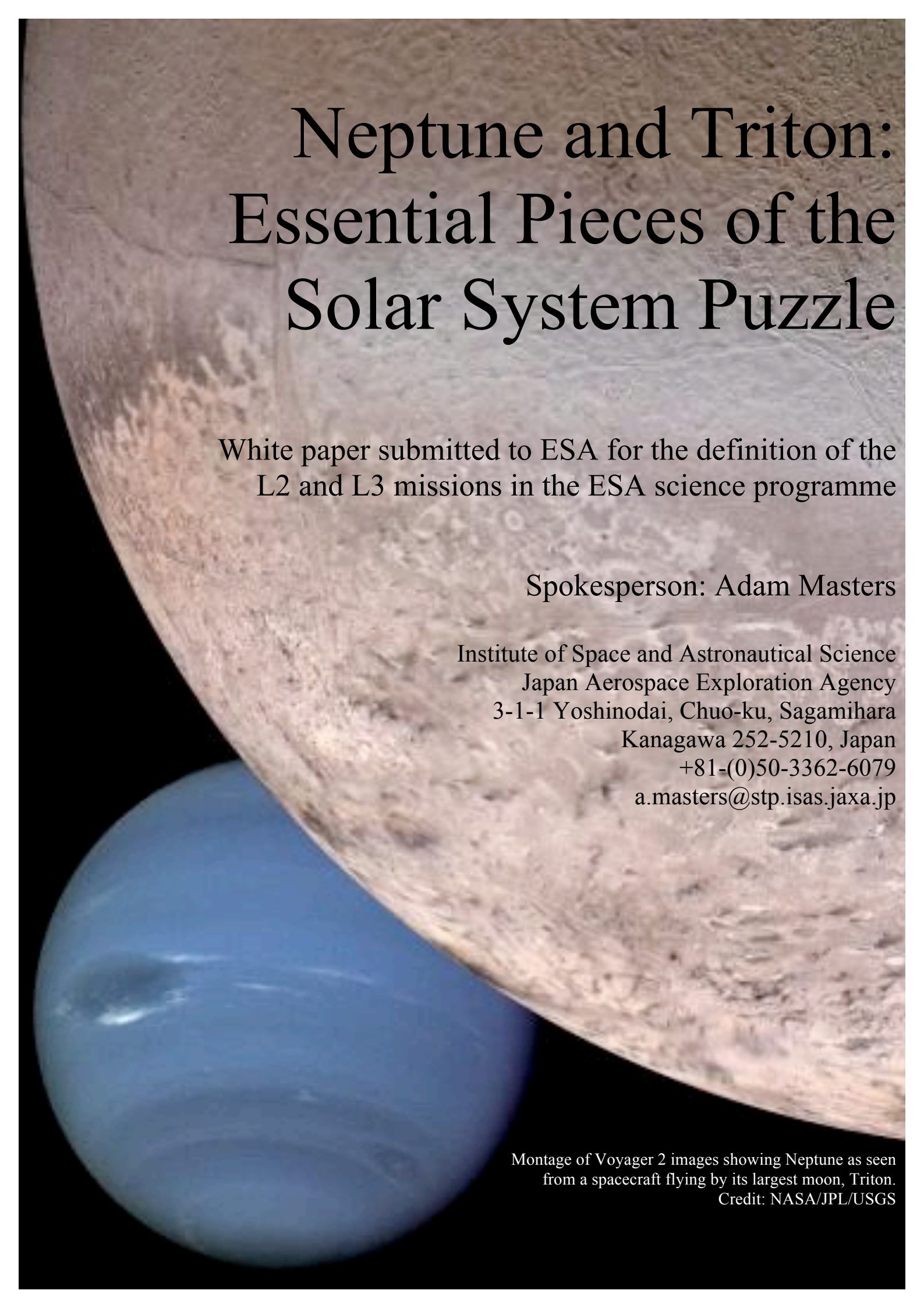
5. Conclusions

In this White Paper, we have shown that the in situ exploration of the giant planets of the Solar System can address two major science themes: *the formation*

6. References

- André N., et al., 2008, *RvGeo*, **46**, 4008
- Atkinson D. H., Pollack J. B., Seiff A., 1998, *JGR*, **103**, 22911
- Attree N. O., Murray C. D., Cooper N. J., Williams G. A., 2012, *ApJ*, **755**, L27
- Atreya S. K., Wong M. H., Owen T. C., Mahaffy P. R., Niemann H. B., de Pater I., Drossart P., Encrenaz T., 1999, *PSS*, **47**, 1243
- Atreya, S.K., 2007, *Proceedings of the 4th International Planetary Probe Workshop, NASA Jet Propulsion Laboratory Document*
- Baines K. H., Carlson R. W., Kamp L. W., 2002, *Icarus*, **159**, 74
- Bercovici D., Schubert G., 1987, *Icarus*, **69**, 557
- Beurle K., Murray C. D., Williams G. A., Evans M. W., Cooper N. J., Agnor C. B., 2010, *ApJ*, **718**, L176
- Boss A. P., 1997, *Science*, **276**, 1836
- Boss A. P., 2001, *Nature*, **409**, 462
- Cao H., Russell C. T., Wicht J., Christensen U. R., Dougherty M. K., 2012, *Icarus*, **221**, 388
- Capone L. A., Dubach J., Whitten R. C., Prasad S. S., 1979, *Icarus*, **39**, 433
- Carbary J., Mitchell D. G., *Rev. Geophys.*, 2013, submitted
- Chambers J. E., Wetherill G. W., 2001, *M&PS*, **36**, 381
- Christensen-Dalsgaard J., 1996, *NuPhS*, **48**, 325
- Colwell J. E., Esposito L. W., Sremcevic M., 2006, *GeoRL*, **33**, 7201
- Connerney J. E. P., 1986, *GeoRL*, **13**, 773
- Cuzzi J. N., et al., 2010, *Science*, **327**, 1470
- Del Genio A. D., Achterberg R. K., Baines K. H., Flasar F. M., Read P. L., Sanchez-Lavega A., Showman A. P., 2009, *sfch.book*, 113
- Dyudina U. A., Ingersoll A. P., Ewald S. P., Porco C. C., Fischer G., Kurth W. S., West R. A., 2010, *GeoRL*, **37**, 9205
- Feuchtgruber H., et al., 2013, *A&A*, **551**, A126
- Fischer G., Gurnett D. A., Kurth W. S., Akalin F., Zarka P., Dyudina U. A., Farrell W. M., Kaiser M. L., 2008, *SSRv*, **137**, 271
- Fletcher L. N., Orton G. S., Teanby N. A., Irwin P. G. J., Bjoraker G. L., 2009, *Icarus*, **199**, 351
- Fletcher L. N., et al., 2011, *Science*, **332**, 1413
- Gaulme P., Schmider F.-X., Gay J., Guillot T., Jacob C., 2011, *A&A*, **531**, A104
- Gautier D., Hersant F., Mousis O., Lunine J. I., 2001, *ApJ*, **550**, L227
- Gulkis S., Janssen M. A., Olsen E. T., 1978, *Icarus*, **34**, 10
- Guillot T., Hueso R., 2006, *MNRAS*, **367**, L47
- Harrington J., de Pater I., Brecht S. H., Deming D., Meadows V., Zahnle K., Nicholson P. D., 2004, *jpsm.book*, 159
- Hedman M. M., Nicholson P. D., 2013, *AJ*, accepted
- Hedman M. M., Burns J. A., Evans M. W., Tiscareno M. S., Porco C. C., 2011, *Science*, **332**, 708
- Helled R., Anderson J. D., Podolak M., Schubert G., 2011, *ApJ*, **726**, 15
- Helled R., Guillot T., 2013, *ApJ*, **767**, 113
- Hersant F., Gautier D., Lunine J. I., 2004, *P&SS*, **52**, 623
- Hersant F., Gautier D., Tobie G., Lunine J. I., 2008, *P&SS*, **56**, 1103
- Ida S., Lin D. N. C., 2004, *ApJ*, **604**, 388
- Janhunen P., Quarta A. A., Mengali G., 2013, *Geoscientific Instrumentation*, **2**, 85
- Janhunen P., et al., 2010, *Review of Scientific Instruments*, **81**, 111301
- Jontof-Hutter D., Hamilton D. P., 2012a, *Icarus*, **218**, 420
- Jontof-Hutter D., Hamilton D. P., 2012b, *Icarus*, **220**, 487
- Kim Y. H., Fox J. L., 1994, *Icarus*, **112**, 310
- Krimigis S. M., et al., 2005, *Science*, **307**, 1270
- Lin D. N. C., Papaloizou J., 1986, *ApJ*, **309**, 846
- Marley M. S., Porco C. C., 1993, *Icarus*, **106**, 508
- Marley M. S., 2013, *Clouds and Hazes in Exoplanet Atmospheres*, arxiv preprint
- Marty B., et al., 2009, *Experimental Astronomy*, **23**, 947
- Marty B., Chaussidon M., Wiens R. C., Jurewicz A. J. G., Burnett D. S., 2011, *Science*, **332**, 1533
- Mizuno H., 1980, *PThPh*, **64**, 544
- Mordasini C., Alibert Y., Klahr H., Henning T., 2012, *A&A*, **547**, A111
- Moses J. I., Fouchet T., Bézard B., Gladstone G. R., Lellouch E., Feuchtgruber H., 2005, *JGRE*, **110**, 8001
- Mousis O., Marboeuf U., Lunine J. I., Alibert Y., Fletcher L. N., Orton G. S., Pauzat F., Ellinger Y., 2009, *ApJ*, **696**, 1348
- Mousis O., Lunine J. I., Madhusudhan N., Johnson T. V., 2012, *ApJ*, **751**, L7
- Nettelmann N., Helled R., Fortney J. J., Redmer R., 2013, *P&SS*, **77**, 143
- Niemann H. B., et al., 1998, *JGR*, **103**, 22831
- O'Donoghue J., Stallard T. S., Melin H., Jones G. H., Cowley S. W. H., Miller S., Baines K. H., Blake J. S. D., 2013, *Nature*, **496**, 193
- Orton G. S., et al., 1998, *JGR*, **103**, 22791
- Owen T., Mahaffy P., Niemann H. B., Atreya S., Donahue T., Bar-Nun A., de Pater I., 1999, *Nature*, **402**, 269
- Owen T., Mahaffy P. R., Niemann H. B., Atreya S., Wong M., 2001, *ApJ*, **553**, L77

- Pauzat F., Ellinger Y., Mousis O., Ali Dib M., Ozgurel O., Zicler E., 2013, *ApJ*, submitted
- Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, *Icarus*, **124**, 62
- Read P. L., Conrath B. J., Fletcher L. N., Gierasch P. J., Simon-Miller A. A., Zuchowski L. C., 2009, *PSS*, **57**, 1682
- Rinnert K., Lanzerotti L. J., Uman M. A., Dehmel G., Gliem F. O., Krider E. P., Bach J., 1998, *JGR*, **103**, 22979
- Russell C. T., Dougherty M. K., 2010, *SSRv*, 152, 251
- Sanchez-Lavega A., et al., 2008, *Nature*, **451**, 437
- Saumon D., Guillot T., 2004, *ApJ*, **609**, 1170
- Seiff A., et al., 1998, *JGR*, **103**, 22857
- Sentman D. D., 1990, *Icarus*, **88**, 73
- Simões F., et al., 2012, *ApJ*, **750**, 85
- Showalter M. R., Hedman M. M., Burns J. A., 2011, *Science*, **332**, 711
- Spilker, L. et al., 2013, White paper for NASA Decadal Survey 2013-2023
- Tiscareno M. S., Burns J. A., Hedman M. M., Porco C. C., Weiss J. W., Dones L., Richardson D. C., Murray C. D., 2006, *Nature*, **440**, 648
- Tiscareno M. S., et al., 2010, *Astrophys. J. Lett.* **718**, L92
- Tiscareno M. S., Hedman M. M., Burns J. A., Weiss J. W., Porco C. C., 2013, *Icarus*, **224**, 201
- von Zahn U., Hunten D. M., Lehmacher G., 1998, *JGR*, **103**, 22815
- Vorontsov S. V., Zharkov V. N., Lubimov V. M., 1976, *Icarus*, **27**, 109
- Ward W. R., 1997, *Icarus*, **126**, 261
- Watkins C., Cho J. Y.-K., 2013, *GeoRL*, **40**, 472
- West R. A., Baines K. H., Friedson A. J., Banfield D., Ragert B., Taylor F. W., 2004, *jpsm.book*, 79
- Wong M. H., Mahaffy P. R., Atreya S. K., Niemann H. B., Owen T. C., 2004, *Icarus*, **171**, 153
- Yair Y., Fischer G., Simões F., Renno N., Zarka P., 2008, *SSRv*, **137**, 29
- Yelle R. V., Miller S., 2004, *jpsm.book*, 185
- Yelle R. V., Lammer H., Ip W.-H., 2009, *coae.book*, 437



Neptune and Triton: Essential Pieces of the Solar System Puzzle

White paper submitted to ESA for the definition of the
L2 and L3 missions in the ESA science programme

Spokesperson: Adam Masters

Institute of Space and Astronautical Science
Japan Aerospace Exploration Agency
3-1-1 Yoshinodai, Chuo-ku, Sagami-hara
Kanagawa 252-5210, Japan
+81-(0)50-3362-6079
a.masters@stp.isas.jaxa.jp

Montage of Voyager 2 images showing Neptune as seen
from a spacecraft flying by its largest moon, Triton.
Credit: NASA/JPL/USGS

Authors

Adam Masters, JAXA/ISAS, Japan
Stefano Campagnola, JAXA/ISAS, Japan
Andrew Coates, University College London, UK
Javier Ruiz, Universidad Complutense de Madrid, Spain
Sébastien Charnoz, Université Paris Diderot, France
Leigh Fletcher, University of Oxford, UK

Francesco Marzari, University of Padova, Italy
Bruno Christophe, ONERA, France
Craig Agnor, Queen Mary, University of London, UK
Nadine Nettelmann, Universität Rostock, Germany
Nicholas Achilleos, University College London, UK
Laurent Lamy, Université Paris Diderot, France

Contributors

Jean-Pierre Lebreton, Université d'Orléans, France
Nick Sergis, Academy of Athens, Greece
Michele Dougherty, Imperial College London, UK
Fran Bagenal, University of Colorado, USA
Georg Moragas-Klostermeyer, Universität Stuttgart
Candice Hansen, NASA/JPL, USA
Ravit Helled, Tel Aviv University, Israel
Tom Nordheim, University College London, UK
Ralf Srama, Universität Stuttgart, Germany
Anil Bhardwaj, Vikram Sarabhai Space Centre, India

Emil Khalisi, Max-Planck-Institute, Germany
Richard Ambrosi, University of Leicester, UK
Abigail Rymer, Johns Hopkins University, USA
Gabor Facsko, Finnish Meteorological Institute, Finland
Martin Volwerk, Austrian Academy of Sciences, Austria
Licia Ray, University College London, UK
Glenn Orton, NASA/JPL, USA
Nicolas André, IRAP, France
Japheth Yates, University College London, UK
Christopher Russell, UCLA, USA

Supporters

Rumi Nakamura, Austrian Academy of Sciences, Austria
Bertrand Bonfond, Université de Liège, Belgium
Katerina Radioti, Université de Liège, Belgium
Denis Grodent, Université de Liège, Belgium
Sebastien Rodriguez, Université Paris Diderot, France
Emmanuel Lellouch, Université Paris Diderot, France
Gabriel Tobie, Université de Nantes, France
Sebastian Hess, Université Paris Diderot, France
Patrick Canu, Observatoire de Saint-Maur, France
Philippe Zarka, Université Paris Diderot, France
Baptiste Cecconi, Université Paris Diderot, France
Mathieu Barthelemy, IPAG, France
Thibault Cavalié, Université Bordeaux, France
Patricia Schippers, Université Paris Diderot, France
Pierre Touboul, ONERA, France
Alessandro Morbidelli, Observatoire de la Cote d'Azur
Michel Dobrijevic, Université de Bordeaux, France
Karl-Heinz Glaßmeier, Max-Planck-Institute, Germany
Martin Paetzold, Universität zu Köln, Germany
Silvia Tellmann, Universität zu Köln, Germany
Tom Andert, UniBw-Munich, Germany
Joachim Saur, Universität zu Köln, Germany
Anna Milillo, INAF, Italy
Satoshi Kasahara, JAXA/ISAS, Japan
Masaki Fujimoto, JAXA/ISAS, Japan
Tomoki Kimura, JAXA/ISAS, Japan
Shingo Kameda, Rikkyo University, Japan

Stas Barabash, Swedish Institute of Space Physics
Christopher Arridge, University College London, UK
Marina Galand, Imperial College London, UK
Nick Teanby, University of Bristol, UK
Jamie Jasinski, University College London, UK
Henrik Melin, University of Leicester, UK
Geraint Jones, University College London, UK
Patrick Guio, University College London, UK
Stan Cowley, University of Leicester, UK
Sarah Badman, University of Leicester, UK
Pat Irwin, University of Oxford, UK
Linda Spilker, NASA/JPL, USA
Kurt Retherford, SwRI, USA
Thomas Greathouse, SwRI, USA
Michael Davis, SwRI, USA
Heidi Hammel, Space Science Institute, USA
Neel Savani, Naval Research Laboratory, USA
Jim Slavin, University of Michigan, USA
Elizabeth Turtle, Johns Hopkins University, USA
Don Banfield, Cornell University, USA
Conor Nixon, NASA/GSFC, USA
Xianzhe Jia, University of Michigan, USA
David Brain, University of Colorado, USA
Jim Raines, University of Michigan, USA
Dan Gerschman, University of Michigan, USA
Todd Smith, Johns Hopkins University, USA

Executive Summary

Following the focus on addressing fundamental scientific questions in the Jupiter system, the Neptune system is a strategic and high-priority target for the European planetary science community. Neptune played a unique and important role in the process of Solar System formation. Within our Solar System Neptune is most similar to the dominant class of (Neptune-sized) detected exoplanets, and Neptune's largest moon Triton is very likely a captured Kuiper Belt Object. Neptune and Triton hold the keys to paradigm-changing advances in multiple fields of planetary science: Solar System and planetary formation, exoplanetary systems, geology and geophysics, atmospheric science, magnetospheric physics, and astrobiology. An L-class Neptune orbiter mission with multiple Triton flybys would lead to major scientific advances, addressing more than half of the ESA Cosmic Vision scientific questions in a single mission.

Neptune's internal heat flux significantly exceeds solar input, producing the most meteorologically active atmosphere in the Solar System despite its great distance from the Sun. This is in stark contrast with Uranus, which has a very low internal heat flux likely due to collisional processes. Sub-Neptune-sized planets are the most numerous class of detected exoplanet, with Neptune the Solar System planet most analogous to the majority of exoplanets. Key science questions include:

- How did the Neptune system form, what role did Neptune play in the dynamics of the early Solar System, and what does this outermost planet tell us about the numerous Neptune-sized exoplanets?
- What are the relative element abundances in Neptune's interior? Is the deep interior convective?
- What powers the most meteorologically active planetary atmosphere in our Solar System?
- How can we explain the unique composition and dynamics of Neptune's rings and small moons?

Neptune's largest moon Triton is very likely a captured Kuiper Belt object, holding the answers to questions about the icy dwarf planets that formed in the outer Solar System. Furthermore, Triton is geologically active, has a tenuous nitrogen atmosphere, and is predicted to have a subsurface ocean, making it a potential habitat of considerable scientific interest. Key science questions include:

- Does Triton retain any physical memory of its origins as an icy dwarf planet?
- What is the state of Triton's interior and surface, and what is the extent of geological activity?
- What is the detailed composition of Triton's atmosphere and what drives the plumes?
- How are Triton's surface, Triton's atmosphere, and Neptune's magnetosphere coupled?
- Does Triton have a subsurface ocean? If so, what are its characteristics and is it a potential habitat?

A Neptune orbiter mission with multiple Triton flybys would address this range of major scientific questions. Enabling technologies are: extended deep space network capability, radioisotope thermoelectric power generators, and solar electric propulsion. Preliminary mission analysis suggests that the interplanetary transfer time to Neptune is 15 years (using an Ariane 5 ECA launcher, and with a Jupiter gravity assist). We present an example 2-year Neptune tour with 55 Triton flybys.

Relevant ESA Cosmic Vision scientific questions:

1. What are the conditions for planet formation and the emergence of life?
 - 1.1 From gas and dust to stars and planets
 - 1.2 From exoplanets to biomarkers
 - 1.3 Life and habitability in the Solar System
2. How does the Solar System work?
 - 2.1 From the Sun to the edge of the Solar System
 - 2.2 The giant planets and their environments
 - 2.3 Asteroids and other small bodies
3. What are the fundamental physical laws of the universe?
 - 3.1 Explore the limits of contemporary physics

1. Introduction

Neptune is classified as one of the gas giant planets, along with Jupiter, Saturn, and Uranus, and is also often grouped with Uranus and referred to as an “ice giant”, because these two planets are both primarily composed of “ices” (volatile elements heavier than hydrogen and helium). However, there are fundamental and important differences between the Uranus and Neptune planetary systems, which their common classification as ice giant planets should not obscure. It is the aim of this white paper to highlight the tremendous importance of the Neptune system for European planetary science, and to outline the wide range of high-priority scientific questions across multiple fields that this planetary system can provide answers to.

Neptune orbits the Sun at a distance ~ 30 times greater than the mean Sun-Earth distance (an Astronomical Unit, AU). The planetary obliquity of $\sim 30^\circ$ leads to seasons over Neptune's ~ 165 -year orbit. A Neptune day is just over 16 hours long, and the planet is surrounded by a system of rings and icy moons (6 regular, 7 irregular). Triton, by far the largest moon, very likely formed as a dwarf planet in the Kuiper belt (like Pluto) before being captured by Neptune. This makes Triton a unique planetary satellite in the Solar System.

Voyager 2 is the only spacecraft that has encountered Neptune to date, flying by the planet on 25 August 1989 when it was summer in Neptune's southern hemisphere. Figure 1 shows *Voyager 2* imaging of Neptune during approach to the planet. The combination of this brief encounter and telescope observing campaigns have shown us that Neptune has the most meteorologically active atmosphere in the Solar System, despite its distance from the Sun, and that Triton has been (and could currently be) geologically active. Neptune is barely explored when compared to other planetary systems, and never with modern spacecraft instrumentation.

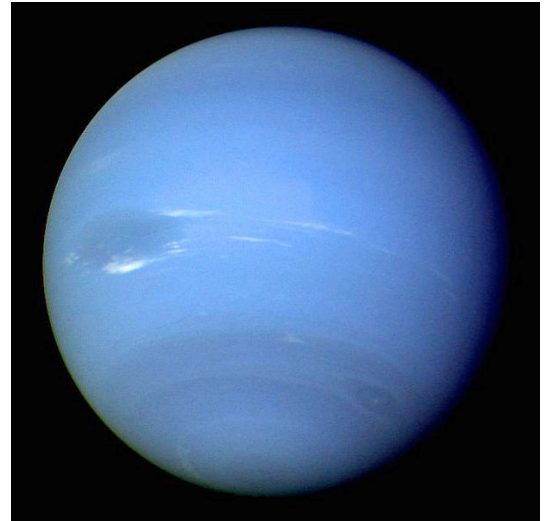


Figure 1. Neptune, captured by the *Voyager 2* narrow-angle camera. Credit: NASA/JPL.

A range of fundamental scientific questions concerning the Neptune planetary system are discussed in Section 2 of this white paper, with those questions concerning Triton outlined separately in Section 3. Because the theme of Neptune-Triton science is highly relevant for multiple scientific fields, both Sections 2 and 3 are divided into subsections on this basis. The important science that could be carried out by a spacecraft during an interplanetary transfer to Neptune is covered in Section 4. Finally, in Section 5 we discuss L-class Neptune orbiter mission concepts including multiple Triton flybys that would address all the major science questions. Identifying enabling technology is a priority.

2. Neptune

Relevant ESA Cosmic Vision scientific questions:

1. What are the conditions for planet formation and the emergence of life?
 - 1.1 From gas and dust to stars and planets
 - 1.2 From exoplanets to biomarkers
2. How does the Solar System work?
 - 2.1 From the Sun to the edge of the Solar System
 - 2.2 The giant planets and their environments
 - 2.3 Asteroids and other small bodies

2.1 Formation and Implications for the Solar System and Exoplanets

While there has been debate about Neptune's formation, a leading theory has now emerged from European scientists (e.g. the Nice model of Solar System dynamical evolution, developed at the Observatoire de la Côte d'Azur [Gomes et al., 2005; Tsiganis et al., 2005; Morbidelli et al., 2005]). It is postulated that Neptune formed at around 12-15 AU via planetesimal accumulation, before migrating to its present orbit at ~ 30 AU through a process of angular momentum exchange with a disk of planetesimals that initially extended out to

30-35 AU, and through interaction with the planets by gravitational scattering [Tsiganis et al., 2005]. This scenario is supported by the higher density of solid material closer to the Sun (typical of protoplanetary disks) that would have lead to a shorter planetary accretion time, and explains the dynamical structure of the Kuiper Belt (~30-50 AU, remnants of the planetesimal disk), the possible occurrence of the cataclysmic late heavy bombardment on the terrestrial planets, and the observed compositional diversity of the asteroid belt.

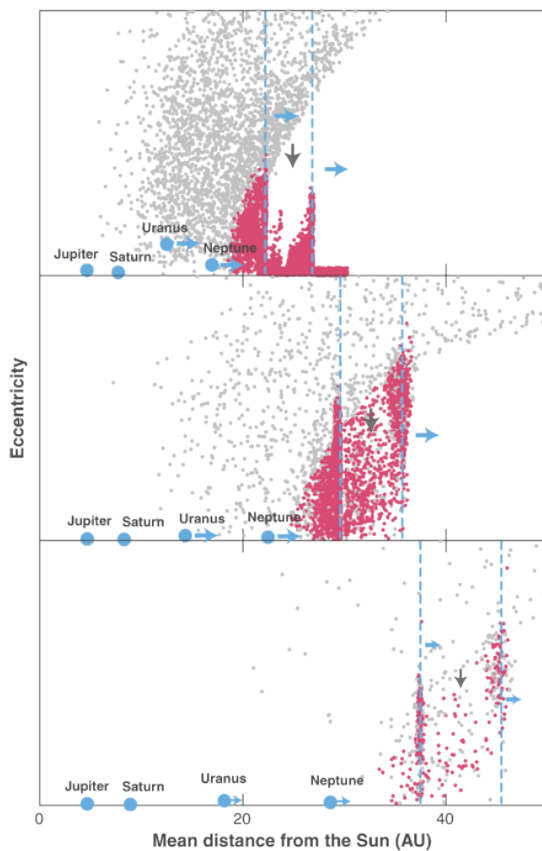


Figure 2. The orbital evolution of the outer Solar System. The three panels show sketches of the beginning, middle, and end of planetary migration. The disk planetesimals are coloured, depending on whether they have had close encounters with Neptune (grey) or not (red). From Morbidelli [2004].

by 3-31% of the Sun-like stars [Fressin et al., 2013]. Therefore Neptune may be typical of the majority of planets beyond our Solar System, unlike Uranus, which appears to have been radically altered by collisional processes. Neptune-sized exoplanets may share a similar evolution with Neptune, and a better knowledge of Neptune's physical properties will shed new light on the formation and characteristics of these exoplanets.

This leading theory highlights the importance of Neptune for Solar System formation and configuration, as illustrated in Figure 2. Neptune effectively pushed the outer boundaries of our Solar System [Morbidelli, 2004]. However, the process by which Neptune formed through accretion of planetesimals is poorly constrained. In addition, present understanding of the composition, configuration, and dynamics of the early Solar System is far from comprehensive, and our best models still cannot explain a number of features of the present day Solar System. Accurate knowledge of the physical properties of Neptune is of paramount importance for progress in all these areas. The size and mass of Neptune's core and its composition (rock/ice fraction) are crucial parameters for the problem of planetary formation, and for revealing the composition of the solar nebula. Knowledge of the properties and composition of interplanetary dust at Neptune's orbit (particularly originating from comets) would also lead to significant progress in this field.

One of the mysteries concerning Neptune's formation stems from the fact that it had to form after Jupiter and Saturn, since it did not accrete as much gas as these two other giant planets. Its core likely reached completion in the later stages of solar nebula evolution, when the gas density was low due to viscous accretion and photoevaporation. How the growth and migration of Jupiter and Saturn delayed the accretion of Neptune's atmosphere is not completely clear [Jakubik et al., 2012]. In this context, a detailed knowledge of the chemistry and composition of Neptune's atmosphere is essential for understanding how and when the planet accreted it.

Focus on Neptune has intensified recently due to the discovery of numerous exoplanets with similar physical characteristics, like Gliese 436 b or GJ 3470 b. In fact, Neptune-sized and sub-Neptune-sized planets are harboured

Key scientific questions: How and where did Neptune form? What role did Neptune play in early Solar System dynamics? What does Neptune tell us about the numerous exoplanets of similar mass?

2.2. Interior

Compared to other planets in the Solar System the composition and structure of Neptune's interior is very poorly constrained. The limited current understanding of the Neptunian interior and the high level of uncertainty is illustrated in Figure 3. Progress in this area is essential for answering one of the major mysteries concerning Neptune: Why is heat flux from the interior so high? Multiple theories have been proposed to answer this fundamental question, but without an improved knowledge of the planetary interior we cannot further constrain them. Note that Neptune's internal heat flux is higher than Uranus', representing an

important difference between the two ice giant planets. Uranus' low internal heat flux is thought to be the result of collisional processes.

The *Voyager 2* encounter with Neptune provides us with some preliminary constraint on Neptune's internal density distribution, rotation rate, and planetary radius. Before the encounter Neptune's interior was thought to be layered in the form of a rocky core, surrounded by an ice shell and a hydrogen/helium envelope. *Voyager* data indicated a light-element component in Neptune's deep interior, and a transition from a hydrogen/helium-rich to an icy/rock-rich interior at about 60-80% of the planetary radius [Hubbard et al., 1995].

Models constrain the light-element mass fraction in Neptune's deep interior to be 0-30% [Nettelmann et al. 2013], but this range allows for a variety of fundamentally different scenarios. For instance, a low light-element mass fraction of up to 10% could be explained by excess hydrogen originating from an initial water, ammonia and methane-rich composition, which was dissociated under high pressures and underwent phase separation into a hydrogen-oxygen phase and a carbon-nitrogen phase. The latter phase may have produced a diamond core. In contrast, a high light-element abundance would indicate simultaneous accretion of small planetesimals and gas, as well as the presence of rock in the deep interior. The question of hydrogen abundance in Neptune's deep interior and where it metallises is central to understanding how the planet generates its magnetic field.

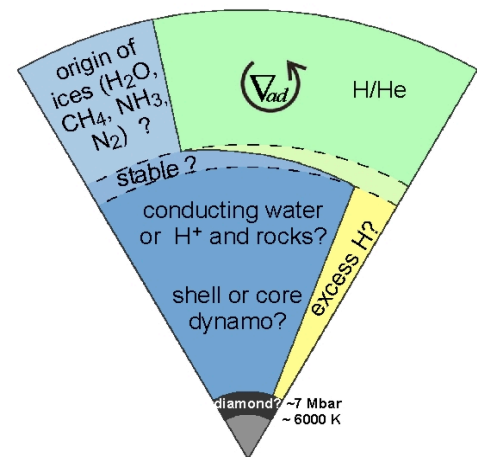


Figure 3. Chart showing Neptune's poorly understood interior.

Another fundamental question about the deep Neptunian interior is whether or not it is convective. All interior models agree in predicting a transition from a hydrogen/helium-rich outer region to a hydrogen/helium-poor inner region. The very deep interior has been suggested to be stably stratified in order to explain what we currently know about the planetary magnetic field thanks to *Voyager*. On the other hand, inefficient internal heat transport over a large fraction of Neptune's interior would yield lower than observed luminosities.

The last of the open issues concerning Neptune's interior that we would like to draw particular attention to is the fraction of heavy elements in the outer envelope, which is related to the fraction of heavy elements in the deep troposphere (the lower atmosphere). Models allow for solar enrichment of oxygen and carbon by a factor ranging between 100 and 400, in agreement with limited measurements of the tropospheric carbon monoxide abundance. In comparison, solar enrichment of heavy elements in the outer envelope of Uranus' interior is thought to be by a factor of 20 or less. It is unclear why the ice giant planets differ in this respect.

Key scientific questions: Why is the heat flux from Neptune's interior so high? What is the size of Neptune's core? What is the rotation rate of Neptune's interior? What is the composition and structure of both the deep interior and outer envelope? Is the deep interior convective?

2.3. Atmosphere

Neptune stands apart from the other giant planets by possessing the most meteorologically active atmosphere in our Solar System, with the fastest winds measured in any planetary atmosphere [e.g. Hammel et al., 1989]. Evidence for Neptune's rapidly evolving atmosphere provided by *Voyager 2* is shown in Figure 4. The drivers of this high level of atmospheric dynamics, given Neptune's distance from the Sun, is a mystery, and is closely related to the puzzle of why the heat flux from Neptune's interior is so high (see Section 2.2). Neptune's level of internal heat flux produces emissions that exceed solar input by a factor of 2.6, the largest of any planet in the Solar System [Pearl & Conrath, 1991]. Comparing Neptune to Uranus once again, Uranus is at the other end of the spectrum, with a ratio of only 1.1.

While Neptune's troposphere is likely driven by this excess of internal heat, it is unknown exactly what processes are involved, and how important surface effects are [Kaspi et al., 2013]. In addition, the power available to drive Neptune's incredibly strong wind systems is 20 times less than that at Jupiter. Some of the

basic dynamical, chemical, and cloud-forming processes at work within Neptune's churning atmosphere, along with the competing influences of seasonally changing insolation and internal heat flux on the atmospheric structure, are unknown.

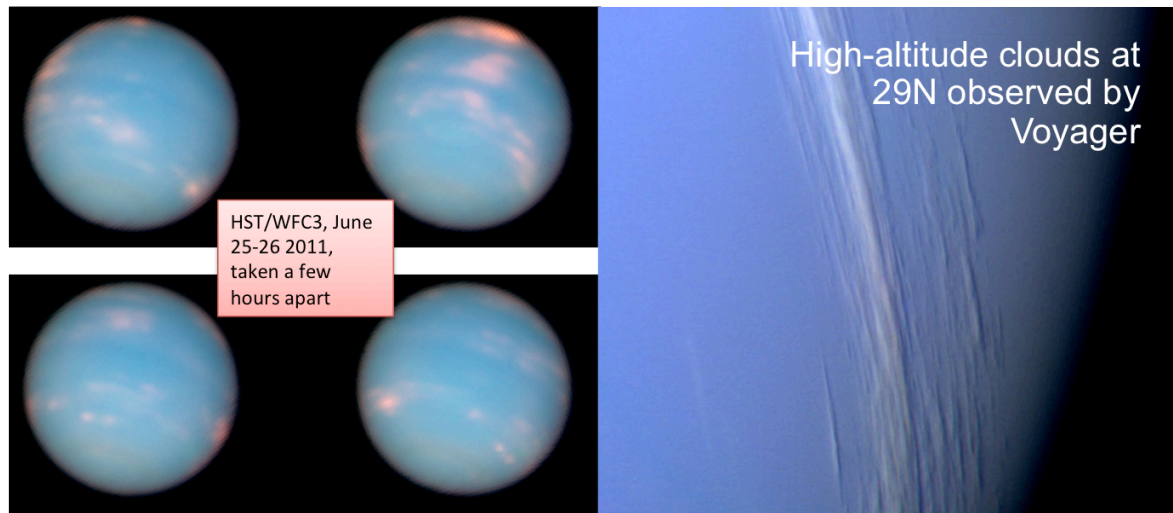


Figure 4. Hubble Space Telescope images of Neptune's rapidly evolving cloud systems, taken just a few hours apart (left). High-altitude clouds seen by Voyager 2 (right), credit: NASA/JPL.

Our present understanding is that Neptune's atmospheric composition is controlled by condensation chemistry, vertical mixing, external influx of oxygenated species from infalling comets and dust, and a rich hydrocarbon photochemistry due to the ultraviolet destruction of methane. Compared to Jupiter and Saturn, Neptune appears to have a different relation between banded cloud structures, atmospheric temperatures, and zonal wind structure. Infalling dust particles (e.g. interplanetary dust) are expected to pollute Neptune's atmosphere.

Rapidly evolving convective cloud activity prevails at cooler mid-latitudes (see Figure 4), with retrograde flow at the warmer equator and a high-latitude prograde jet confining a seasonally variable polar vortex of unusually high temperatures and unique chemical composition. Dark ovals (such as the Great Dark Spot observed by *Voyager 2*) are enormous storm systems, and are sometimes associated with bright white orographic clouds at higher altitudes. Neptune's strong zonal winds suggest a possibly low level of atmospheric turbulence that leads to energy dissipation. A reduced level of atmospheric turbulence would be in stark contrast to the fully turbulent atmospheres of Jupiter and Saturn.

The three-dimensional composition and structure of the atmosphere is unclear, and is central to understanding atmospheric chemistry [e.g. Lellouch et al., 1994] and cloud formation. Profiles of temperature, density, gaseous composition and aerosols would hold the key to understanding the balance between internal heating, convective mixing, latent heat release, and radiative heating and cooling throughout Neptune's atmosphere. In addition, the poles are important regions for the Neptune atmosphere problem, as they are the sites of coupling to the planet's magnetosphere (see Section 2.5). As is the case with other giant planets in our Solar System, the high temperature of Neptune's upper atmosphere remains to be explained.

Key scientific questions: What drives dynamics in the most meteorologically active atmosphere in our Solar System? What is the composition and structure of Neptune's atmosphere? What powers storm systems in Neptune's atmosphere? What is the nature of atmospheric chemistry and cloud formation?

2.4. Rings and Small Icy Satellites

Although all giant planets shelter a ring system, Neptune's ring system is unique because it consists of a collection of concentric and semi-transparent ringlets embedded in a tenuous sheet of dust. The Neptunian rings are tightly gravitationally coupled to a rich system of moonlets. Between the ringlets orbit a number of small moons (Naiad, Thalassa, Despina, Galatea). Both the rings and moons are especially dark, and the coupling between them is likely to be of key importance. The rings contain up to 70% dust in some regions [Smith et al., 1989], which makes them fundamentally different from Saturn's rings, which contain less than

1% dust. The origin of this difference in composition is still a mystery, and could be the signature of different formation/evolutionary processes.

High-resolution imaging carried out by *Voyager 2* suggests that some rings have sharp edges despite viscous spreading, suggesting gravitational confinement effects. Other rings appear to be broken into arc-like structures, as shown in Figure 5, which are somehow able to survive despite tidal forces and collisions between ring particles. The confinement effect of one or several nearby moons has been invoked to explain this. Earth-based observations have revealed the dynamical nature of the rings, and showed in 1999 that some arcs had shifted significantly from their expected location [Sicardy et al., 1999], while others seem to have fluctuated strongly in brightness since the *Voyager* era. Although the Jovian and Saturnian systems have moon-driven, extended, diffuse ring systems, currently no data exists about the Neptunian environment [Krivov et al., 2002; Srama et al., 2006].

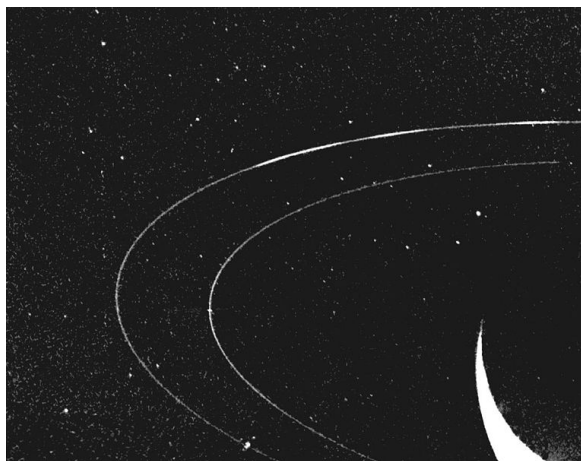


Figure 5. Image of the Adams and Leverrier ring (outer and inner curve, respectively) taken by the Voyager 2 wide-angle camera. The brightest parts of the Adams ring are the ring arcs. Credit: NASA/JPL.

The driver(s) of ring dynamics are unclear, and widely debated. It is thought that Neptune's rings evolve under the coupled action of sunlight, gravity, and collisional processes, but why their evolution is so different from other planetary ring systems is unknown. One of the most exciting perspectives about their origin is that they could be the result of disrupted satellites, either by tides [Leinhardt et al., 2012] or by cometary impacts [Colwell & Esposito, 1990]. A re-accretion process might currently be operating.

Neptune has 6 regular moons orbiting within 5 planetary radii, forming a compact system reminiscent of Saturn's mid-sized moons. A good fraction of them seem to orbit inside Neptune's Roche limit for ice, which implies that the small moons may be denser than ice [Tiscareno et al., 2013]. Tidal disruption of the weakest moons could give birth to narrow rings [Leinhardt et al., 2012]. Neptune's regular satellites are barely characterised, and their mass and densities are simply inferred from model-dependent arguments concerning the evolution of the rings. The surface of Proteus, the largest of Neptune's inner satellites,

appears to be densely cratered, and its non-hydrostatic shape may be the signature of past collisions, as illustrated by its large crater Pharos. The surfaces of the four innermost moons have never been imaged, representing a serious gap in our knowledge of the Neptune planetary system.

Satellite surfaces are continuously exposed to the interplanetary and interstellar meteoroid background, and ejecta from moon surfaces generates surrounding dust clouds, potentially creating ring systems [Krivov et al., 2002], and it has been proposed that the rings might have played a role in building the satellites themselves [Crida & Charnoz, 2012]. What is clear about this barely understood inner region of the Neptune system is that answering the many open questions about either the rings or inner moons would have important implications for the other.

Key scientific questions: Why is the composition of Neptune's rings different to that of any other planetary ring system? How do the ring arcs survive? Does Neptune have extended, dusty rings like Jupiter and Saturn? How did Neptune's inner satellites form? How does the coupled ring-moon system work?

2.5. Magnetic Environment

Neptune's magnetic field has a complex geometry. There is a large angle of $\sim 47^\circ$ between the magnetic dipole and rotation axes of the planet, the field is significantly offset from the centre of the planet by ~ 0.5 Neptune radii (R_N), and non-dipolar components are significant [e.g. Connerney et al., 1991]. The single *Voyager 2* flyby provides us with only a basic understanding of the field structure. Combining this with the fundamental unknowns concerning the planetary interior (see Section 2.2) makes the origin of Neptune's unusual magnetic field unclear. Solving the problem of how Neptune generates its magnetic field is a major challenge for

dynamo theorists, with broad implications for the field of planetary magnetism [e.g. Stanley & Bloxham, 2004; Soderlund et al., 2013].

The nature of Neptune's magnetic field leads to a highly irregular magnetosphere surrounding the planet [Bagenal, 1992]. The competition between the pressure exerted by the flow of solar wind plasma from the Sun and the pressure exerted by Neptune's magnetic field produces a substantial magnetospheric cavity in the solar wind flow that envelopes most of the Neptunian satellites, including Triton. Neptune's large dipole tilt angle leads to dramatic diurnal changes in the magnetosphere, unlike any other magnetosphere explored in detail to date. The changing configuration of Neptune's magnetosphere is illustrated in Figure 6.

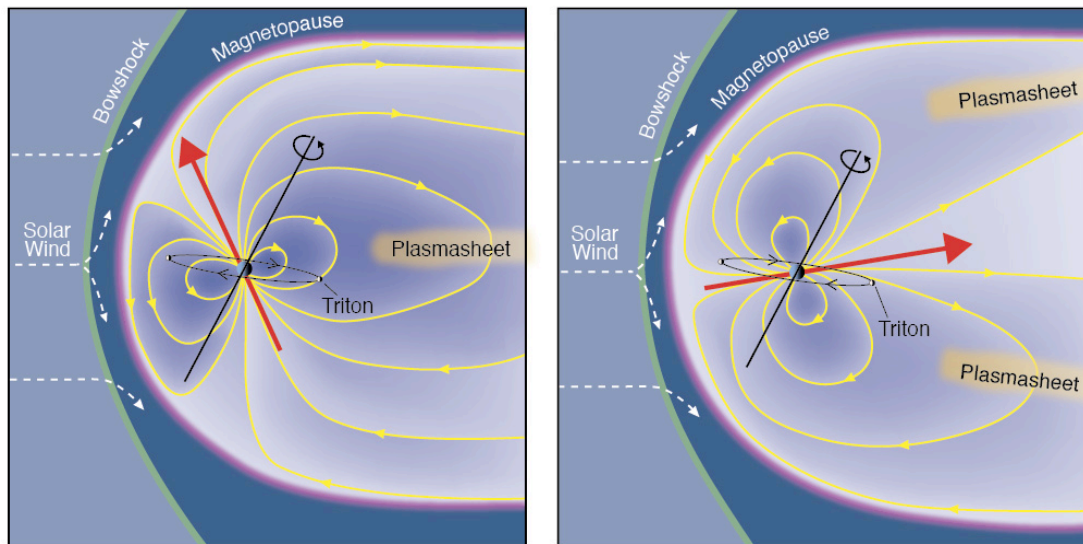


Figure 6. The changing configuration of Neptune's magnetosphere under solstice (southern summer) conditions. The noon-midnight plane is shown, with the planetary dipole (red arrow) captured at positions separated by half a planetary rotation period. Credit: Steve Bartlett & Fran Bagenal.

There are numerous important questions about how Neptune's magnetosphere works, which are also highly relevant for understanding the planetary atmosphere, rings, and satellites. Uncertainty surrounds the question of how the magnetosphere changes so dramatically in only half a planetary rotation period (~8 hrs), and what this means for the coupling between various parts of the system. This dynamic nature means that Neptune's magnetosphere may be our best Solar System laboratory for studying charge separation and equilibration due to variable magnetic fields, and the timescales associated with different particle acceleration mechanisms.

The relative importance of sources and sinks of plasma in Neptune's magnetosphere is also unknown. Triton is thought to be an important source [Richardson et al., 1991] (see Section 3), as well as dust particles. Triton makes the Neptunian magnetosphere a vital link between magnetospheres with similar internal sources of plasma but simpler internal fields (Jupiter and Saturn), and those with similar magnetic complexity but lacking these sources (Uranus). In particular, the presence/absence of a Triton plasma torus may explain the mysterious lack of a clear torus in Saturn's magnetosphere due to the moon Titan. Strong dust-plasma interactions may produce charged dust streams like those at Jupiter and Saturn [e.g. Kempf et al., 2005].

Auroral radio emission with distinct components has been unambiguously identified [e.g. Zarka et al., 1995] as well as emission at other wavelengths [Bhardwaj & Gladstone, 2000]. Otherwise, Neptune's auroral emissions are among the most mysterious in the Solar System, yet essential for understanding Neptune's magnetospheric system and the atmospheric energy budget. As the furthest planet from the Sun, how Neptune's dynamic magnetosphere interacts with the solar wind is of great interest [e.g. Schulz et al., 1995]. The planetary bow shock wave that stands upstream of the magnetosphere in the solar wind flow is expected to be the strongest (highest Mach number) in the heliosphere, and the magnetopause boundary of Neptune's magnetosphere is a unique laboratory in which to study fundamental processes (e.g. magnetic reconnection).

Key scientific questions: What is the origin and structure of Neptune's complex magnetic field? How does the magnetosphere re-configure on such short timescales? What are the sources and sinks of magnetospheric plasma? How does Neptune's aurora work? How does Neptune's magnetosphere interact with the solar wind?

3. Triton

Relevant ESA Cosmic Vision scientific questions:

1. What are the conditions for planet formation and the emergence of life?
 - 1.1 From gas and dust to stars and planets
 - 1.3 Life and habitability in the Solar System
2. How does the Solar System work?
 - 2.2 The giant planets and their environments
 - 2.3 Asteroids and other small bodies

3.1 Origin and Implications for the Neptune System

Triton, by far the largest of Neptune's moons, dominates Neptune's satellite system, and is an object of tremendous scientific interest. Triton's inclined (157°) retrograde orbit strongly suggests that it was captured by Neptune at some point during its history, as illustrated in Figure 7 [Goldreich et al., 1989; McKinnon et al., 1995; Agnor & Hamilton, 2006]. Thus, Triton likely formed orbiting the Sun in a similar region as other icy dwarf planets and primitive bodies in the outer Solar System, such as Eris, Pluto, Makemake, Haumea, Sedna, Orcus, and Quaoar.

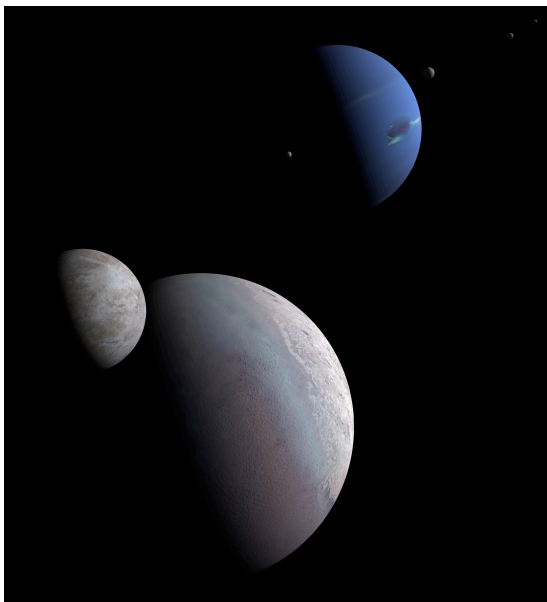


Figure 7. Triton and its binary companion as they approached Neptune. This encounter lead to Triton's capture by Neptune, an event that catastrophically altered the Neptune satellite system. In the image Neptune is orbited by several primordial satellites that may have existed prior to the encounter, but were destroyed in its aftermath.

This makes Triton the only large moon in the Solar System that did not form around its host planet. The physical characteristics (e.g. composition) of Triton hold the key to understanding the icy dwarf planets of the distant Kuiper Belt, an opportunity that no other planetary system can claim. Triton is subject to the tidal, radiolytic, and collisional environment of an icy satellite, but with the initial composition of a Kuiper Belt object.

Triton's capture must have left it on an orbit that was much larger (orbital radius: $\sim 80\text{--}1,000 R_N$) and more eccentric (eccentricity: $>\sim 0.95$) than its current one (orbital radius: $14 R_N$, eccentricity: 0). Triton's post-capture evolution likely dominated the subsequent evolution of the Neptunian system, and subjected the planetary satellite system to extreme processing via catastrophically disruptive collisions, gravitational scattering and tidal heating.

Driven to crossing orbits by Triton's perturbations, Neptune's inner satellites would collide at such large velocities that they would suffer catastrophic disruption and grind each other down into a debris disk [Goldreich et al., 1989]. In this view, Neptune's inner satellites are either the shards left over from this process or second-generation satellites that accreted from the rings and debris disk [Crida & Charnoz, 2012] (see Section 2.4). In either case, the inner satellite system has experienced extreme collisional processing. Neptune's distant irregular satellites were gravitationally sculpted by Triton following its capture with satellite material being exchanged

between the inner and outer regions through a variety of dynamical mechanisms.

Triton itself may have accumulated a significant portion of its mass ($\sim 20\%$) from the debris disk [Cuk & Gladman, 2005]. The accretion of this material would have hastened Triton's orbital decay, and would suggest that it may be a composite of heliocentric and planetocentric material. Triton's orbital decay was dominated by tidal friction, and the heating during this epoch is expected to be sufficient for global melting of Triton, and the formation of subsurface oceans [McKinnon et al., 1995].

Key scientific questions: Does Triton retain any physical memory of its origins as an icy dwarf planet? How did Triton evolve after it was captured, and how did Triton affect the Neptune planetary system? What are the similarities and differences between Triton and the dwarf planets of the Kuiper Belt?

3.2 Interior and Surface

The current state of our knowledge of Triton is based on very few observations (*Voyager 2*) and models. As a result, everything we think we know is subject to significant uncertainty, and there are fundamental questions that we have no answer to at present. What little we know implies that Triton is composed of a high proportion of rock and metal (~65-70%) compared to ice. Triton's orbital history and surface geology suggest an important role for tidal heating in the past [e.g., McKinnon et al., 1995] (see Section 3.1), which may have produced a differentiated interior with separation of ices, rocks, and metals. Triton could have a metallic core, silicate mantle, and internal liquid ocean between ice layers [Husmann et al., 2006; McKinnon & Kirk, 2007].

The brittle lithosphere (the outermost region of Triton's interior) is estimated to be ~10-15 km thick [Ruiz, 2003], which implies heat flows at the time when the surface was deformed that were clearly higher than those associated with the total radioactive heat production in the rocky portion of the satellite. Thus, observed resurfacing, geological activity, and the relatively thin lithosphere could have been caused by the heat generated during the capture of Triton, or by later release of the remaining heat.

Triton's surface is composed of ices, mostly N_2 (which includes CO , CH_4 , and C_2H_6 in solution), with seasonal polar deposits, plus H_2O , and CO_2 [Quirico et al., 1999]. Triton's surface has a young appearance, indicated by the sparseness and limited size of unambiguous impact craters, and by a variety of terrains unique to icy satellites. Crater counts indicate a surface age of several tens to hundreds of millions of years, but that in places the surface age could be as young as a few million years [Stern & McKinnon, 2000; Schenk & Zahnle, 2007]. Triton's surface is one of the younger surfaces in the Solar System, and shows that Triton is a geologically active satellite.

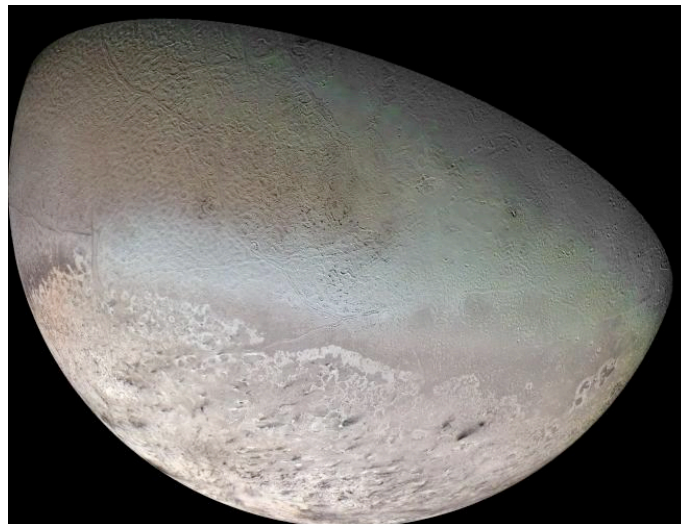


Figure 8. Global mosaic of Triton's surface. The southern polar cap covers the lower part of the imaged region of the surface. At lower latitudes the cantaloupe terrain and plains are in the West and the East, respectively. Credit: NASA/JPL/USGS.

On Triton's surface there are two major types of geological terrains [Smith et al., 1989; Croft et al., 1995], and a large polar cap of solid nitrogen ice covers a significant fraction of the southern hemisphere. Figure 8 shows *Voyager 2* imaging of the different terrain types. A substantial portion of the surface away from the polar cap that could be imaged by *Voyager 2* during its flyby appears to be occupied by expanses of regularly spaced, nearly circular depressions, dubbed cantaloupe terrains. The depressions are a few tens of kilometres wide and have a complex morphology. The other terrain type is undulating or smooth plains that show a variety of landforms, including depressions filled with smooth materials and "ice lakes". The surface is also deformed by a global network of ridges and troughs, more visible on the cantaloupe terrains and partly flooded at some locations on the plains [Croft et al., 1995]. The ridges morphologically resemble those seen at Jupiter's moon Europa [Prockter et al., 2005], although they are much less numerous.

Voyager 2 observed at least two plumes of nitrogen gas and dust at Triton's southern polar cap, which erupted from beneath the surface, extended up to 8 km above it, and were then dragged by atmospheric winds [Soderblom et al., 1990] (see Section 3.3). The polar dark streaks may be a result of such plume activity.

Key scientific questions: What are the composition, structure, and heat flow from Triton's interior? What is the age of features on Triton's surface? How geologically active is Triton and what drives the plumes?

3.3 Atmosphere

Triton's tenuous atmosphere was discovered by *Voyager 2*, although more distant remote sensing provided indirect evidence for an atmosphere before the flyby. We know only basic properties of the atmosphere, and how Triton's atmosphere interacts with both the surface of the moon below, and Neptune's magnetosphere above, remains unclear. Yet these properties are essential for understanding energy flow through the coupled planet-moon system.

Triton's atmosphere appears to be nitrogen-rich, and sustained by ices at the surface in vapour pressure equilibrium with the atmosphere. It has been likened to the atmosphere of Pluto. Currently known additional species in Triton's atmosphere are trace amounts of volatile gases, including methane and carbon monoxide. Trace amounts of CH₄, less than those in the atmospheres of Saturn's moon Titan or Pluto, were discovered

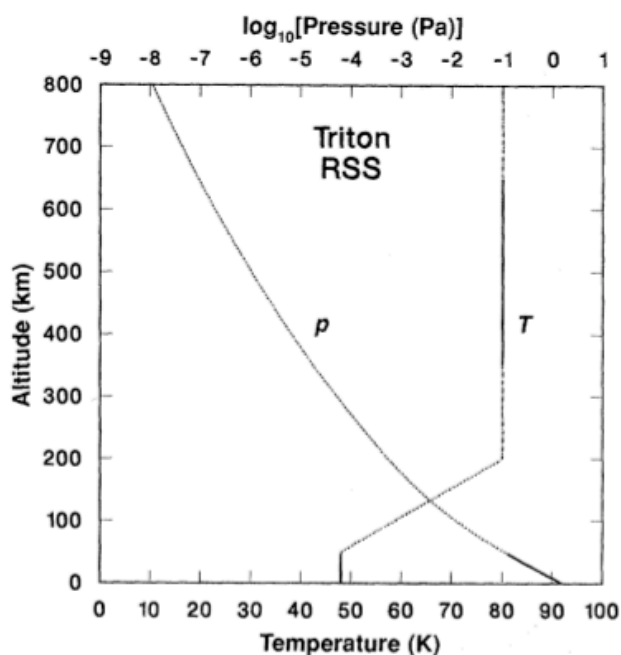


Figure 9. Profile of Triton's atmosphere based on radio data and models. From Tyler et al. [1989].

radiation and precipitation from Neptune's magnetosphere [Broadfoot et al., 1989]. A haze permeates most of Triton's troposphere, which may be largely composed of hydrocarbons and nitriles created by the action of sunlight on methane. The Triton atmosphere also appears to possess clouds of condensed nitrogen that lie between 1 and 3 km from the surface [Smith et al., 1989].

Key scientific questions: What molecular species are present in Triton's atmosphere? What is the distribution and source of aerosols in the atmosphere? How do winds affect the structure of Triton's atmosphere? What are the properties of the nitrogen plumes? What is the rate of dust infall to Triton's atmosphere?

3.4 Interaction with the Magnetosphere

Triton is thought to be the major source of plasma in Neptune's dynamic and irregular magnetosphere [Richardson et al., 1991] (see Section 2.5); however, the relative strength of Triton as a source compared to the solar wind and Neptune's ionosphere is unclear. Because of Triton's remarkable retrograde and highly inclined orbit, coupled with the dramatic diurnal reconfigurations of the planetary magnetosphere, the interaction between Triton and Neptune's magnetosphere is unique in the Solar System, and may be key to understanding the electrodynamics of moon-magnetosphere interactions in other planetary systems.

Triton has an ionosphere at the top of its tenuous atmosphere with a peak density at ~340km, as determined by radio science observations. One surprise revealed by these data was the observed high ionospheric density of

using ultraviolet observations made by *Voyager* [Broadfoot et al., 1989]. CO was first observed using the European Southern Observatory Very Large Telescope [Lellouch et al., 2010].

A profile of Triton's atmosphere is shown in Figure 9. Surface atmospheric pressure is thought to be ~1.4-1.9 Pa (14-19 μ bar) [Broadfoot et al., 1989, Tyler et al., 1989]. Pressure equilibrium in the nitrogen-rich atmosphere implies an upper limit for the surface temperature of Triton of ~38 K. Triton's atmosphere is seasonally variable, as the CH₄ abundance observed recently was several times that observed by *Voyager* [Lellouch et al., 2010].

Turbulence at Triton's surface creates a troposphere (lower level of the atmosphere) up to 8 km. Streaks on Triton's surface left by plumes (see Section 3.2) suggest that the troposphere is driven by seasonal winds capable of moving material over ~1 μ m in size [Smith et al., 1989]. Triton lacks a stratosphere, but has a thermosphere between ~8 and ~950 km, and an exosphere above. The temperature of the upper atmosphere is ~95 K, higher than that at the surface, which is thought to be due to heat absorbed from solar

$\sim 46,000 \text{ cm}^{-3}$ [Tyler et al., 1989]; this is higher than that in the ionosphere of Saturn's moon Titan, which also has a nitrogen-based atmosphere. This is surprising because the solar illumination is a factor of ~ 10 lower at Triton than at Titan. The high density has been suggested to be due to the impact of energetic ($>10 \text{ keV}$) precipitating particles from Neptune's magnetosphere [Strobel et al., 1990]. The measured energy flux of $>22 \text{ keV}$ particles well away from Triton is ~ 2 orders of magnitude greater than sunlight [Krimigis et al., 1989], but this will reduce significantly when Triton is far from the planetary magnetic equator.

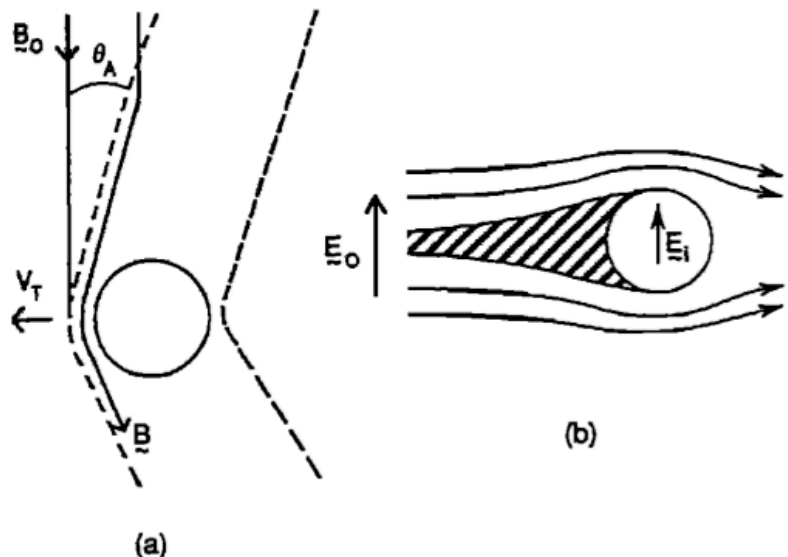


Figure 10. Triton's magnetospheric interaction, showing the expected Alfvén wings. From Strobel et al. [1990].

Due to the geometry and closest approach distance of the *Voyager 2* encounter with Triton, the moon-magnetosphere interaction has never been measured directly. Triton regularly visits different regions of Neptune's magnetosphere (magnetic L-shells between 14.3 and $>>40 R_N$ [Ness et al., 1989]) and is subject to different particle fluxes, and thus different coupling between the magnetosphere, atmosphere, and possibly Triton's surface. There is also a complex seasonal cycle, which must provide interesting and possibly significant effects.

Triton's orbital speed (4.4 km s^{-1}) and the expected local speed of magnetospheric plasma flow ($\sim 40 \text{ km s}^{-1}$) mean that Triton's interaction is likely to be transonic and sub-Alfvénic [Neubauer, 1990; Strobel et al., 1990]. These conditions are similar to those at Jupiter's moon Io. As a result, Alfvén wings are anticipated at Triton, as illustrated in Figure 10. Any intrinsic or induced (e.g. due to a subsurface ocean) magnetic fields at Triton would clearly affect this interaction with the magnetosphere.

Key scientific objectives: Why is Triton's ionosphere so dense, and what production and loss processes are involved? What is the nature of the Triton-magnetosphere interaction, and how does it respond to constantly changing external conditions? How important is Triton as a source of magnetospheric plasma? Does Triton have an internal magnetic field or aurorae? To what extent do energetic particles penetrate the atmosphere?

3.5 Habitability

Since the era of the *Voyager* planetary encounters subsurface oceans have been identified at three of Jupiter's moons (Europa, Ganymede, and Callisto), and there is indirect evidence for two of Saturn's moons (Enceladus and Titan) [e.g. Kivelson, 2004]. Subsurface oceans may be a common feature of icy moons in the Solar System, and a subsurface water ocean is predicted at Triton [McKinnon et al., 1995; Hussmann et al., 2006; McKinnon & Kirk, 2007]. Water is thought to be a key requirement for the habitability of such an ocean. *Cassini* observations at Saturn's moon Enceladus have demonstrated that dust in the surrounding environment can potentially reveal the composition of any subsurface ocean [Postberg et al., 2011].

As we have seen in Section 3.2, Triton has a young surface, with active cryovolcanism likely. This is evidence for the interplay between tidal dissipation, heat transfer, and tectonics which provides the energy for resurfacing of Jupiter's satellites Europa and Ganymede and at Saturn's satellite Enceladus. Such a source of energy is another expected requirement for the habitability of a subsurface ocean. Remaining expected habitability requirements are the right chemical environment, and time. Our limited knowledge of Triton's surface and atmospheric composition are the major constraint in our assessment of Triton as a potential habitat. Whether a subsurface ocean exists as predicted and whether there is any chemical evidence for this on the surface or in the atmosphere are major open questions concerning Triton, highly relevant for the field of astrobiology.

Key scientific questions: Does Triton have a subsurface ocean, and, if so, what are its properties and composition? Is the chemical environment favourable for habitability? How does Triton compare to other Solar System moons of astrobiological interest?

4. Science During an Interplanetary Transfer to the Neptune System

Relevant ESA Cosmic Vision scientific questions:

2. How does the Solar System work?

2.1 From the Sun to the edge of the Solar System

3. What are the fundamental physical laws of the Universe?

3.1 Explore the limits of contemporary physics

Answering the fundamental questions across multiple fields of planetary science that are outlined in Sections 2 and 3 require further spacecraft exploration of the Neptune system, as will be discussed in Section 5. Because of this, in this section we highlight the important additional scientific questions that could be addressed by a spacecraft on an interplanetary transfer to Neptune. This further extends the broad range of scientific fields covered by the theme of Neptune-Triton science.

Interplanetary and interstellar dust pervade the Solar System, but their distribution is not well known. Revealing the properties of this dust from 1 to 30 AU would have implications for Solar System formation and evolution (see Section 2.1), providing information about the Kuiper Belt. Interstellar dust grains are of particular interest as they are expected to preserve the conditions of star formation [Altobelli et al., 2003].

The continuous flow of solar wind plasma away from the Sun leads to significant energy flux through our entire Solar System, and this plasma flow eventually encounters its heliopause boundary. However, very few solar wind measurements have been made in the outer Solar System, beyond 10 AU. How solar wind structures (e.g. coronal mass ejections), evolve from the Sun to Neptune is therefore a largely open question in heliospheric physics. In addition, energetic neutral atoms have never been detected in the distant Solar System where Neptune resides, and would shed light on the global structure of the heliosphere itself.

Interplanetary space approaching Neptune is of great importance as an environment in which we can test the limits of contemporary physics. General Relativity (GR), the current theoretical formulation of gravitation, is in good agreement with most experimental tests [Will, 2006]. However, GR is a classical theory, and all attempts to merge it with the quantum description of the other fundamental interactions suggest it cannot be the final theory of gravitation. Meanwhile, experimental tests leave open windows for deviations from GR at small [Adelberger et al., 2009] and large distances [Reynaud & Jaekel, 2005].

GR is also challenged by observations at galactic and cosmic scales. The rotation curves of galaxies and the relation between redshifts and luminosities of supernovae deviate from the predictions of the theory. These anomalies are interpreted as revealing the presence of new components of the Universe, so-called “dark matter” and “dark energy” [Copeland et al., 2006; Frieman et al., 2008] which are thought to constitute respectively 25.8% and 69.4% of the energy content of the Universe according to most recent estimates [Ade et al., 2013]. The nature of both dark matter and energy remains unknown, and, despite their contribution to total energy content, they have not been detected up to now by means other than gravitational measurements.

A crucial question when addressing the nature of dark matter and dark energy is whether or not GR is the correct description of gravity at large scales, like distances approaching that between the Sun and Neptune. Addressing this question is essential in order to bridge the gap between experiments in the Solar System and astrophysical or cosmological observations. Probing the limits of current gravitation theory is also clearly related to the problem of Solar System formation and evolution, including the formation of the Neptune planetary system (see Section 2.1).

Key scientific questions: How do dust properties vary from Earth to Neptune? Do solar wind properties in the outer Solar System agree with model predictions? How do solar wind transients evolve from the Sun to ~30 AU, and what does this mean for Neptune’s magnetospheric dynamics? Is general relativity the correct description of gravity at scales approaching the Sun-Neptune distance? If not, how does this change our understanding of Solar System formation and evolution, and the dark matter/dark energy problem?

5. Neptune-Triton Mission Concepts

The driver of this white paper is the broad range of fundamental scientific questions that can be addressed in the Neptune planetary system (highly relevant for ESA's Cosmic Vision), which make Neptune and Triton strategic targets for planetary scientists. In this section we present a high-level discussion of the possibility of an ESA L-class mission to Neptune. Identifying enabling technology is a priority. The cost of an ESA Neptune mission would likely be similar to *JUICE*, within the L-class mission framework.

The tremendous scientific potential of Neptune and Triton has been known for many years, and has led to multiple mission concepts being studied in great detail. Heritage is provided by the most recent NASA/JPL mission concept study, which considered a range of mission architectures [Marley et al., 2010], and also from the Outer Solar System Mission that was submitted to ESA in response to the most recent call for M-class mission proposals [Christophe et al., 2012]. There is a great deal of scope for international collaboration, and also potential to use *JUICE* hardware in a spacecraft bound for Neptune.

5.1 Mission Architecture and Payload Options

To address all the diverse scientific questions outlined in Sections 2 to 4 a mission to Neptune requires a Neptune-orbiting spacecraft that makes multiple Triton flybys. If equipped with a payload of modern spacecraft instrumentation, such a spacecraft in orbit around Neptune would allow all questions to be addressed. Possible additional mission elements are a Neptune atmospheric descent probe, a Triton lander, and additional spacecraft; however, the inclusion of these enhancing elements is likely to be limited by mission cost and technical feasibility, and so they are not considered here. Such elements could potentially be considered in the framework of an international collaboration programme.

There are similarities between the Neptune orbiter discussed here and the *Galileo* and *Cassini-Huygens* missions to Jupiter and to Saturn and Titan, respectively. In both these cases the first spacecraft to orbit each planet lead to/continues to provide a hugely significant, paradigm-changing scientific return. A Neptune orbiter carrying a similar payload of scientific instruments would cover the wide range of Neptune-Triton science themes. Table 1 lists payload options. All modern spacecraft instrumentation included in Table 1 have a high Technology Readiness Level (TRL) and significant flight heritage. All values for instrument mass and power consumption are estimates. Specific measurement requirements for a Neptune orbiter mission are not discussed here, although the measurement ranges of heritage instruments would very likely be appropriate.

Instrument	Mass (kg)	Power (W)	Heritage
Narrow-angle camera (NAC)	9.8	14.0	Mars Express (SRC), New Horizons (LORRI), JUICE (JANUS)
Visible-infrared imager (VIR)	10.1	7.5	New Horizons (Ralph), Mars Express (OMEGA), Rosetta (VIRTIS), BepiColombo (SIMBIO-SYS)
Ultraviolet imaging spectrometer (UVIS)	5.0	12.0	BepiColombo (PHEBUS), Mars Express (SPICAM-UV), JUICE (UVS)
Accelerometer (ACC)	3.5	3.0	GOCE, GRACE, BepiColombo (ISA)
Radio science experiment (including ultrastable oscillator) (RSE)	3.5	45.5	Rosetta (RSI), New Horizons (REX), BepiColombo (MORE), JUICE (3GM)
Magnetometer (MAG)	3.3	3.0	Cassini (MAG), Double Star (MAG), Rosetta (RPC), BepiColombo (MERMAG), JUICE (J-MAG)
Thermal imager (TMI)	7.0	20.0	BepiColombo (MERTIS)
Particle package (plasma, neutrals, energetic neutral atoms) (PP)	23.0	50.0	Cassini (CAPS, MIMI), New Horizons (SWAP, PEPSSI), JUICE (PEP)
Radio and plasma wave system (RPWS)	5.7	7.1	Cassini (RPWS), JUICE (RPWI)
Dust Analyser (DA)	3.2	8.0	Cassini (CDA), Stardust (CIDA)

Table 1. Neptune orbiter payload options. All values of instrument mass and power consumption are estimates based on heritage instruments.

Science theme	NAC	VIR	UVIS	ACC	RSE	MAG	TMI	PP	RPWS	DA
Neptune interior (2.2)										
Neptune atmosphere (2.3)										
Neptune rings and icy satellites (2.4)										
Neptune magnetic environment (2.5)										
Triton interior and surface (3.2)										
Triton atmosphere (3.3)										
Triton-magnetosphere interaction (3.4)										
Cruise science (4)										

Table 2. Matrix relating science themes to payload options. Numbers given in brackets after each science theme indicate the relevant section/sub-section of this white paper.

Including all these instruments would result in a total payload mass of ~70 kg, compared to the ~60 kg value associated with a Neptune orbiter mission architecture previously studied by NASA [Marley et al., 2010]. Instrument development between now and the L2/L3 timeframe will likely reduce both mass and power consumption. Note that the instrument radiation shielding requirements are significantly lower for a Neptune orbiter compared to ESA's *JUICE* mission. We suggest that a programme of instrument studies within Europe's potential payload-providing countries be considered. Table 2 shows how each instrument included in Table 1 would address the various Neptune-Triton science themes discussed in this white paper (dependent on instrument specifications). Neptune formation (Section 2.1), Triton origin (Section 3.1), and Triton habitability (Section 3.5) science themes are not included in Table 2 because, while each is a distinct science theme, answering major science questions in these areas is dependent on answering those in other (included) areas.

5.2 Enabling Technology

Extended Deep Space Network (DSN) capability

Ka and X bands would be used for data and telemetry for a Neptune orbiter mission. The previous Neptune orbiter study by NASA [Marley et al., 2010] showed that a Ka-downlink to a single 34-m antenna yields 1-6 kbps at Neptune. A suggested solution to improve the data rate consisted of using four arrayed 34 m antennas. Although technology studies have been performed by ESOC, plans do not currently exist for multiple 35-m antennas in a single location of the European Tracking Network. However, plans exist within NASA's Deep Space Network (DSN). Use of the future DSN capability by ESA under a cooperation agreement would allow a data rate sufficient for a Neptune orbiter mission.

Radioisotope Thermoelectric Generators (RTGs) or Stirling Radioisotope Generators (SRGs)

The issue of electrical power for any mission beyond Jupiter makes RTGs or SRGs an enabling technology for a Neptune orbiter. European RTG development activities are currently targeting a maximum electrical power output of 50 W, with SRGs targeting 100 W. The European program to develop RTGs is currently at TRL ~3 [Ambrosi et al., 2013]. The radioisotope chosen for the European space nuclear power program is Americium-241 [Sarsfield et al., 2013], which has a longer half-life than the Plutonium-238 that has been used in almost all past RTGs employed in space. The current European RTG lifetime requirement is 20 years. Although there are differences between past and present RTG systems and the European units under development, we note that all past space RTGs have exceeded lifetime requirements (e.g. *Cassini-Huygens*). If we take the nominal power requirement of a Neptune orbiter mission to be 500 W, 10 European RTGs would be sufficient, producing a total electric power of 500 W for a total mass of ~250 kg. This assumes a nominal specific power of 2.0 W/kg, which is the current target of a study led by a UK team [Ambrosi et al., 2013]. Assuming a 20% maturity margin, the total mass would be ~300 kg. For comparison, a previous Neptune orbiter study by NASA [Marley et al., 2010] included 2 ASRGs (and 1 redundant) for a total power of ~280 W (100 kg). *JUICE* uses solar arrays to produce ~640 W at end of life (~350 kg for the entire subsystem) [Dougherty et al., 2011].

Solar Electric Propulsion (SEP)

An RTG lifetime comparable to interplanetary transfer time leads to a third enabling technology for a Neptune orbiter mission. Options to reduce the interplanetary transfer time are an SEP module, an Electric Sail (E-sail),

and aerocapture at Neptune Orbit Insertion (NOI). The option with the highest TRL is SEP, which would provide large Delta-V with small propellant mass in the earlier part of an interplanetary transfer to Neptune (see Section 5.3), before module ejection prior to NOI. An SEP module with four QinetiQ T6 Gridded Ion Engines (3 nominal and 1 redundant) would be sufficient, each providing 155 mN of thrust and requiring 5.5 kW. These high-TRL engines will fly on *Alphabus*, the new European GEO platform to be launched later this year, as well as on *BepiColombo*. The power for a Neptune mission EP system would be provided by solar arrays. With current technology, the specific power provided is 75 W/kg at 1 AU (*Dawn* for instance achieved 82 W/kg at 1 AU). The total 1 AU power output of the envisaged Neptune SEP module solar arrays would be similar to that of *Alphabus*. An estimate of the total mass of an SEP module for a Neptune orbiter is 1,500 kg, including solar arrays, tanks, structure, and 640 kg of propellant. A Neptune orbiter SEP module would not be subject to degradation at high temperatures, unlike the *BepiColombo* SEP module. Although not favoured here, we note that the lower TRL alternative options of an E-sail [Janhunen et al., 2013] and aerocapture at NOI have potential to become more attractive options in future.

5.3 Preliminary Mission Analysis

A detailed Neptune orbiter mission analysis is beyond the scope of this white paper. However, even the high-level discussion of mission concepts presented here requires a preliminary mission analysis to be carried out, for reasons explained below. This preliminary analysis is intended to serve as a starting point for future, detailed analysis of an ESA L-class mission to Neptune and Triton. Full details of the results of our preliminary mission analysis (not given here due to length constraints) are available on request.

Interplanetary transfer to Neptune

Issues such as RTG lifetime (20 years, including pre-launch ground phase) make the duration of an interplanetary transfer to Neptune an essential aspect of any discussion of Neptune orbiter mission concepts. We investigated trajectory options involving a launch from Kourou centred on the 2028-2034 timeframe. Rather than project future Ariane launcher performance, we assume an Ariane 5 ECA launcher for this preliminary analysis. Interplanetary transfer to Neptune requires a Gravity Assist (GA) by either Jupiter or Saturn a few years after launch because of RTG lifetime and to mitigate propellant requirements. However, a Jupiter GA is more effective than a Saturn GA for a Neptune orbiter mission [Landau et al., 2009].

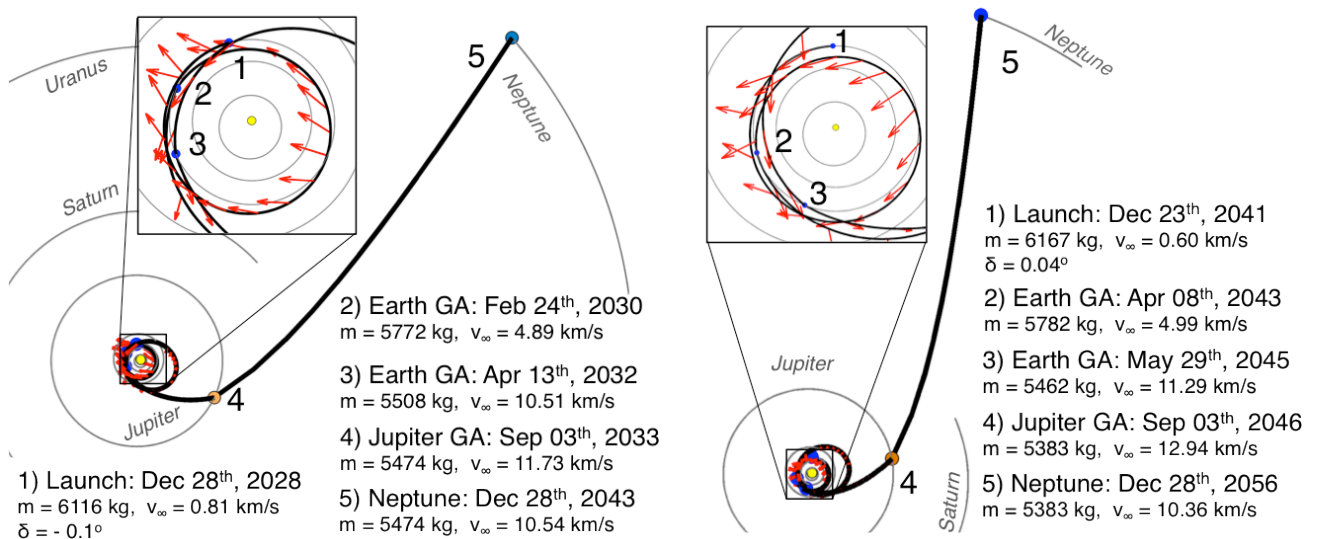


Figure 11. Example interplanetary transfer with launch in 2028 (left) and 2041 (right). Trajectory arcs where SEP is employed are modelled by small impulsive Delta-V (represented by red arrows).

Favourable opportunities for a Jupiter GA will exist in 2033 and in 2046 (separated by a Jupiter-Neptune synodic period of ~ 13 years). We thus studied interplanetary transfers that take advantage of each of these opportunities. One or more Earth GAs and orbital manoeuvres are required prior to the Jupiter GA in both cases, with mission-enabling SEP employed in this phase since chemical propulsion would require large amounts of fuel (>4 tons, neglecting use of low-TRL aerocapture for Neptune orbit insertion). Figure 11 shows an example interplanetary transfer for each Jupiter GA opportunity. Launch is in 2028 and Neptune arrival is in 2043 in the first example, and launch is in 2041 and Neptune arrival is in 2056 in the second example. The

transfer duration is ~ 15 years in both examples. These transfer options deliver $\sim 1,800$ kg dry mass into Neptune orbit, in line with estimates provided by the past Neptune orbiter study by NASA [Marley et al., 2010], and similar to the JUICE dry mass of 1,800 kg (including radiation shielding not required at Neptune).

Neptune orbital tour

Another essential aspect of a Neptune orbiter mission concepts discussion is the question of whether a spacecraft tour would allow the necessary observation opportunities. The frequency and geometry of Triton flybys is crucial. The key point we would like to highlight is that Triton is an effective “tour engine”, allowing a wide range of orbit trajectories and observation opportunities. We present one example Neptune tour here, which is essentially a proof of concept. Although not optimised, this tour would address all scientific questions. Our example tour is 2 years in duration, starting with interplanetary transfer arrival conditions given by the first stage of this preliminary analysis. At the beginning of the tour the spacecraft flies between the inner rings and executes NOI at 3,000 km altitude, following previous NASA mission concepts [Marley et al., 2010]. The tour is shown in Figure 12. In two years there are 55 Triton flybys, with groundtracks shown in Figure 13. Total chemical Delta-V for the whole mission is ~ 3 km/s, similar to *JUICE* (~ 2.6 km/s).

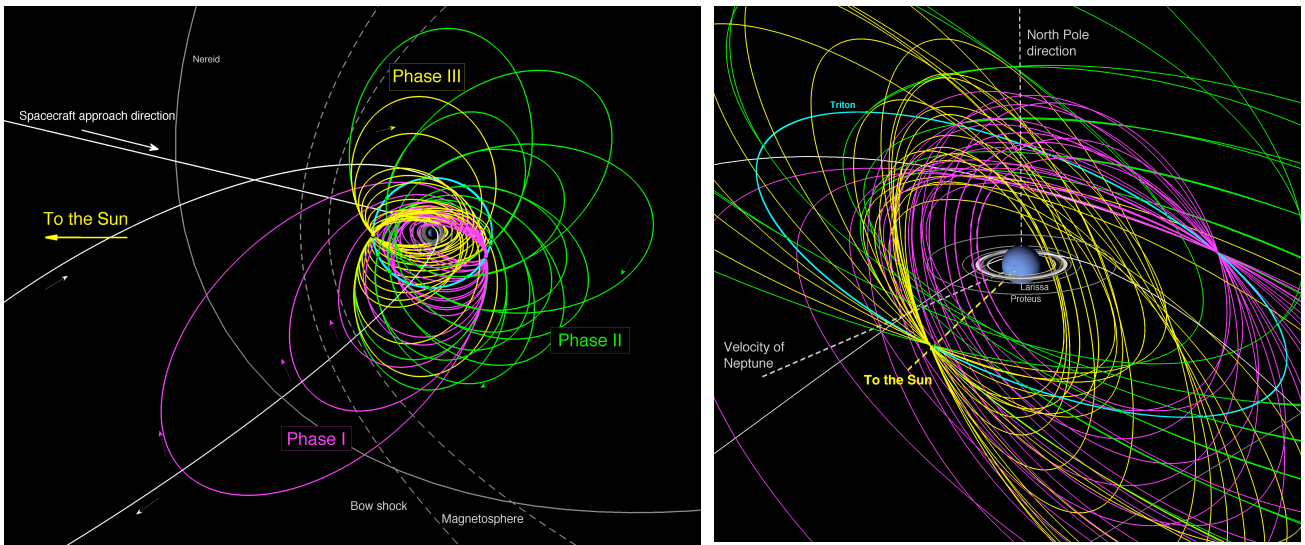


Figure 12. Example Neptune orbital tour. Left: Viewed from Neptune's north pole. Right: Close-up of the tour.

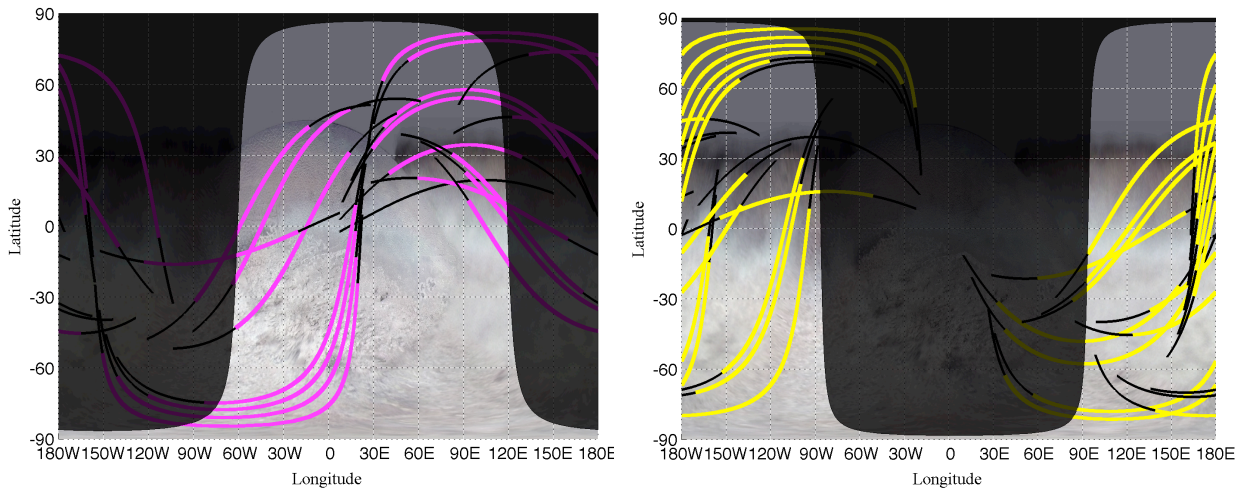


Figure 13. Groundtracks of Triton flybys during Phase I (left) and Phase III (right). Below 5000 km (black), below 1000 km (coloured).

During the three phases of this example tour there are inclined Neptune orbits and orbits in Triton's orbital plane. Triton flybys occur over the full range of Triton orbital locations, and at altitudes between ~ 150 and $\sim 1,000$ km. There is significant flexibility in, for example, Triton flyby altitudes, which can be raised or lowered as necessary. Our preliminary analysis suggests that a Triton orbit phase could be included at a Delta-V cost of ~ 300 m/s, using a transfer similar to that planned for *JUICE* [Campagnola et al., 2012]. Close flybys at Neptunian moons other than Triton are also possible.

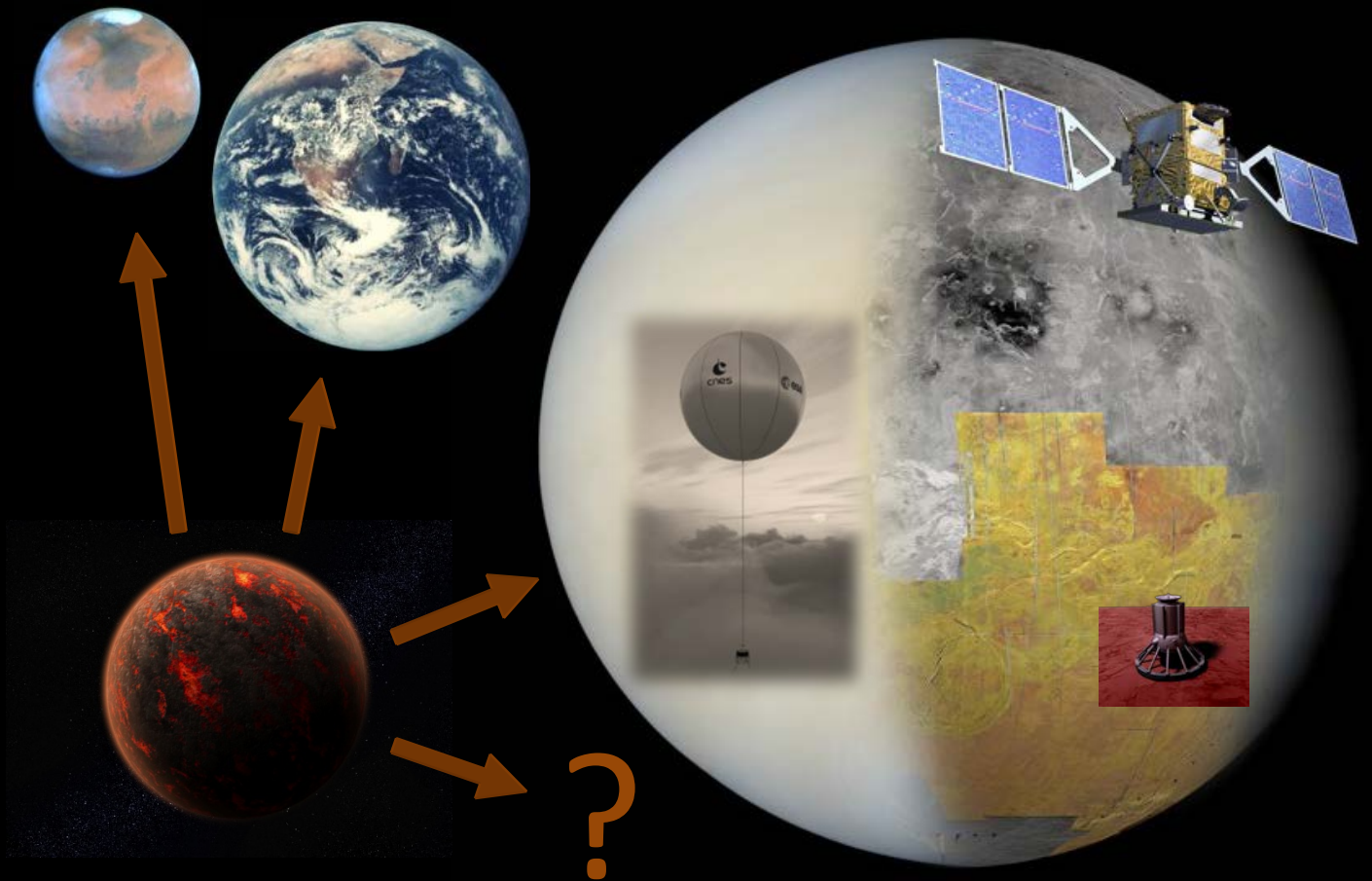
References

- Ade, P. A. R., et al. (2013), Planck 2013 results. XVI. Cosmological parameters, arXiv:1303.5076.
- Adelberger, E., et al. (2009), Torsion balance experiments: A low-energy frontier of particle physics, *Progress in Particle and Nuclear Physics*, 62, 102.
- Agnor, C. B., and D. P. Hamilton (2006), Neptune's capture of its moon Triton in a binary-planet gravitational encounter, *Nature*, 441, 192.
- Altobelli, N., et al. (2003), Cassini between Venus and Earth: Detection of interstellar dust, *J. Geophys. Res.*, 108, A10, 7-1.
- Ambrosi, R. M., et al. (2013), Development and testing of an Americium-241 radioisotope thermoelectric generator, *Proceedings of Nuclear and Emerging Technologies for Space*, Albuquerque, NM, February 25-28.
- Bagenal, F. (1992), Giant Planet Magnetospheres, *Ann. Rev. Earth Planet. Sci.*, 20, 289.
- Bhardwaj, A., and G. R. Gladstone (2000), Auroral emissions of the giant planets, *Rev. Geophys.*, 38, 295.
- Broadfoot, A. L., et al. (1989), Ultraviolet spectrometer observations of Neptune and Triton, *Science*, 246, 1459.
- Campagnola, S., et al. (2012), Tisserand-leveraging transfers, *Advances in the Astronautical Sciences*, 143, 1205.
- Christophe, B., et al. (2012), OSS (Outer Solar System): A fundamental and planetary physics mission to Neptune, Triton and the Kuiper Belt, *Exp. Astron.*, 34, 203.
- Colwell, J. E., and L. W. Esposito (1990), A numerical model of the Uranian dust rings, *Icarus*, 86, 530.
- Connerney, J. E. P., et al. (1991), The magnetic field of Neptune, *J. Geophys. Res.*, 96, 19023.
- Copeland, E. J., et al. (2006), Dynamics of dark energy, *Int. J. Mod. Phys.*, 15, 1753.
- Crida, A., and S. Charnoz (2012), Formation of regular satellites from ancient massive rings in the Solar System, *Science*, 338, 1196.
- Croft, S. K., et al. (1995), The geology of Triton, in: Cruikshank (Ed.), *Neptune and Triton*, Univ. of Arizona Press, Tucson, pp. 879.
- Cuk, M., and B. J. Gladman (2005), Constraints on the orbital evolution of Triton, *Ap. J.*, 626, L113.
- Dougherty, M. K., et al. (2011), JUICE, Exploring the emergence of habitable worlds around gas giants, *Assessment Study Report*, European Space Agency, ESA/SRE(2011)18.
- Fressin, F., et al. (2013), The false positive rate of Kepler and the occurrence of planets, *Ap. J.*, 766, 81.
- Frieman, J. A., et al. (2008), Dark energy and the accelerating universe, *Annu. Rev. Astron. Astrophys.*, 46, 385.
- Goldreich, P., et al. (1989), Neptune's story, *Science*, 245, 500.
- Gomes, R., et al. (2005), Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets, *Nature*, 435, 466.
- Hammel, H. B., et al. (1989), Neptune's wind speeds obtained by tracking clouds in Voyager images, *Science*, 245, 1367.
- Hubbard, W. B., et al. (1995), The interior of Neptune, In: Cruikshank (Ed.), *Neptune and Triton*. University of Arizona, Tucson, pp. 109.
- Hussmann, H., F., et al. (2006), Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-neptunian objects, *Icarus*, 185, 258.
- Jakubik, M., et al. (2012), The accretion of Uranus and Neptune by collisions among planetary embryos in the vicinity of Jupiter and Saturn, *A&A*, 540, 16.
- Janhunen, P., et al. (2013), Electric solar wind sail mass budget model, *Geosci. Instrum. Method. Data Syst.*, 2, 85.
- Kaspi, Y., et al. (2013), Atmospheric confinement of jet streams on Uranus and Neptune, *Nature*, 497, 344.
- Kempf, S., et al. (2005), High-velocity streams of dust originating from Saturn, *Nature*, 433, 289.
- Kivelson, M. G. (2004), Moon-magnetosphere interactions: a tutorial, *Adv. Space Res.*, 33, 2061.
- Krimigis, S. M., et al. (1989), Hot plasma and energetic particles in Neptune's magnetosphere, *Science*, 246, 1483.
- Krivov, A. V., et al. (2002), Dust on the outskirts of the Jovian system, *Icarus*, 157, 436.
- Landau, D. F., et al. (2009), Broad search and optimization of solar electric propulsion trajectories to Uranus and Neptune, *Advances in the Astronautical Sciences*, 153, 2093.
- Leinhardt, Z. M., et al. (2012), Tidal disruption of satellites and formation of narrow rings, *MNRAS*, 424, 1419.

- Lellouch, E., et al. (1994), The vertical Distribution and Origin of HCN in Neptune's Atmosphere, *Icarus*, 108, 112.
- Lellouch, E., et al. (2010), Detection of CO in Triton's atmosphere and the nature of surface-atmosphere interactions, *A&A*, 512, L8, doi: 10.1051/0004-6361/201014339.
- M. Marley, et al. (2010), Planetary Science Decadal Survey JPL Rapid Mission Architecture Neptune-Triton KBO Study Final Report, <http://solarsystem.nasa.gov/2013decadal/whitepapers.cfm?Category=MS>.
- McKinnon, W. B., et al. (1995), Origin and evolution of Triton, in: Cruikshank (Ed.), *Neptune and Triton*, Univ. of Arizona Press, Tucson, pp. 807.
- McKinnon, W. B., and R. L. Kirk (2007), Triton, in: L. A. McFadden, P. Weissman, T. Johnson (Eds.), *Encyclopedia of the Solar System*. Academic Press, pp. 483.
- Morbidelli, A. (2004), How Neptune pushed the outer boundaries of our Solar System, *Science*, 306, 1302.
- Morbidelli, A., et al. (2005), Chaotic capture of Jupiter's Trojan asteroids in the early Solar System, *Nature*, 435, 462.
- Ness, N. F., et al. (1989), Magnetic fields at Neptune, *Science*, 246, 1473.
- Nettelmann, N., et al. (2013), New indication for a dichotomy in the interior structure of Uranus and Neptune from the application of modified shape and rotation data, *Planet. Space Sci.*, 77, 143.
- Neubauer, F. M. (1990), Satellite plasma interactions, *Adv. Space Res.*, 10, 25.
- Pearl, J. C., and B. J. Conrath (1991), The albedo, effective temperature, and energy balance of Neptune, as determined from Voyager data, *J. Geophys. Res.*, 96, 18921.
- Postberg, F., et al. (2011), A salt-water reservoir as the source of a compositionally stratified plume on Enceladus, *Nature*, 474, 620.
- Prockter, L. M., et al. (2005), A shear heating origin for ridges on Triton, *Geophys. Res. Lett.*, 32, L14202, doi: 10.1029/2005GL022832.
- Quirico, E., et al. (1999), Composition, physical state, and distribution of ices at the surface of Triton, *Icarus* 139, 159.
- Reynaud, S., and M. T. Jaekel (2005), Testing the Newton law at long distances, *Int. J. Mod. Phys.*, 20, 2294.
- Richardson, J. D., et al. (1991), Low-energy ions near Neptune, *J. Geophys. Res.*, 96, 18993.
- Ruiz, J. (2003), Heat flow and depth to a possible internal ocean on Triton, *Icarus*, 166, 436.
- Sarsfield, M. J., et al. (2013), Progress on ²⁴¹Am production for use in Radioisotope Power Systems, *Proceedings of Nuclear and Emerging Technologies for Space*, Albuquerque, NM, February 25-28.
- Shenk, P.M., and K. Zahnle (2007), On the negligible surface age of Triton, *Icarus*, 192, 135.
- Schulz, M., et al. (1995), Magnetospheric Configuration of Neptune, in *Neptune and Triton*, ed. D. P. Cruikshank, M. S. Matthews, A. M. Schumann, University of Arizona Press.
- Sicardy, B., et al. (1999), Images of Neptune's ring arcs obtained by a ground-based telescope, *Nature*, 400, 731.
- Smith, B. A., et al. (1989), Voyager 2 at Neptune: Imaging Science Results, *Science*, 246, 1422.
- Soderblom, L. A., et al. (1990), Triton's geyser-like plumes: Discovery and basic characterization, *Science* 250, 410.
- Soderlund, K. M., et al. (2013), Turbulent models of ice giant internal dynamics: Dynamos, heat transfer, and zonal flows, *Icarus*, 224, 97.
- Srama, R., et al. (2006), In situ dust measurements in the inner Saturnian system, *Planet. Space Sci.*, 54, 967.
- Stanley, S., and J. Bloxham (2004), Convective-region geometry as the cause of Uranus' and Neptune's unusual magnetic fields, *Nature*, 428, 151.
- Stern, S. A., and W. B. McKinnon (2000), Triton's surface age and impactor population revisited in light of Kuiper belt fluxes: evidence for small Kuiper belt objects and recent geological activity, *Astron. J.*, 119, 945.
- Strobel, D. F., et al. (1990), Magnetospheric interaction with Triton's ionosphere, *Geophys. Res. Lett.*, 17, 1661.
- Tiscareno, M. S., et al. (2013), Compositions and origins of outer planet systems: insights from the Roche critical density, *Ap. J. Lett.*, 765, 5.
- Tsiganis, K., et al. (2005), Origin of the orbital architecture of the giant planets of the Solar System, *Nature*, 435, 459.
- Tyler, G. L., et al. (1989), Voyager Radio Science Observations of Neptune and Triton, *Science*, 246, 1466.
- Will, C. M. (2006), The confrontation between general relativity and experiment, *Living Rev. Relativity*, 9(3), 1.
- Zarka, P., et al. (1995), Radio emission from Neptune, in *Neptune and Triton*, ed. D. P. Cruikshank, M. S. Matthews, A. M. Schumann, University of Arizona Press.

VENUS: KEY TO UNDERSTANDING THE EVOLUTION OF TERRESTRIAL PLANETS

A RESPONSE TO ESA'S CALL FOR WHITE PAPERS FOR THE DEFINITION OF SCIENCE THEMES FOR L2/L3 MISSIONS IN THE ESA SCIENCE PROGRAMME.



Spokesperson: Colin Wilson

Atmospheric, Oceanic and Planetary Physics,
Clarendon Laboratory,
University of Oxford, UK.
E-mail: wilson@atm.ox.ac.uk

EXECUTIVE SUMMARY

In this White Paper, we advocate L2/L3 science themes of understanding the diversity and evolution of habitable planets, and emphasize the importance of Venus to these science themes.

Why are the terrestrial planets so different from each other? Venus should be the most Earth-like of all our planetary neighbours. Its size, bulk composition and distance from the Sun are very similar to those of the Earth. Its original atmosphere was probably similar to that of early Earth, with large atmospheric abundances of carbon dioxide and water. Furthermore, the young sun's fainter output may have permitted a liquid water ocean on the surface. While on Earth a moderate climate ensued, Venus experienced runaway greenhouse warming, which led to its current hostile climate. How and why did it all go wrong for Venus? What lessons can we learn about the life story of terrestrial planets/exoplanets in general, whether in our solar system or in others?

ESA's Venus Express mission has proved tremendously successful, answering many questions about Earth's sibling planet and establishing European leadership in Venus research. However, further understanding of Venus and its history requires several further lines of investigation. Entry into the atmosphere will be required in order to measure noble gas isotopic signatures of past history and to understand the role of clouds in the climate balance. Radar mapping at metre-scale spatial resolution, and surface height change detection at centimetre scale, would enable detection of current volcanic and Aeolian activity, and would revolutionise comparative geology between the terrestrial planets. The tessera highlands of Venus are thought to be the oldest terrain type found on Venus but have not yet been visited by spacecraft; a lander in these regions would measure surface composition to provide clues as to the earliest geologic record available on Venus.

Individually, these investigations could be carried out by separate low cost missions, and there is ample scope for international collaboration as individual payload or mission elements could be provided or even separately launched by different space agencies. However, there is strong synergy to be achieved by having all of these elements operating at the same time at Venus in close co-ordination. An orbiter permits much higher data return from balloons or landing probes by acting as data relay, and also provides positioning and context imaging for in situ measurements, as has been demonstrated at Mars and Titan. More broadly speaking, the combination of geological data from surface and radar elements with new geochemical constraints from isotopic ratio measurements and in situ atmospheric data would reveal the evolution of Venus and its climate, with relevance to terrestrial planets everywhere.

To address these themes **we propose a strawman mission based on a combination of an in situ balloon platform, a radar-equipped orbiter, and (optionally) a descent probe.** These mission elements are modelled on the 2010 EVE M3 mission proposal, on, the 2010 EnVision M3 proposal (which proposes re-use of ESA's GMES Sentinel-1 radar technology), and on Russia's Venera-D entry probe, respectively. The science themes described in the present white paper are phrased specifically in terms of Venus, but could equally be included in broader science themes such as comparative planetology or solar system evolution, both of which are central to our understanding of our own planet and of the diversity of exoplanets being discovered.

Written by Colin Wilson (Oxford University, UK) with contributions from K. Baines (U. Wisconsin & JPL, USA), E. Chassefière (Univ. Paris-Sud, France), R. Ghail (Imperial College, UK), R. Ghent (U. Toronto, Canada), J. Helbert (DLR, Germany), and the EVE and Envision Science Teams.

1 INTRODUCTION

1.1 THE IMPORTANCE OF VENUS FOR COMPARATIVE PLANETOLOGY

One of the central goals of planetary research is to find our place in the Universe. Investigation the evolution of solar systems, planets, and of life itself, is at the heart of this quest. The three terrestrial planets in our solar system - Earth, Mars, and Venus - show a wide range of evolutionary pathways, and so represent a “key” to our understanding of planets and exoplanets.

Earth and Venus were born as twins – formed at around the same time, with apparently similar bulk composition and the same size. However, they have evolved very differently: the enormous contrast between these planets today challenge our understanding of how terrestrial planets work. The atmosphere is surprising in many ways – its 400 km/h winds on a slowly rotating planet; its enormous surface temperature, even though it absorbs *less* sunlight than does the Earth; its extreme aridity, with sulphuric acid as its main condensable species instead of water. The solid planet, too is mysterious: its apparent lack of geodynamo and plate tectonics, the uncertainty of its current volcanic state, the apparent young age of much of its surface. How and why does a planet so similar to Earth end up so different?

In an era where we will soon have detected hundreds and then thousands of Earth-sized exoplanets, planetary science must seek to characterise these planets and to explain the diverse outcomes which may befall them. Are these exoplanets habitable? Many efforts have been made to define the edges of the ‘habitable zone’, i.e. the range of distances from a parent star at which a planet can sustain liquid water on its surface. The inner edge of the habitable zone has been estimated to lie anywhere from 0.5 to 0.99 AU – this latter figure, from Kopparapu et al., 2013, should be a cause of concern for us Earth-dwellers! Detailed study of Venus is indispensable if we are to understand what processes determine the inner edge of the habitable zone.

The habitable zone’s boundaries will evolve over a planet’s lifetime due to the evolution of the star’s output as well as changes in the planet and its atmosphere. There are only three terrestrial planets at which we can study geophysical and evolutionary processes: Venus, Earth, and Mars. Exploration of the latter two is firmly established, while in contrast there are no Venus missions currently planned after Venus Express.

1.2 CONTEXT: THE STATE OF VENUS SCIENCE AFTER VENUS EXPRESS

Venus Express has been a tremendously successful mission. Since its arrival at Venus in April 2006 it has made a wealth of discoveries relating to the atmosphere at all altitudes from the surface up to the exosphere. It has mapped cloud motions to reveal wind

velocities at different altitudes; it has measured the spatial distributions of key chemical species, and discovered new ones; it has measured the rate at which oxygen and hydrogen are being lost to space; it has found signs of frequent lightning; it has found signs

of recent lava flows on the surface using surface mapping at 1 μm wavelength.

In short Venus Express has provided much-needed data for atmospheric dynamics, chemistry, and radiative transfer, and for understanding of its ionosphere & induced magnetosphere. These new data are invaluable for constraining models of how Venus works today. However, the history of Venus still remains enigmatic. In this paper we propose a new set of investigations that focus on understanding the evolution of Venus through a combination of surface and atmospheric investigations.

Thanks to Venus Express, Europe is at the forefront of Venus research. Research groups across Europe have participated in the construction of and analysis from the scientific payload; dozens of researchers have completed doctoral theses based on Venus Express research. Europe is thus well-placed to lead a future Venus mission. ESA can now build on this position, capitalising on investments made in Earth Observation programme and advanced satellite technologies, to address fundamental questions about the evolution of terrestrial planets and the appearance of life.

2 SCIENCE THEMES

2.1 GEOLOGY (INTERIOR, TECTONISM, VOLCANISM, GEOMORPHOLOGY, MINERALOGY)

The major unknowns in Venus geological science are associated with its resurfacing history and establishing whether it is currently geologically active.

Resurfacing history

The age of Venus' surface is poorly known. Unlike Mercury, the Moon, and Mars, Venus has a thick atmosphere that represents a powerful filter to small impactors. As a result, its crater population is limited to a few large craters; there are very few craters with diameters <20 km, and there are fewer than 1000 craters in total. The observed crater population offers poor constraints on surface ages, allowing a number of different production and resurfacing scenarios. These include catastrophic global lithospheric overturn which take place every 500 to 700 My [Turcotte et al., 1999], equilibrium resurfacing models more similar to those

found on Earth [Stofan et al. 2005], as well as many models in between.

The community has used the existing Magellan radar data to attempt to resolve these fundamental conflicts by applying mapping techniques to establish stratigraphic relationships among surface units and structures. NASA's Magellan orbiter, launched in 1989, obtained global radar maps with a spatial resolution of 100 – 200 m. An important limitation to using radar images for geological mapping is that geological mapping requires the ability to identify distinct rock units, whose formation represent geological processes (e.g., distinct lava flows or sedimentary units). Magellan imagery provides the opportunity to identify some units; for example, it is possible to map lava flow boundaries to high precision in some terrains. But in many, many other cases, the materials being mapped have been affected by later tectonic structures; moreover, the

highly deformed tessera terrain, thought to be the oldest on Venus, is characterized by overlapping structures whose relationships are ambiguous in currently available observations. Different methods for accommodating this complication have led to widely divergent mapping styles, which in turn have resulted in a range of surface evolution models, mirroring the range of interior evolution models (see e.g. review by Guest & Stofan, 1999). Many remaining debates over, for instance, the sequence and relative timing of tectonic deformation in the complex tesserae cannot be resolved using currently available radar data, because of the limitations represented by the spatial resolution and single polarization of those data.

Current geological activity

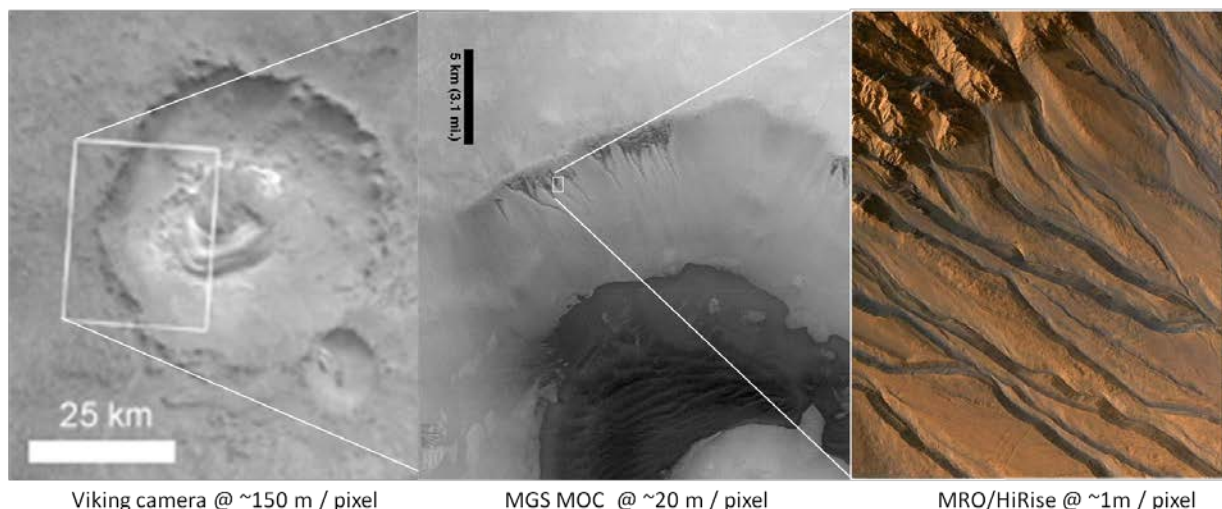
Venus is thought to have similar internal heat production to Earth, but it is not clear how the internal heat is lost to space. Is heat lost solely by crustal conduction, or does volcanic activity play an important role? Is the loss rate sufficient to maintain an equilibrium or is heat building up in the interior, potentially leading to an episodic resurfacing scenario? Understanding how Venus loses its internal heat is important for understanding both

Earth's earliest history and for understanding those exoplanets larger than Earth, both of which share its problem of a buoyant lithosphere; these mechanisms at work may profoundly affect the atmosphere and climate, and prove catastrophic for life.

There are some hints of recent geological activity particularly from Venus Express data (Smrekar et al., 2010, Marcq et al., 2012), but these analyses are indirect. A new radar dataset would enable not only better understanding of current surface weathering and alteration processes, which is needed in order to calculate ages for geologically recent changes such as lava flows and dune movements, but would also enable direct searching for surface change.

Case for next-generation radar:

Because of the extreme surface conditions and opaque clouds on Venus, geological investigation requires orbital remote sensing, with techniques including interferometric synthetic aperture radar (InSAR), gravimetry, altimetry and infrared observation using nightside infrared windows. The value of radar mapping at Venus was demonstrated by NASA's Magellan orbiter, launched in 1989, which obtained global radar

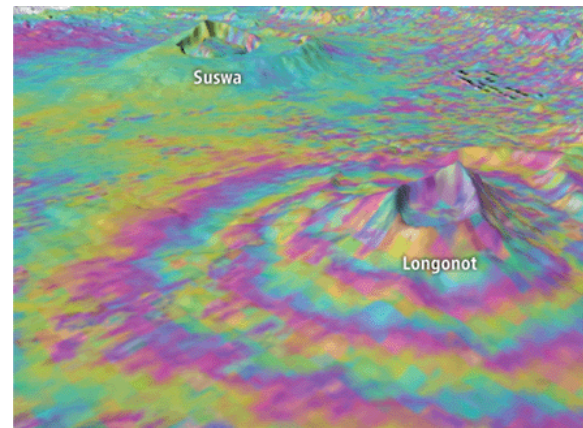


Images of the same region of Mars illustrate the revolutions in understanding which are enabled by increasing spatial resolution by an order of magnitude, particular for surface processes. Quoted pixel resolution is that of the displayed image used rather than the full resolution of the original. Image credits: NASA/JPL.

maps with a horizontal resolution of 100 – 200 m and altimetry with a vertical resolution of 100 m. Advances in technology, data acquisition and processing, and satellite control and tracking, mean that the spatial resolutions in the 1 – 10 m range are now possible.

This high-resolution radar mapping of Venus would revolutionise geological understanding. Generations of Mars orbital imagery have seen successive order-of-magnitude improvements, as illustrated below. As imagers progressed from the 50 m resolution of Viking towards the 5m resolution of MGS/MOC and the higher resolutions of MEx/HRSC and MRO/HiRise, our conception of Mars as a frozen, inactive planet was followed by hypotheses that geologically recent flow had occurred, to actual detection of current surface changes (e.g. gullies & dune movement). For Venus, metre-scale imagery will enable study of Aeolian features and dunes (only two dune fields have been unambiguously identified to date on Venus); will enable more accurate stratigraphy and visibility of layering; will constrain the morphology of tesserae enough that their stress history and structural properties can be constrained; will enable detailed study of styles of volcanism by enabling detailed mapping of volcanic vents and lava flows; and will enable direct search for surface changes due to volcanic activity and Aeolian activity. Metre-scale resolution would even enable search for changes in rotation rate due to surface-atmosphere momentum exchange, which could constrain internal structure (Karatekin et al., 2011).

The revolutions in radar performance go beyond just spatial resolution. Differential InSAR allows surface change detection at centimetre scale. This technique has been used to show surface deformations after earthquakes and volcanic eruptions on Earth



Differential InSAR revealed altimetry changes in this volcano in Kenya which previously had been thought to be dormant. [Sparks et al., 2012; observations from Envisat ERS-1 & ERS-2].

(see figure); similar results on Venus could provide dramatic evidence of current volcanic or tectonic activity.

The maturity of InSAR studies on Earth is such that data returned from Venus can confidently be understood within a solid theoretical framework developed from coupled terrestrial InSAR data and ground-truth observations, given the absence of ground-truth data on Venus. Note that the radar resolutions may be high enough to permit identification of Venera and Pioneer Venus landers, which serves as some ground truth even before a new generation of landers is taken into account. Radar mapping with different polarisation states constrains surface roughness and dielectric properties; mapping at different look angles and different wavelengths will provide further new constraints on surface properties.

The proximity of Venus to Earth, the relatively calm (if extreme) surface conditions and lack of water, the absence of a large satellite and its moderately well-known geoid and topography, all help to ease the technical demands on the mission. Europe is the world leader in radar systems and could, with minimal cost and development, adapt a GMES Sentinel-1 or NovaSAR-S modular array antenna for use at Venus, providing higher-resolution imagery, topography, geoid and

interferometric change data that will revolutionise our understanding of surface and interior processes.

In addition to radar techniques, the surface can also be observed by exploiting near-infrared spectral window regions at wavelengths of 0.8 – 2.5 μm . On the nightside of Venus, thermal emission from the surface escapes to space in some of these spectral windows, allowing mapping of surface thermal emissivity, as demonstrated by VEx/VIRTIS [Mueller et al., 2008]. A new instrument optimised for this observation could map mineralogy and also monitor the surface for volcanic activity [Ghail et al., 2012].

Case for Venus in situ geological investigations:

The Venera and Vega missions returned data about the composition of Venus surface materials, but their accuracy is not sufficient to permit confident interpretation. The Venera and VEGA analyses of major elements (by XRF) did not return abundances of Na, and their data on Mg and Al are little more than detections at the 2 σ level. Their analyses for K, U, and Th (by gamma rays) are imprecise, except for one (Venera 8) with extremely high K contents (~4% K₂O) and one (Venera 9) with a non-chondritic U/Th abundance ratio. The landers did not return data on other critical trace and minor elements, like Cr and Ni. In addition, the Venera and VEGA landers sampled only materials from the Venus lowlands – they did not target sites in any of the highland areas, the coronae, tesserae, nor the unique plateau construct of Ishtar Terra. Currently available instruments could provide much more precise analyses for major and minor elements, even within the engineering constraints of Venera-like landers.

A new generation of geologic instrumentation should be brought to the surface of Venus,

including Raman/LIBS and XRF/XRD; this would allow mineralogical, as well as merely elemental, composition. Such precise analyses would be welcome for basalts of Venus' lowland plains, but would be especially desirable for the highland tesserae and for Ishtar Terra. The tesserae may well represent ancient crust that predates the most recent volcanic resurfacing event and so provide a geochemical look into Venus' distant past. Ishtar Terra, too, may be composed (at least in part) of granitic rocks like Earth's continental crust, which required abundant water to form. Coronae samples will reveal how magmatic systems evolve on Venus in the absence of water but possibly in the presence of CO₂, SO₂ or other volatiles. Surface geological analysis would benefit from high temperature drilling/coring and sample processing capabilities, although further investigation will be needed to assess the extent to which this can be within the scope of an L-class mission opportunity. Long-lived seismological stations would also be valuable in the longer term but they too are considered outside the scope of the current call.

Descent imaging has not yet been performed by any Venus lander. Descent imaging of any landing site would be useful, particularly so for the tessera highlands where it would reveal the morphology of the highland surfaces and yield clues as to what weathering processes have been at work in these regions. Multi-wavelength imaging in near-infrared wavelengths would yield compositional information to provide further constraints on surface processes. In particular, descent imaging can establish whether near-surface weathering or real compositional differences are the root cause for near-IR emissivity variations seen from orbit, (Helbert et al. 2008) providing important ground truth for these orbital observations.

Profiles of atmospheric composition in the near-surface atmosphere would reveal which chemical cycles are responsible for maintaining the enormously high carbon dioxide concentrations in the atmosphere, and would also reveal details about surface-atmosphere exchanges of volatiles. Several mechanisms have been invoked for buffering the observed abundance of carbon dioxide, including the carbonate (Fegley & Treimann

1992) or pyrite-magnetite (Hashimoto & Abe, 1997) buffer hypotheses. Measurements of near-surface abundances and vertical gradients of trace gases, in particular SO_2 , H_2O , CO and OCS , would enable discrimination between different hypotheses. Correlating these data with lander and orbiter data will reveal how important and widespread the sources and sinks of these species might be.

2.2 PLANETARY EVOLUTION AS REVEALED BY ISOTOPE GEOCHEMISTRY

Geology is a powerful witness to history, but it does not provide answers about evolution in the time before the formation of the oldest rocks, which on Venus were formed only about a billion years ago. For constraints on earlier evolution, we must turn to isotope geochemistry.

Radiogenic noble gas isotopes provide information about the degassing history of the planet. ^{40}Ar , produced from the decay of long-lived ^{40}K , has been continuously accumulated over >4 billion years and so its current abundance constrains the degree of volcanic/tectonic resurfacing throughout history. ^{129}Xe and ^{130}Xe are produced from the now extinct ^{129}I and ^{244}Pu respectively within the first 100Ma of the solar system's history. The depletion of these isotopes in the atmospheres of Mars and Earth reveal that these two planets underwent a vigorous early degassing and blow-off, although the mechanisms of this blow-off and of subsequent deliveries of materials from comets and meteorites vary according to different scenarios. Neither the bulk Xe abundance nor the abundances of its eight isotopes have ever been measured at Venus. Measurements of these abundances would provide entirely new constraints on early degassing history. ^4He , produced in the mantle from long-lived U and Th decay, has an

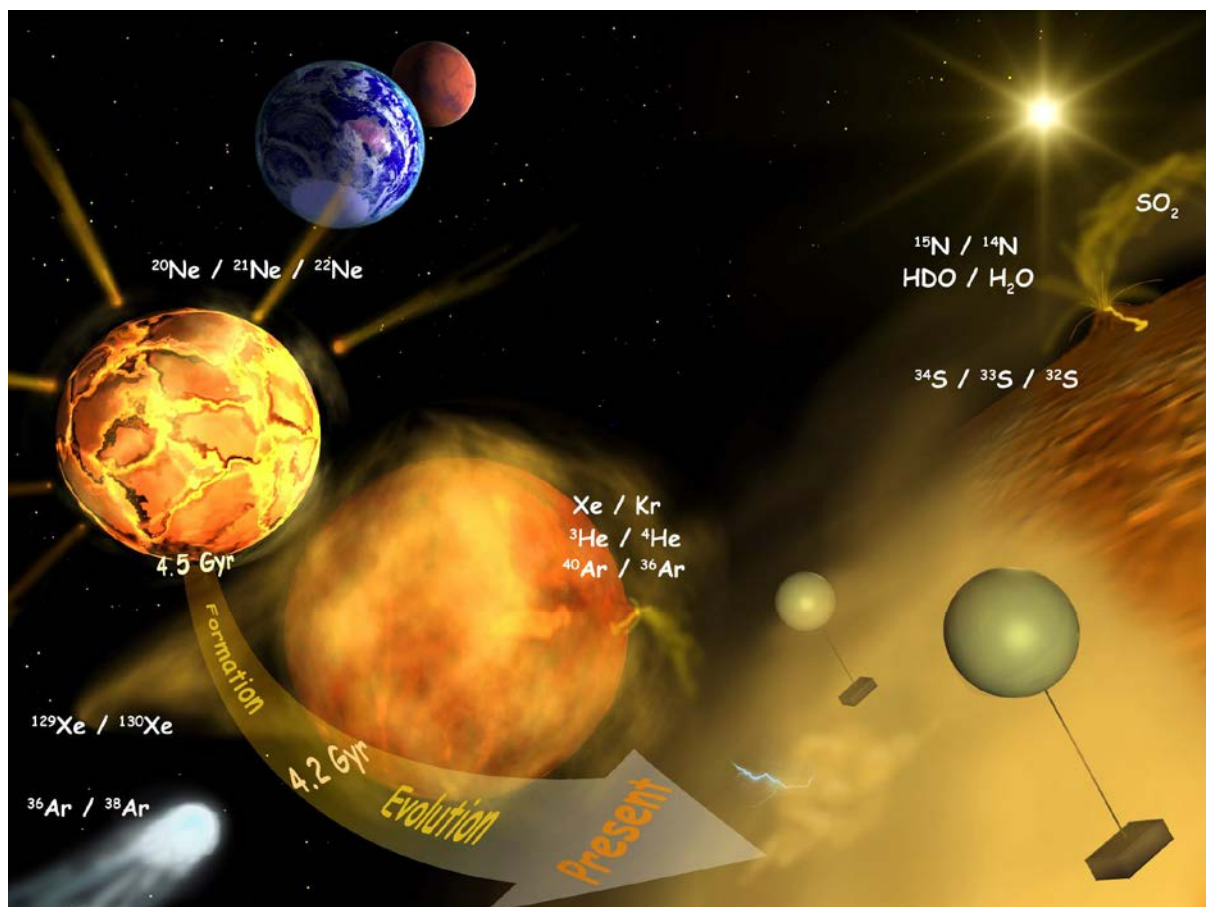
atmospheric lifetime of only a few hundred million years before it is lost by escape to space. Therefore its current atmospheric abundance provides constraints on recent outgassing and escape rates within the last $10^8 - 10^9$ years.

Non-radiogenic noble gas isotopes provide information about acquisition and loss of planet-forming material and volatiles. Venus is less depleted in Neon and Argon isotopes than are Earth and Mars, but its Xe and Kr isotopic abundances are still unknown (Xe isotopic abundances have not yet been measured, and past measurements of Kr abundance vary by an order of magnitude, providing little useful constraint). The significant fractionation of xenon on Earth and Mars can be attributed to massive blowoffs of the initial atmospheres in the period after the radiogenic creation of xenon from its parent elements, ~50-80 Myr after planet formation. However, it could also be that the fractionation is reflecting that of a source material delivered late in planetary formation, perhaps from very cold comets. If Venus has the same xenon fractionation pattern as Earth and Mars, this would support the idea that a common source of fractionated xenon material was delivered to all three planets, and weaken the case that these reflect large blowoffs. If we see a pattern with less Xe

fractionation then that would support the blow-off theory for Earth and Mars. Considered together with other non-radiogenic isotopic abundances, this allows determination of the relative importance of EUV, impact-related or other early loss processes. These measurements would also allow tighter constraints on the how much of the gas inventory originated from the original accretion disk, how much came from the solar wind, and how much came later from planetesimals and comets. Late impacts such as the Earth's moon-forming impact also have an effect on noble gas isotopic ratios so can also be constrained through these measurements.

Light element isotopic ratios, namely H, C, O N etc, provide further insights into the origin

and subsequent histories of planetary atmospheres. Measurement of $^{20}\text{Ne}/^{21}\text{Ne}/^{22}\text{Ne}$ and/or of $^{16}\text{O}/^{17}\text{O}/^{18}\text{O}$ would enable determination of whether Earth, Venus and Mars came from the same or from different parts of the protoplanetary nebula, i.e. are they truly sibling planets of common origin. Venus' enhanced Deuterium to hydrogen ratio, 150 times greater than that found on Earth, suggests that hydrogen escape has played an important role in removing water from the atmosphere of Venus, removing more than a terrestrial ocean's worth of water during the first few hundred million years of the planet's evolution (Gillmann et al., 2009). $^{14}\text{N}/^{15}\text{N}$ ratios have been found to vary considerably in the solar system, with the solar wind, comets,



Isotopic ratios provide keys to constrain planetary origins (Ne isotopes), early atmospheric loss processes (non-radiogenic Xe, Kr, Ar), late impact scenarios, recent mantle outgassing and late resurfacing, and escape of water. Modified from Baines et al., 2007.

and meteorites all exhibiting isotopic ratios different from the terrestrial values; it also is affected by preferential escape of ^{14}N . Measurement of the nitrogen isotopic ratio therefore helps establish whether the gas source was primarily meteoritic or cometary, and constrains history of escape rates.

Taken together, measurements of these isotopic abundances on Venus, Earth and Mars are needed to provide a consistent picture of the formation and evolution of these planets and their atmospheres, and in particular the history of water on Venus. Early Venus would have had an atmosphere rich in carbon dioxide and water vapour, like that of Hadean Earth. Hydrodynamic escape from this early steam atmosphere would have been rapid – but would it have been rapid enough to lose all of Venus' water before the planet had cooled enough to allow water to

condense? If Venus did have a liquid water ocean, how long did this era persist before the runaway greenhouse warming 'ran away', with the oceans evaporating and the resulting water vapour being lost to space? The nature of early escape processes is as yet too poorly constrained to answer these questions. An early Venus with a liquid water ocean would have arguably have been more Earthlike than was early Mars and could have taken steps towards development of life. A habitable phase for early Venus would have important consequences for our understanding of astrobiology and the habitable zones of exoplanets.

More detailed treatments of Venus isotope geochemistry goals and interpretation can be found in Chassefière et al., 2012 and Baines et al., 2007.

2.3 ATMOSPHERIC SCIENCE (DYNAMICS, CHEMISTRY & CLOUDS, STRUCTURE & RADIATIVE BALANCE)

The terrestrial planets today have very different climates. Study of the fundamental processes at work on these three planets will lead to a deeper understanding of how atmospheres work, and of climates evolve.

Dynamics and thermal structure

One of the crucial factors determining planetary habitability is the redistribution of heat around the planet. The solid planet of Venus rotates only once every 243 days but the atmosphere above exhibits strong super-rotation, circling the planet some 40-50 times faster than the solid planet below. General Circulation Models are now able to reproduce super-rotation, but are very sensitive not only to model parameters but also to the details of how those models operate. Efforts to improve modelling of the Venus atmosphere have led to improvements in how Earth handle details

like conservation of angular momentum and temperature dependent specific heat capacities [Bengtsson et al., 2013]. Exchanges of momentum between surface and atmosphere, an important boundary condition for the atmospheric circulation, depend sensitively on the thermal structure in the lowest 10 km of the atmosphere. Many probes experienced instrument failure at these high temperatures so the atmospheric structure in the lowest parts of the atmosphere is not well known. The atmospheric circulation is driven by solar absorption in the upper cloud; most of this energy absorption is by an as yet unidentified UV absorber, which is spatially and temporally variable. Efforts to identify this UV absorber by remote sounding have failed, so in situ identification will be necessary.

Knowledge of the wind fields on Venus is currently achieved by tracking cloud features or by tracking descent probes. Tracking by cloud features returns information about winds at altitudes from 48 to 70 km altitude, depending on the wavelength used. However, it is not clear at what altitude to assume that the derived cloud vectors apply; the formation mechanism for the observed contrasts is not known so it cannot be ascertained to what extent the derived velocity vectors represent true air motion rather than the product of, for example, wave activity. Furthermore, cloud tracking on the day- and night-sides of Venus is accomplished using different wavelengths, referring to different altitudes, so global averages of wind fields are not achievable by wind tracking, frustrating attempts to understand global circulation.

Direct measurement of mesospheric wind velocities (or at least their line-of-sight components) from orbit can be achieved by using Doppler sub-millimetre observations, Doppler LIDAR or other such instruments, and this would help to constrain circulation models. Tracking of descent probes will yield direct measurement of the vertical profile of horizontal winds in the deep atmosphere, which will provide important constraints on the mechanisms of super-rotation. Balloon elements are ideal for measuring vertical wind speeds but also provide information about wave and tidal activity at constant altitude.

Chemistry and clouds

Venus has an enormous atmosphere with many complex chemical processes at work. Processes occurring near the surface, where carbon dioxide becomes supercritical and many metals would melt, are very different from those at the mesopause where temperatures can be below -150°C , colder than any found on Earth. A diversity of

observations is clearly required to understand this diversity of environments. While chemical processes in the mesosphere and lower thermosphere are now being studied by Venus Express, processes in the clouds and below are very difficult to sound from orbit and require in situ investigation.

The dominant chemical cycles at work in Venus's clouds are those linking the sulphuric acid and sulphur dioxide: Sulphuric acid is photochemically produced at cloud-tops, has a net downwards transport through the clouds, and then evaporates and then thermal dissociates below the clouds; this is then balanced by net upwards transport through the clouds of its stable chemical precursors (SO_2 and H_2O). Infrared remote sensing observations of Venus can be matched by assuming a cloud composition entirely of sulphuric acid mixed with water, but Vega descent probe XRF measurements found also tantalising evidence of P, Cl and even Fe in the cloud particles [Andreichikov et al., 1987]. If confirmed these measurements would provide important clues as to exchanges with the surface: are these elements associated with volcanic, Aeolian or other processes? An in situ chemical laboratory floating in the clouds, or multiple descent probes, would be needed to address these measurement goals.

As to lower atmosphere chemistry, it is poorly understood because the rapidly falling descent probes did not have time to ingest and fully analyse many atmospheric samples during their brief descent. Sub-cloud hazes were detected by several probes but their composition is unknown. Ground- and space-based observations in near-infrared window regions permit remote sounding of only a few major gases, but many minor species which may play important catalytic or intermediate roles in chemical cycles cannot be probed remotely. Spatial variation of volcanic gases

would be a direct way of finding active volcanism on Venus but would be very difficult to achieve. However, as mentioned above, near surface abundances and their vertical profiles will permit determination of whether there are active surface-atmosphere exchanges taking place and of what surface reactions are buffering the atmospheric composition.

Thermal structure and radiative fluxes

The cloud layer of Venus is highly reflective so Venus presently absorbs less power from the sun than does the Earth. Its high surface temperature is instead caused by its enormous greenhouse warming effect, caused by carbon dioxide, water vapour and other gases. Its clouds, too, have a net warming effect because they prevent thermal fluxes from escaping the deep atmosphere.

1-D radiative and radiative-convective models for the determination of climate are suddenly widespread as researchers worldwide attempt to determine the likely climate of exoplanets.

Venus offers a proving ground for these models much closer to home, one where the conditions are much better known than on exoplanets. Radiative transfer calculations on Venus are difficult: uncertainties in the radiative transfer properties of carbon dioxide at high temperatures and pressures are the main unknown, particularly in the middle- and far- infrared where there are no spectral window regions to allow empirical correction. As on Earth, clouds play an important role, reflecting away sunlight but also trapping upwelling infrared radiation. The state-of-the-art Venus radiative balance are still only 1-D models representing an average over the whole planet. However, we now know that the clouds are very variable; the vertically integrated optical thickness (as measured at 0.63 μm) can vary by up to 100% [Barstow et al 2012] and the vertical structure of clouds varies strongly with latitude. In-situ measurements of cloud properties with co-located radiative flux measurements are needed to determine the diversity of cloud effects on the global radiative balance.

3 MISSION ELEMENTS

A satellite in low, near-circular polar orbit is required for radar mapping, and for LIDAR and Doppler sub-mm measurement of wind speeds. Wide angle cameras like the MRO/MARCI camera can be used to obtain continuous imaging coverage of UV cloud-top features. Recent examples of proposed low circular Venus orbiters include the Envision radar mapper [Ghail et al., 2012], the RAVEN radar mapper [Sharpton et al., AGU 2009], VERITAS radar mapper [Hensley, Smrekar et al., AGU 2012], MuSAR radar mapper [Blumberg, Mackwell et al., URSI 2011], and the Vesper sub-mm sounding orbiter [Allen et al., DPS 1998].

A satellite in a highly elliptical orbit can provide synoptic views of an entire hemisphere at once; One example of this is Venus Express, whose polar apocentre allows it to study the vortex circulation of the South polar region; a second example is the nearly equatorial 40-hour orbit of Japan's Akatsuki orbiter, which allows it to dwell over low-latitude cloud features for tens of hours at a time. The large range of altitudes covered is also useful for in situ studies of thermosphere and ionosphere and thus for studies of solar wind interaction and escape.

Balloons are ideally suited for exploring Venus because they can operate at altitudes where pressures and temperatures are far more

benign than at the surface. Deployment of two small balloons at 55 km altitude, in the heart of the main convective cloud layer, was successfully demonstrated by the Soviet VeGa mission in 1984. At this altitude, the ambient temperature is a comfortable 20° C and the pressure is 0.5 atm. The main environmental hazard is the concentrated sulphuric acid which makes up the cloud particles; however, effects can be mitigated by choosing appropriate materials for external surfaces. Balloons at this altitude can take advantage of the fast super-rotating winds which will carry the balloon all the way around the planet in a week or less (depending on latitude and altitude). Horizontal propulsion (with motors) is not advised because of power requirements and the difficulty of countering the fast (250 km/h) zonal windspeed. A cloud-level balloon is an ideal platform for studying interlinked dynamical chemical and radiative cloud-level processes. It also offers a thermally stable long-lived platform from which measurements of noble gas abundances and isotopic ratios can be carefully carried out and repeated if necessary (in contrast to a descent probe, which offers one chance for making this measurement, in a rapidly changing thermal environment).

Balloons can be used to explore a range of altitudes. Operation in the convectively stable upper clouds, above 63 km, would be optimal for identification of the UV absorber, but the low atmospheric density leads to a relatively small mass fraction for scientific payload. Operation below the main cloud deck at 40 km, has been proposed by Japanese researchers, with a primary goal of establishing wind fields below the clouds. Balloons can also be used to image the surface, if they are within the lowest 1-10 km of the atmosphere, but high temperatures here require exotic designs such as metallic bellows which are beyond the scope of this

paper. An intriguing possibility for revealing winds in the lower atmosphere is to use passive balloons, reflective to radio waves, which could be tracked by radar – this possibility should be studied further if a radar + entry probes architecture were to be studied further.

Descent probes provide vertical profiles of composition, radiation, chemical composition as a function of altitude, and enable access to the surface. Science goals for a descent probe include: cloud-level composition and microphysical processes; near-surface composition, winds, and temperature structure; surface composition and imaging; and noble gas abundances and isotopic ratios. If the scientific focus of the probe is measurements at cloud level then a parachute may be deployed during the initial part of the entry phase in order to slow the rate of descent during the clouds. This was carried out, for example, by the Pioneer Venus Large Probe. Alternatively, if the main focus of the probe is measurements in the lower atmosphere or surface then the probe may dispense with a parachute completely. Descent imagery of impact/landing site at visible wavelengths will be invaluable for contextualising surface results; Rayleigh scattering limits the altitudes from which useful surface imagery can be obtained to ~1 km if imaging in visible wavelengths, or to 10 km if imaging at 1 µm wavelength [Moroz, PSS 2002].

Surface elements – including static landers and rovers – must cope with the harsh conditions of ~450 °C at the surface of Venus. Rovers and landers which require sustained surface operations would require nuclear power due to the low levels of sunlight reaching the surface, and are considered beyond the scope of this proposal. However, a surface element using passive thermal control – relying on thermal inertia and thermal

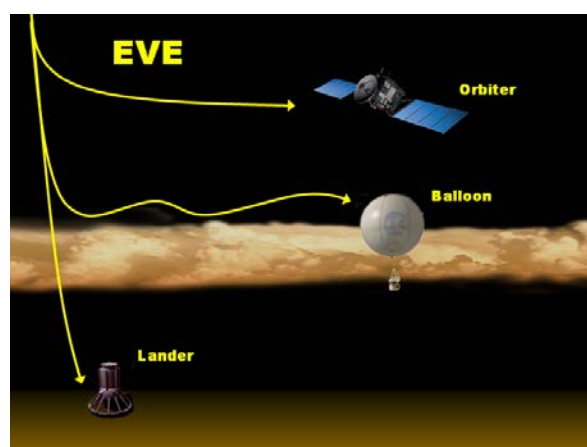
insulation to keep a central electronics compartment cool – allows operation times of hours or even days [see e.g. Venera-D mission, Vorontsov et al., Solar System Research 2011]. The principal science payload of a surface element would include surface imagers and non-contact mineralogical sensors such as Gamma spectrometer with Neutron activation, capable of measuring elemental abundances of U, Th, K, Si, Fe, Al, Ca, Mg, Mn, Cl (Li, Mitrofanov et al., EPSC 2010) and/or Raman/LIBS (Clegg et al., LPSC 2011). Inclusion of surface sample ingestion via a drill/grinder/scoop would allow further

analysis techniques (e.g. mass spectroscopy; X-ray fluorescence (XRF) spectroscopy) but would require significant technology development and verification. Gamma- and XRF spectroscopy have been performed on Venera and Vega landers, but modern equivalents of these instruments would provide much improved accuracy; also, repeating the composition analyses at a tessera region (not before sampled) would reveal whether these tessera regions are chemically differentiated from the lava plains where previous analyses have been conducted.

4 A STRAWMAN MISSION ARCHITECTURE

A Large mission to Venus should include both orbital and in situ science measurements. One possible strawman mission concept which would could address this theme would be a combination of **an orbiter, a cloud-level balloon platform, and (optionally) a Russian descent probe**. As a strawman payload, we suggest the balloon element be modelled on the 2010 EVE M3 proposal [Wilson et al., 2012]. The radar orbiter may be based on a reuse of ESA's GMES Sentinel-1 InSAR technology, whose application at Venus was first described in the 2010 Envision M3 proposal [Ghail et al., 2012]. Finally, the landing probe envisaged is based on the lander component of the Venera-D mission [Vorontsov et al., 2011]

It is very important to have multiple mission elements working together simultaneously at Venus. An orbiter is necessary both to increase vastly the volume of data returned from the in situ elements, but also to place those in situ measurements into atmospheric and geological context. The in situ measurements are required to measure parameters, like noble gas abundances and surface mineralogy, which cannot be



determined from orbit. The whole mission is greater than the sum of its parts. This has been amply demonstrated by the constellation of missions at Mars, and by the Cassini/Huygens collaboration at Titan.

Not all of these mission elements need be provided by ESA; There is ample scope for international co-operation in creating this mission architecture. In particular, Russia has unequalled heritage in providing Venus descent probes from its Venera and Vega descent probes, and will gain new heritage from its Venera-D lander, planned for the coming decade. A range of mission proposals have been developed in the USA for orbiters,

balloons and descent probes, from Discovery-class to Flagship-class, many of which could form parts of a joint NASA-ESA exploration programme should a high-level agreement be reached. Japan has an active Venus research community, with its Akatsuki (Venus Climate Orbiter) spacecraft still in flight, and has been developing prototypes for a Venus sub-cloud balloon [Fujita et al., IPPW 2012]. After its recent successful launches of radar satellites RISAT-1 and RISAT-2, ISRO has also investigated possibilities of an Indian Venus mission [Anurup et al., 2012], by 2030 this may be a real possibility. Israel's TECSAR satellites enable 1-m scale radar mapping with a 300 kg satellite, in the frame of future ESA-Israel agreements, collaborations on Venus radar could be fruitful. In this time frame collaborations with China are also feasible.

These could be launched as a stack on a single launcher, or it may prove convenient to use separate launchers, for example in order to insert the orbiter into a low circular orbit before the arrival of the in situ elements for optimal data relay and context remote sounding for the in situ measurements.

This scenario, Orbiter + balloon + Descent Probe, was proposed to ESA in 2007 in response to the M1/M2 mission Call for Ideas as a joint European Russian mission, with a European-led orbiter and balloon, and a Russian descent probe. For EVE 2007, the entire mission was to be launched on a single Soyuz launch, however subsequent studies revealed that this scenario was not consistent with a single Soyuz launcher, and would be more consistent with an L-class rather than an M-class opportunity.

5 TECHNOLOGY DEVELOPMENTS NEEDED

Much of the technology required for a Large Venus mission already exists at a high Technology Readiness Level, but further technology development both for spacecraft technologies and for science payload technologies would be useful to maximise science return and de-risk mission aspects.

Many of the mission-enabling technologies required for Venus exploration are shared with other targets. Aerobraking/aerocapture would improve Δv and mass budgets for Venus orbiters. Further improvement in deep space communications, including development of Ka-band or optical communications, would provide increased data return from orbiters which would be particularly useful for the large volumes of data generated by high-resolution radar instrumentation at Venus – 100 Mbps or better would enable the maximum scientific return from a radar orbiter. High

speed entry modelling, thermal protection systems and parachute development will all be useful for Venus entry probes. Nuclear power systems are not necessary for the strawman Venus mission described here, but would enable longer lifetimes and increased nightside operations for the balloon element of the mission, and will be necessary for long-lived surface stations (in the even more distant future) due to the scarcity of sunlight reaching the Venus surface and long night-time duration.

Two technology areas specific to Venus exploration are balloon technology, and high-temperature components. The balloons proposed in this strawman mission are helium superpressure balloons, which are designed to float at constant altitude. Although ESA has little familiarity with this technique, thousands of helium superpressure balloons have been launched on Earth, and two were

successfully deployed on Venus in 1985 as part of the Russian VeGa programme, so this is a very mature technology. Air-launching of balloons – deploying them from a probe descending under parachute – was achieved by the Russian VeGa balloons, was demonstrated by CNES in VeGa development programmes, and demonstrated recently by JPL engineers in their own Venus balloon test programme; nevertheless, a new demonstration programme in Europe would be required to obtain recent European experience for this technology. Balloon envelope design and sulphuric acid resistance verification would also be valuable to conduct in Europe. Feasibility studies on other forms of aerial mobility, including phase change fluid balloons, <5 kg microprobes and fixed wing aircraft, would also be useful for expanding the possibilities of long-term future exploration programmes. Compact X-band phased array antennae are well suited for mobile atmospheric platforms like balloons (or indeed rovers), investment in these systems would be valuable.

High temperature technologies are only needed if Europe is to provide mission elements or payloads which need to operate in the lower atmosphere, below 40 km altitude. The descent probe proposed in the present strawman mission proposal is a classical Venera-type design with all electronics and most sensors inside a layer thermal insulation in a central compartment with high thermal inertia. This kind of passive thermal design permits operation on the surface for hours or even days. The more components can be placed on the outside of the probe, the longer the mission lifetime can be extended. Development programmes in high-temperature electronic components like amplifiers for telecommunications systems would therefore be valuable.

Further investment in scientific payload development is also needed to get the most science return from a large Venus mission. The radar is a critical element of the mission and further optimisation with respect to that proposed in the Envision proposal is still possible. Thanks to technology developments in the last ten years (including ESA's Sentinel-1A mission, Israel's TECSAR mission and the UK's NovaSar-S developments), metre-scale InSAR mapping has become not only possible but also affordable and deployable in small spacecraft. Further development would be valuable to adapt these systems for Venus, including further work on surface height change detection through differential interferometric SAR. Investment in optical and sub-mm heterodyne receivers would hasten the development of instruments capable of measuring mesospheric wind velocities using Doppler techniques.

For atmospheric in situ measurements, a key instrument is a mass spectrometer with getters and cryotrap to isolate and precisely measure the noble gas and light element isotope abundances. These technologies have been developed for Mars (e.g. MSL/SAM, ExoMars/PALOMA proposal), but further development to maximise the precision of the measurements and to optimise the development for the thermal environment of Venus balloons and descent probes will be needed. In situ GC+MS characterisation atmospheric chemistry is another mature field, but further development in particular of an aerosol collector system to allow detailed characterisation of cloud particle composition would be valuable.

Specific payload developments for a landing probe should also include surface characterisation instruments – gamma ray and neutron spectrometers, XRF/XRD, Raman instruments for mineralogical identification. High-temperature drilling and sample

ingestion systems are not currently proposed for the Venera-D lander included in the strawman mission, but some feasibility studies in this area would be a useful investment. Finally, high temperature chemical, meteorological and seismological sensors using silicon carbide semiconductor technology would usefully complement a lander's payload.

In summary, most of the technologies needed for a Large class Venus mission already have high levels of heritage either from Earth or from other planetary missions, but a well-targeted development programme would de-risk mission elements and maximise the science return. Many of the development areas identified would also benefit other space missions and have spin-out potential on Earth.

6 CONCLUSIONS

As we become aware of Earth's changing climate, and as we discover terrestrial planets in other solar systems, we gain ever more reasons to study the Earth's nearest neighbour and closest sibling.

For the scientific and programmatic reasons outlined in this document, Venus is a compelling target for exploration. The science themes important for Venus research – comparative planetology and planetary evolution – are common to all of planetary and exoplanetary science, and many of the instrumentation required – in situ mass spectrometry, radar and atmospheric remote sensing – are found in mission proposals for many other solar system targets. Venus is close to the Earth, which leads to a short cruise phase (typically only 5 months) and high data rates, and also close to the sun, resulting in plentiful solar power with modest solar arrays. These factors ensure that it will be a highly attractive target for emerging space nations to send missions, so is an ideal arena for inter-agency collaboration.

Venus is thus an excellent proving ground for international collaboration on large space missions; an excellent proving ground for new instrumentation before it is sent on long journeys to the distant reaches of the solar system; an excellent proving ground for techniques of analysis of exoplanets; an excellent proving ground for fundamental understanding of geophysical processes of terrestrial planets; an indispensable part of our quest to understand the evolution of Earthlike planets. For all these reasons, a Venus mission would be a strong candidate for an ESA Large-class mission in the coming decades and we therefore propose science themes of planetary evolution and comparative planetology for the L2/L3 opportunities.

REFERENCES

- Allen, M. et al. (1998), The VESPER Mission to Venus. American Astronomical Society, DPS meeting abstract #48.P08.
- Andreichikov, B.M. et al. (1987), VEGA 1 and 2 X-ray radiometer analysis of the Venus cloud aerosol. *Kosmicheskie Issledovaniia* (ISSN 0023-4206), vol. 25, Sept.-Oct. 1987, p. 721-736.
- Anurup, M.S. et al. (2012), Indian Scientific Mission to Venus – Future Prospects and Experiments. National Space Science Symposium 2012, Tirupati, India.
- Baines, K. H. et al. (2007), Experiencing Venus: Clues to the origin, evolution, and chemistry of terrestrial planets via in-situ exploration of our sister world. In *Exploring Venus as a Terrestrial Planet*. Geophysical Monograph 176, American Geophysical Union, Washington, DC. pp. 171-189.
- Barstow, J.K. et al (2012), Models of the global cloud structure on Venus derived from Venus Express Observations. *Icarus*, [doi:10.1016/j.icarus.2011.05.018](https://doi.org/10.1016/j.icarus.2011.05.018) .
- Bengtsson, L. et al. (2013), Towards Understanding the Climate of Venus: Applications of Terrestrial Models to Our Sister Planet. International Space Science Institute Scientific Report, [doi:10.1007/978-1-4614-5064-1](https://doi.org/10.1007/978-1-4614-5064-1).
- Blumberg, D.G. et al (2011), MuSAR: a novel SAR mission to Venus. Int. Union of Radio Science General Assembly 2011, abstract J05.1.
- Chassefière, E. et al. (2012), The evolution of Venus: Present state of knowledge and future exploration. *Planetary and Space Science*, [doi:10.1016/j.pss.2011.04.007](https://doi.org/10.1016/j.pss.2011.04.007) .
- Fegley B. Jr. & Treiman A.H. (1992) Chemistry of Atmosphere-Surface Interactions on Venus and Mars. in *Venus and Mars: Atmospheres, Ionospheres, and Solar Wind Interactions*. AGU, Geophysical Monograph No. 66, 7–71.
- Gillmann, C., et al. (2009), A consistent picture of early hydrodynamic escape of Venus atmosphere explaining present Ne and Ar isotopic ratios and low oxygen atmospheric content. *Earth Planet. Sci. Lett.* 286, 503–513.
- Guest, J.E. & Stofan, E.R. (1999), A New View of the Stratigraphic History of Venus. *Icarus*, [doi:10.1006/icar.1999.6091](https://doi.org/10.1006/icar.1999.6091).
- Fujita et al., (2012), An overview of Japan's planetary probe mission planning. International Planetary Probe Workshop, Toulouse, France, 2012.
- Ghail, R.C. et al. (2012), EnVision: taking the pulse of our twin planet. *Experimental Astronomy*, [doi:10.1007/s10686-011-9244-3](https://doi.org/10.1007/s10686-011-9244-3).
- Hashimoto, G.L. & Abe, Y. (1997). Venus' surface temperature controlled by a coupled mechanism of chemical and albedo feedback. *Bull. Amer. Astron. Soc.* 29, 1043.
- Helbert, J. et al. (2008), Surface brightness variations seen by VIRTIS on Venus Express and implications for the evolution of the Lada Terra region, Venus. *Geophysical Research Letters*, doi: 10.1029/2008GL033609
- Hensley, S. et al. (2012), VERITAS: A Mission Concept for the High Resolution Topographic Mapping and Imaging of Venus. AGU Fall Meeting 2012 abstract #P33C-1950.
- Karatekin, Ö. et al. (2011), Atmospheric angular momentum variations of Earth, Mars and Venus at seasonal timescales. *Planetary and Space Science*, [doi:10.1016/j.pss.2010.09.010](https://doi.org/10.1016/j.pss.2010.09.010) .
- Kopparapu, R.K.. et al. (2013), Habitable zones around main-sequence stars: new estimates. *Astrophys. J.*, doi: [doi:10.1088/0004-637X/765/2/131](https://doi.org/10.1088/0004-637X/765/2/131)
- Marcq, E. et al., Variations of sulphur dioxide at the cloud top of Venus's dynamic atmosphere. *Nature Geoscience*, [doi:10.1038/NCEO1650](https://doi.org/10.1038/NCEO1650) .

Mitrofanov et al., (2010), Neutron-Activated Gamma Ray Spectrometer (NAGRS) for the Venus Surface and Atmosphere Geochemical Explorer (SAGE) mission. European Planetary Science Congress 2010, abstract #264.

Moroz, V.I. (2002), Estimates of visibility of the surface of Venus from descent probes and balloons. *Planetary and Space Science*, [doi:10.1016/S0032-0633\(01\)00128-3](https://doi.org/10.1016/S0032-0633(01)00128-3).

Mueller, N. et al. (2008), Venus surface thermal emission at 1 μm in VIRTIS imaging observations: Evidence for variation of crust and mantle differentiation conditions, *Journal of Geophysical Research*, doi: 10.1029/2008JE003118

Sharpton, V.L. et al. (2009), RAVEN - High-resolution Mapping of Venus within a Discovery Mission Budget. AGU Fall Meeting 2009 abstract #P31D-04.

Smrekar, S.E. et al. (2010), Recent Hotspot Volcanism on Venus from VIRTIS emissivity data, *Science*, 328, p. 605-608, [doi:10.1126/science.1186785](https://doi.org/10.1126/science.1186785).

Sparks, R.S.J. et al. (2012), Monitoring volcanoes. *Science*, [doi:10.1126/science.1219485](https://doi.org/10.1126/science.1219485).

Stofan, E.R. & Smrekar, S.E., in *Plates, Plumes, and Paradigms*, (Special Volume 388, Geological Society of America, Denver, CO, 2005), pp. 861–885.

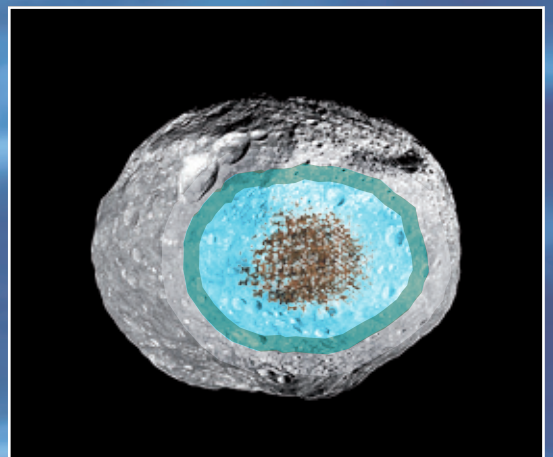
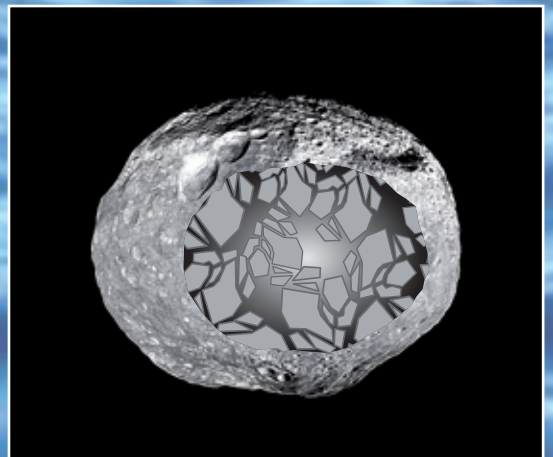
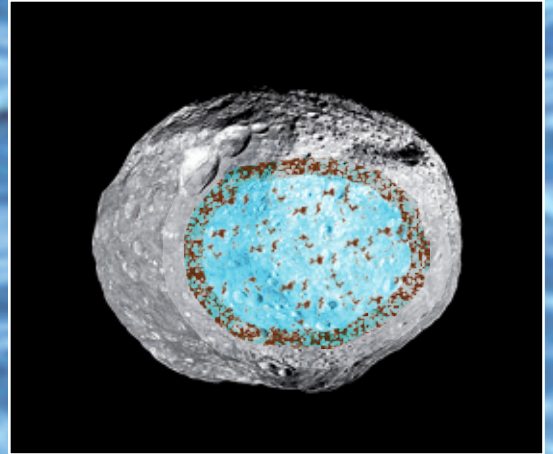
Turcotte, D. L. et al. (1999), Catastrophic resurfacing and episodic subduction on Venus. *Icarus*, [doi:10.1006/icar.1999.6084](https://doi.org/10.1006/icar.1999.6084).

Vorontsov, V.A. et al. (2011), Prospective Spacecraft for Venus Research: Venera-D Design. *Solar System Research*, [doi:10.1134/S0038094611070288](https://doi.org/10.1134/S0038094611070288).

Wilson, C.F. et al. (2013), The 2010 European Venus Explorer (EVE) mission proposal. *Experimental Astronomy*, [doi:10.1007/s10686-011-9259-9](https://doi.org/10.1007/s10686-011-9259-9).

INSIDER

Interior of Primordial Asteroids and the Origin of Earth's Water



A white paper submitted in response to ESA's L2/L3 call for ideas
May 24th, 2013

INSIDER:

Interior of Primordial Asteroids and the Origin of Earth's Water

ESA contact:

Pierre Vernazza
Laboratoire d'Astrophysique de Marseille, France
Email : pierre.vernazza@oamp.fr
Tel : +33491055911

Core team:

Philippe Lamy (FR), Matthieu Gounelle (FR), Benoit Carry (FR), Simone Marchi (US), Alessandro Morbidelli (FR), Aurelie Guilbert (NL)

Science community supporting the proposal:

Jessica Agarwal (DE), David Baratoux (FR), Antonella Barucci (FR), Pierre Beck (FR), Jim Bell (US), Philippe Bendjoya (FR), Sebastien Besse (NL), Jens Biele (DE), Mirel Birlan (FR), Phil Bland (AU), Bill Bottke (US), Daniel Britt (US), Rosario Brunetto (FR), Alberto Cellino (IT), Marc Chaussidon (FR), Francois Colas (FR), Audrey Delsanti (FR), Francesca DeMeo (US), Maria Cristina De Sanctis (IT), Alain Doressoundiram (FR), Rene Duffard (ES), Christophe Dumas (ESO), Josef Durech (CZ), Siegfried Eggl (FR), Josh Emery (US), Francesca Ferri (IT), Luigi Folco (IT), Sonia Fornasier (FR), Daniele Fulvio (DE), Jerome Gattacceca (FR), Maria Gritsevich (FI), Olivier Groussin (FR), Daniel Hestroffer (FR), Roger Hewins (FR), Albert Jambon (FR), Laurent Jorda (FR), Mikko Kaasalainen (FI), Tomáš Kohout (FI), Pedro Lacerda (UK), Claes-Ingvar Lagerkvist (SE), Jeremie Lasue (FR), Monica Lazzarin (IT), Jean-Pierre Lebreton (FR), Franck Marchis (US), Francesco Marzari (IT), Mario Melita (AR), Patrick Michel (FR), Alessandra Migliorini (IT), Olivier Mousis (FR), Frédéric Moynier (US), Thomas Mueller (DE), Alin Nedelcu (RO), Ernesto Palomba (IT), Paolo Paolicchi (IT), Nuno Peixinho (PT), Davide Perna (FR), Noemí Pinilla-Alonso (ES), Marcel Popescu (RO), Petr Pravec (CZ), Yoann Quesnel (FR), Nicolas Rambaux (FR), Henri Reme (FR), Andy Rivkin (US), Sebastien Rodriguez (FR), Philippe Rousselot (FR), Pablo Santos Sanz (ES), Colin Snodgrass (DE), François Robert (FR), Pierre Rochette (FR), Pascal Rosenblatt (BE), Mathieu Roskosz (FR), Sara Russell (UK), Rachel Soja (DE), Jessica Sunshine (US), Paolo Tanga (FR), Imre Toth (HU), Josep Maria Trigo (ES), Stephan Ulamec (DE), Esa Vilenius (DE), Brigitte Zanda (MNH)

Support from Thales Alenia Space (Cannes): Joel Poncy (FR)

Support from: Thierry Botti (FR), Melody Didier (FR)

Table of contents

1) Executive summary	3
2) Small Bodies at the crux of the formation and evolution of the solar system	4
3) The asteroid belt: a condensed version of the primordial solar system	5
4) The exploration of the diversity of the asteroid belt	8
4.1 Asteroids unconnected to meteorites	8
4.2 Metallic asteroids	9
4.3 Multiple asteroid systems	9
5) The first investigation of the internal structure of asteroids	9
5.1 Interior of primordial planetesimals: fingerprint of the time of formation	9
5.2 Which asteroid sizes should be targeted?	10
5.3 Which asteroid types should be targeted?	11
6) The origin of water on Earth	11
6.1 Asteroid and comets as plausible sources	11
6.2 From the D/H ratio to models of the solar system formation	12
7) Mission profile and instruments	14
7.1 Instrument payload description for Main spacecraft (MSC)	14
7.2 Lander	15
7.3 Mission	15
7.4 Flight system	16
8) Reference	17

1) Executive summary

Today's asteroid belt may not only be populated by objects that formed in situ, typically between 2.2 and 3.3 AU, but also by bodies that formed over a very large range of heliocentric distances. It is currently proposed that both the early (<5 Myrs after Solar System formation) and late (>700 Myrs after Solar System formation) dynamical evolution of the Solar System was governed by giant planet migrations that led to the insertion of inner (1–3 AU) as well as outer (4–13 AU) small bodies in the asteroid belt. Taken together, current dynamical models have the advantage to explain many striking features of the asteroid belt including i) its incredible compositional diversity deduced mainly from spectroscopic observations and meteorites measurements and ii) the evidence of radial mixing experienced by the various asteroid classes (e.g., S-, C-types) after their formation. In a broad stroke, the idea that the asteroid belt is a condensed version of the primordial Solar System is progressively emerging.

The asteroid belt, although still in the Earth's neighbourhood, therefore appears the ideal place for testing Solar System formation models by exploring the building blocks predicted by models of i) the telluric planets, ii) the giant planet cores, iii) the giant planets' satellites, and iv) outer small bodies such TNOs and comets. It also appears as an ideal place to search for the origin of Earth's water.

Up to now, only a few asteroid classes (e.g., several S-types) have been visited by spacecraft and the focus of these in situ measurements has been mainly to give a geological context to ground based observations as well as strengthen/validate their interpretation. Most of the tantalizing discoveries of asteroid missions have been realized via images of the objects surfaces. Time has come for asteroid space science to reach a new milestone by extending the reconnaissance of the Belt's diversity and addressing new science questions.

The scientific objectives of the proposed INSIDER mission require the exploration of diverse primordial asteroids - possibly the smallest surviving protoplanets of our Solar System - in order to constrain the earliest stages of planetesimal formation thus avoiding the effect of destructive collisions, which produce highly transformed rubble piles. Our science objectives that justify in situ measurements in the context of an L-class mission and that are expected to lead to significant breakthroughs include:

- 1) The exploration of the diversity of the asteroid belt
- 2) The first investigation of the internal structure of asteroids
- 3) The origin of water on Earth

The proposed mission scenario consists i) in successive rendez-vous followed by orbit insertion of multiple large ($D > 100$ km) objects (typically 4), ii) of a lander/rover that will study the composition (D/H ratio in particular) of a water-rich asteroid. The potential targets would therefore include one or two such bodies (e.g., 24 Themis), a binary or triple asteroid system, and one or two main belt interlopers (e.g., metallic and D-type asteroids).

Meeting our science objectives requires instruments (such as radar and magnetometer) not flown so far on past asteroids missions along with the traditional powerhouses, such as cameras and spectrometers.

2) Small Bodies at the crux of the formation and evolution of the Solar System

Small bodies in our Solar System are rocky and/or icy objects, usually ranging in size from a few meters to a few hundreds of kilometers. They comprise Asteroids, giant planet trojans, Trans-Neptunian Objects (TNOs) and Comets. Their physical nature, distribution, formation, and evolution are fundamental to understand how planets formed and ultimately, why water and life are present on Earth.

In our current Solar System, they are the most direct remnants of the original building blocks that formed the terrestrial planets and the solid cores of the giant planets. As such, they contain a relatively pristine record of the initial conditions that existed in our solar nebula some 4.6 Gyrs ago. The small bodies that have survived since that epoch, however, have experienced numerous collisional, dynamical, and thermal events that have shaped their present-day physical and orbital properties. Interpreting this record via observations, laboratory studies, and theoretical/numerical modeling can tell us much about the primordial state of these bodies and how they have evolved thereafter. In fact, even though small bodies represent only a tiny fraction of the total mass of the planets, their large numbers, diverse compositions, and orbital distributions provide powerful constraints for planet formation models.

The classical model for planetary formation involves three steps (Morbidelli et al. 2009):

- In Step 1, small bodies form. The small dust grains embedded in the gas-rich protoplanetary disk coalesce to form bodies roughly 1–1000 km in size called small bodies, which are large enough to acquire and retain additional material gravitationally [Chambers 2004, Morbidelli et al. 2009]. Asteroids, Comets, Trojans and TNOs are the only surviving witnesses of this step (Fig. 1).

- In Step 2, planetary embryos/cores form. Collisional coagulation among the small bodies allows the latter to agglomerate into massive bodies that are the building blocks of both the terrestrial planets and the cores of the giant planets. In the inner Solar System, the growth produces a population of planetary embryos, with lunar to martian masses. In the outer Solar System, beyond the so-called snowline, it is generally accepted that the end result is the formation of a few super-Earth cores [Chambers 2006] that, by accretion of a massive gaseous atmosphere from the disk, become giant planets [Alibert et al. 2004, 2005].

- In Step 3, the terrestrial planets form. In the inner Solar System, the system of embryos – whose initial formation locations can span several AU in width [Raymond et al. 2006] – becomes unstable and the embryos start to collide with each other, forming the terrestrial planets on a timescale of several $\sim 10^7$ to $\sim 10^8$ years [Raymond et al. 2007]. It remains to be understood whether the Earth acquired its water budget during this period or later on. Such information would, in turn, help constraining the necessary conditions for a planet to be hospitable for life.

There is now growing evidence that all three steps have been heavily affected by migration pre- and post-formation.

- a) The simultaneous presence of low- and high-temperature components believed to have formed respectively far from and close to the Sun in both chondritic (primitive) meteorites and in comets (Stardust mission, Spitzer space telescope observations; Zolensky et al. 2006, Lisse et al. 2006, Nakamura et al. 2008) implies that radial mixing in the protoplanetary disk during the early stage of step 1 played a prominent role in shaping the composition of small bodies (Ciesla 2007).

- b) Today's orbital architecture of Solar System small bodies further implies that migrations of the giant planets governed the dynamical evolution of the Solar System during steps 2 and 3 and even after step 3. It is currently proposed that both early (< 5 Myrs after Solar System formation) and late (> 700 Myrs after Solar System formation) migrations led to the insertion of inner (1–3 AU) as well as outer (4–13 AU) small bodies in the asteroid belt (Walsh et al. 2011, Morbidelli et al. 2005, Gomes et al. 2005, Tsiganis et al. 2005). giant planet migrations also help explaining both the orbital architecture of the Kuiper Belt and the existence of the Oort cloud (see Figure 1). Note that giant planet migrations appear as a natural consequence of the evolution of a planetary system as observations of exoplanets often reveal the existence of hot Jupiters.

In short, the study of the populations of solar system small bodies has been instrumental in

establishing the re-arrangement of the solar system architecture as a function of time, which could not have been inferred from the observations of the giant and/or telluric planets themselves. These objects thus hold the keys to our understanding of the dynamical evolution of the solar system for each of the above 3 steps. Their physical properties also conceal a chronology of processes, which took place from the end of step 3 until today. As such, small bodies are time capsules of the history of the solar system, imposing a detailed chronology of the major dynamical events.

At this stage, an important point is worth mentioning: the solar system is the only planetary system where small bodies AND the dynamical evolution of a planetary system as a function of time can be properly studied. We can indeed observe and study exo-planets on the one hand and protoplanetary/debris disks whose dust component is either primordial or produced by the collisional grinding of planetesimals on the other hand, but small bodies remain inaccessible. Concerning the dynamical evolution, observations of extra-solar systems offer only snapshots of their architecture at a given time limiting their study to statistically quantify the outcome of planetary formation. Whereas migrations appear ubiquitous as demonstrated by the detection of many hot Jupiters, we are presently unable to decipher the temporal evolution and the physical processes at the origin of the presently observed architectures of the exo-planetary systems.

Therefore, our solar system remains – via the study of small bodies – the only complete benchmark of planetary formation and evolution to which other planetary systems can be compared.

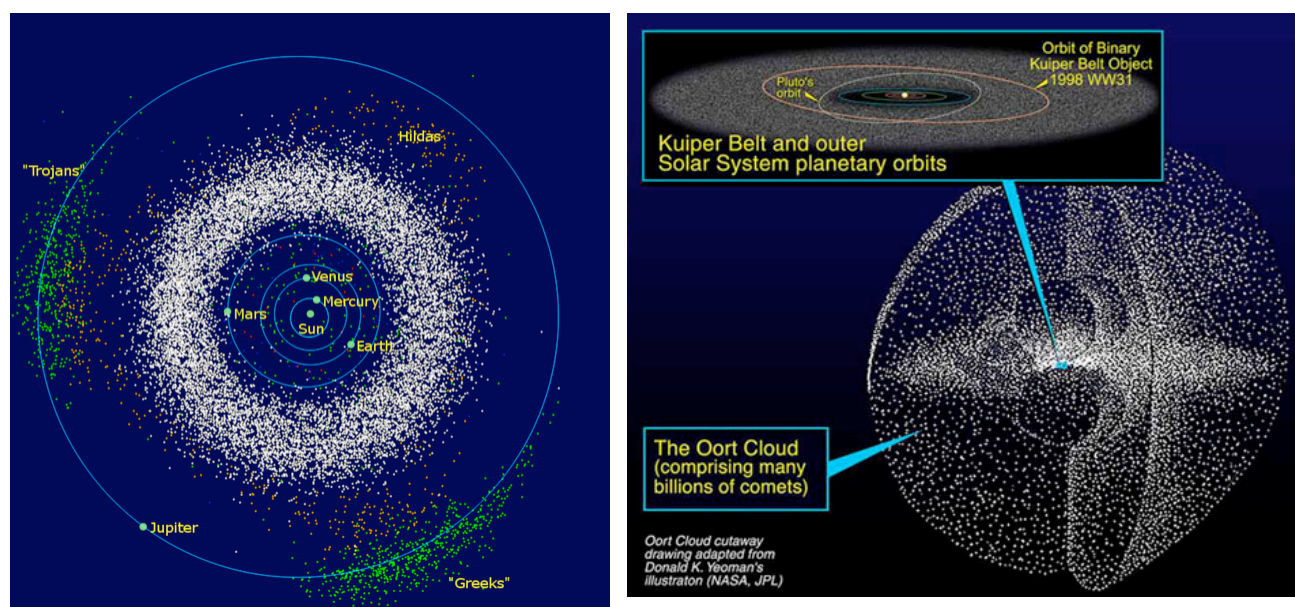


Figure 1: Inventory of the small bodies of the solar system. Left: The inner solar system, from the Sun to Jupiter. It shows the main asteroid belt (the white donut-shaped cloud), the Hildas (the orange "triangle" just inside the orbit of Jupiter) and the Jovian Trojans (green). Right: Artists rendering of the outer solar system, including both the Kuiper Belt and the Oort Cloud.

3) The asteroid belt: a condensed version of the primordial solar system

We summarize below our current understanding of the formation and evolution of the asteroid belt, an estimated population of $\sim 10^6$ objects larger than 1 km in diameter, in orbit between Mars and Jupiter. Among them, asteroid (1) Ceres is the largest one with a diameter of ~ 950 km. This knowledge is mostly based on i) ground based observations, ii) dynamical simulations, and iii) meteorite measurements, and to a lesser extent, on in situ observations by space missions as only eight asteroids have been visited by a spacecraft (5 flybys and only 3 rendez vous).

Spectroscopy in the visible and near-infrared range (VNIR, $0.4\text{--}2.5\ \mu\text{m}$) has proven to be a powerful

tool for constraining the surface composition of solar system bodies, asteroids in particular. As of today, more than 2000 main belt asteroids (Bus & Binzel 2002) have been observed via visible spectroscopy (0.4-0.9 μm) whereas a few hundred objects have been observed in the near-infrared (0.9-2.5 μm). **Their spectral properties in the combined VNIR range (0.4-2.5 microns) indicate the presence of an incredible number of compositional groups (24, see Figure 2). No other population of small bodies in our solar system exhibits such compositional diversity.** Only two compositional groups have been identified among Jupiter Trojans (Emery et al. 2011), TNOs seem to comprise four to five compositional groups only and most of the giant planet' satellites seem to be very similar in composition (water-rich surfaces).

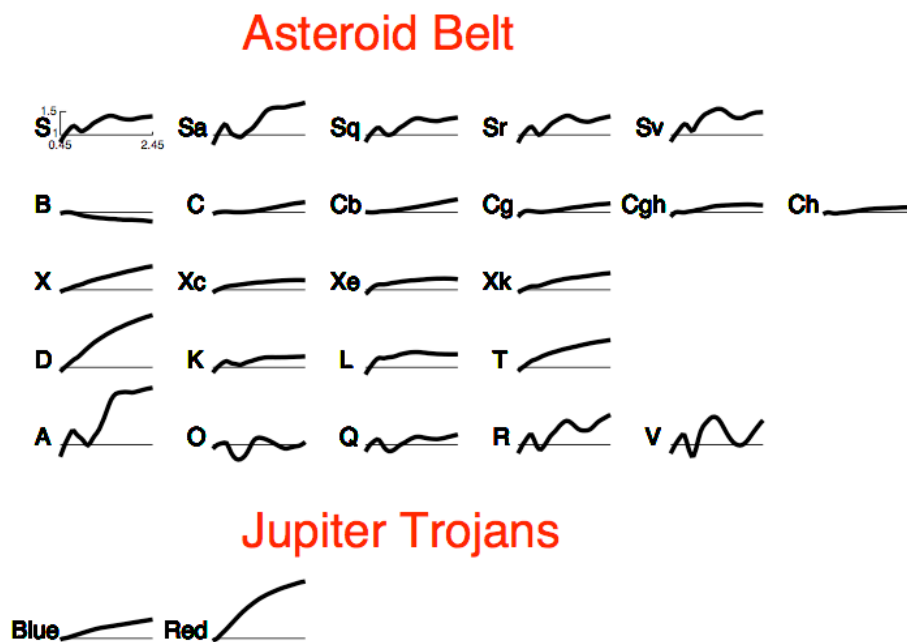


Figure 2: Compositional diversity of the asteroid belt and of the Jupiter Trojans. There are 24 spectral classes within the Main Belt and only two spectral classes among Jupiter Trojans.

When looking at the compositional distribution across the asteroid belt, there are three fundamental properties which directly constrain its formation and evolution:

- i) There are two main asteroid populations, the so-called S-types (composition: silicates including olivine and pyroxene) and C-types (composition: phyllosilicates) accounting for more than 50% of all main belt asteroids and several minor populations (compositions include iron-nickel, basalt, iron-poor silicates);
- ii) A heliocentric gradient (Gradie and Tedesco 1982) with water-free S-type asteroids being preferentially located in the inner belt while water-rich C-types are the dominant population in the outer belt (see Fig. 3);
- iii) A compositional overlap (for example, a large number of C-types are currently located where S-types are the most abundant and vice versa; see Fig. 3).

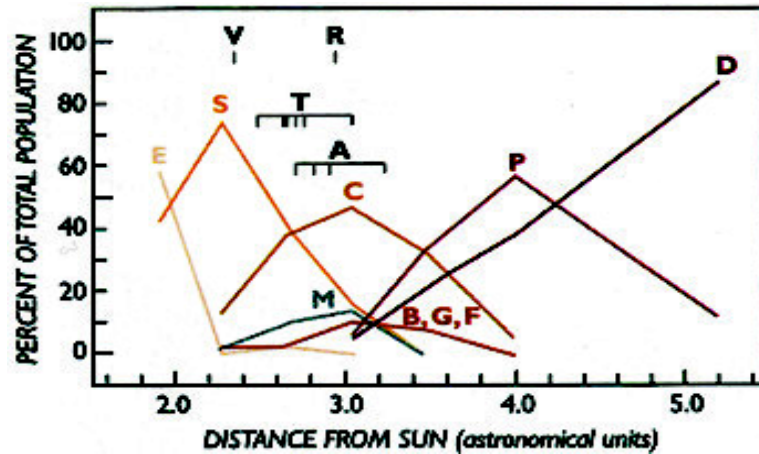


Figure 3: Compositional gradient within the asteroid belt (Gradie and Tedesco, 1982).

These three properties have been used as key tracers of the early dynamical evolution of the solar system. Bottke et al. (2006) proposed that a small population of planetesimals that formed in the terrestrial planet region has been scattered into the main belt by emerging protoplanets early in its history, thus predicting that some main belt asteroids are interlopers. Later, Levison et al. (2009) showed that the violent dynamical evolution of the giant-planet orbits required by the so-called Nice model (Gomes et al., 2005; Tsiganis et al., 2005; Morbidelli et al., 2005) which took place 700 Myrs after solar system formation led to the insertion of primitive trans-neptunian objects into the outer belt. Both these models help explaining the first property, namely the presence of several minor populations in the asteroid belt.

Recently, a new dynamical scenario proposed by Walsh et al. (2011) – invoking an early inward migration of Jupiter and Saturn to 1.5 AU in order to explain Mars' low mass – shows that the asteroid belt region may comprise bodies formed in the inner (1–3 AU) as well as the outer (4–13 AU) regions. In particular, their scenario helps explaining the second and third properties: the S- and C-types formed on each side of the snowline (<3 AU for S-types; >4 AU for C-types), explaining why S-types are water-free while C-types are water-rich. It further explains why S- and C-types overlap in heliocentric distance as a natural outcome of their respective migrations to their current locations.

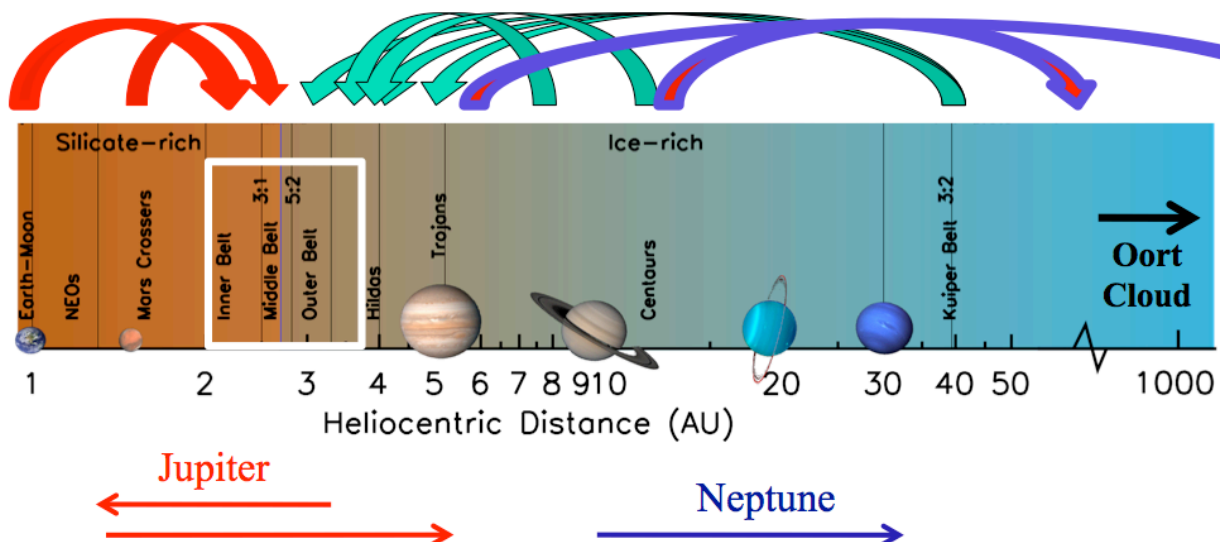


Figure 4: This schematic summarizes in a simple way the distribution of material across the solar system. The planets are pictured at the bottom and the location of different small body populations is labeled. Arrows show the migrations experienced by both the giant planets (bottom) and small bodies (top) as predicted by the most recent dynamical models. The result of those migration processes is that the asteroid belt may contain objects that formed throughout the solar system (DeMeo 2010).

In conclusion, the idea that the asteroid belt is a condensed version of the primordial solar system is progressively emerging (see Fig. 4). The asteroid belt (i.e., the Earth's neighbourhood) therefore appears an ideal place for testing solar system formation models and for exploring the building blocks predicted by models of i) the telluric planets, ii) the giant planet cores, iii) the giant planets' satellites, and iv) outer small bodies such as TNOs and comets. It also appears as an ideal place to search for the origin of Earth's water.

Whereas reaching a full comprehension of the solar system formation and evolution eventually requires collecting detailed constraints of the physical properties of ALL populations of small bodies, it therefore appears that the diversity of the main belt asteroids offers a first, simple approach to this global investigation. A large fraction of these constraints can certainly be inferred from both ground-based observations and meteorite measurements, but several key constraints require in situ measurements which, in the foreseeable future, are only reachable for relatively close objects. An example of the contribution of in situ observations come from the overall low number of large impact craters observed on the recently visited asteroids Lutetia and Vesta; they suggest that the primordial main belt was not drastically more massive than today (Marchi et al. 2012) in apparent contradiction with one of the prediction of the giant planet migration that leads to considerable depletion.

A collective brainstorming exercise between ground and space observers, meteoricists and dynamicists on the future in situ measurements that would not only lead to significant breakthroughs but also deliver constraints that can not be obtained from Earth neither today nor in the foreseeable future identified the following top level science objectives that justify an L-class mission:

- 1) The exploration of the diversity of the asteroid belt
- 2) The first investigation of the internal structure of asteroids
- 3) The origin of water on Earth

The sections below consider in detail these three science themes.

4) The exploration of the diversity of the asteroid belt

Only five asteroid types out of the identified 24 types (DeMeo's classification) have been observed at close range by space mission: four S-types (Gaspia, Ida, Eros, Itokawa), one V-type (Vesta), one Xe-type (Steins), one Xc-type (Lutetia), and one Cb-type (Mathilde). In particular, only three objects have been orbited (Eros, Itokawa, Vesta) for detailed investigation. It thus appears that we are only at the dawn of asteroid exploration. We review below specific questions which justify a major in situ observational effort.

4.1 Asteroids unconnected to meteorites

A large fraction of asteroid spectra are well matched by meteorite ones. Specifically, ground-based observations have revealed the following associations between asteroid and meteorite types:

- i) Ch and Cgh types - CM meteorites
- ii) K types - CV, CO, CR, CK meteorites
- iii) X types - iron meteorites
- iv) V types - HED meteorites
- v) Xc types - ECs and aubrites
- vi) T types - Tagish Lake meteorite
- vii) Xk types - mesosiderites
- viii) A types- pallasites and brachinites
- ix) S types - ordinary chondrites (~80% of the falls).

There are however several asteroid types which are not represented in our meteorite collections (e.g., B, C, Cb, Cg, D, L). Considering that these asteroids have nearly featureless spectra in the visible and near-infrared ranges, their surface composition has been and still is an open issue. In situ measurements (for instance, via Gamma spectroscopy) would help solving the question of their composition and thus of their origin.

4.2 Metallic asteroids

General speaking, metallic asteroids (X-types) and their meteoritic counterparts (iron meteorites) are among the most perplexing and mysterious objects among solar system small bodies. It is still not understood how these asteroids with diameters up to 250 km (Psyche) formed and what they actually represent. **Are they remnant cores of primordial differentiated protoplanets or did they directly form as we see them today?** In the first scenario, Psyche for instance would require that the mantle of a parent body of the size of Vesta (since Psyche's size is similar to Vesta's metallic core) has been totally blown off; if so, where has such a huge mantle gone? Whereas ground-based observations are unable to provide constraints, in situ observations will unambiguously help clarifying the question.

Another perplexing feature of metallic asteroids is their apparent low density $\sim 4\text{-}5\text{ g/cm}^3$ to be compared with that of their associated metal-rich meteorites $\sim 7.8\text{ g/cm}^3$. This implies significant porosity, comparable to that found in any other asteroid. How can this be the case? The remnant cores of differentiated bodies are not expected to have significant porosity but may still include a substantial fraction of silicates; however, this would be in contradiction with the existence of purely metallic meteorites. Clearly these questions can only be solved by in situ observations.

In situ observations of metallic asteroids will further allow studying for the first time their response to collisional impacts via the properties of the craters (size, depth, morphology). The cratering process has been investigated in the case of the silicate-rich asteroids of typical densities of $1.5\text{ to }3.5\text{ g/cm}^3$ visited by previous space missions already indicating significant differences in crater morphology between the low density (Mathilde) and higher densities (Lutetia, Vesta). **We therefore expect to discover a new type of crater morphology and thus surface response to impacts on those metallic asteroids.**

4.3 Multiple asteroid systems

Following the in situ discovery of the Dactyl satellite of asteroid Ida during the Galileo fly by, 83 multiple systems have been compiled in the main belt. The vast majority is binary (primary + one satellite) but four are triple (primary + two satellites). It is currently estimated that $\sim 5\%$ of all main belt asteroids are multiple systems. Those systems present an outstanding potential for tackling new science questions. First, the orbit of the satellite(s) allows to directly determining the density of the primary. Second, their physical characterization sheds light on the overall process of satellite formation (probably post-collisional accretion as rubble piles) and their gravitational evolution (synchronous rotation? tidally controlled shape?)

No space mission has ever studied in detail a binary or even triple asteroid system. Visiting such a system therefore appears as a top priority for future missions.

5) The first investigation of the internal structure of asteroids

5.1 Interior of primordial planetesimals: fingerprint of the time of formation

Very little is known about the early thermal evolution and internal structure of asteroids. Meteorite measurements have provided some constraints (evidence of aqueous metamorphism, dehydration, thermal

metamorphism, incipient melting and widespread melting with core formation) whereas density measurements performed either in situ or from the ground have provided the remaining elements of answer that are summarized below.

- These measurements indicate that a significant fraction (>95%) of the smaller asteroids ($D < \sim 60$ km) are under dense (Carry 2012): their density is lower than that of their surface composition. These under-dense bodies have been interpreted as being pervaded by large cracks and voids in their interiors, resulting from cataclysmic impacts and subsequent uneven re-accumulation of material. The fraction of volume occupied by these voids is called *macroporosity*. Our current census of density and macroporosity for about 300 asteroids indicates that some asteroids may have macroporosities up to 50%.

- Density measurements also indicate that large asteroids ($D > \sim 100$ km) are far less porous (macroporosity <10%) than smaller ones; when coupled with the fact that large asteroids are prominently primordial protoplanets, this implies that their **internal compositional structure holds invaluable clues on the earliest stages of their formation. By studying these objects, we can learn whether they are differentiated or not. Such information, in turn, holds invaluable constraints on their time of formation that are otherwise only accessible from meteorite measurements (i.e., differentiated bodies formed earlier than undifferentiated ones). In case of differentiation, we can learn about their internal compositional structure such as the thickness and nature of the various layers.** Of direct interest to the present topic, several researchers have recently shown that the distinction between primitive bodies and differentiated bodies might not be as simple as once thought. A paleofield in the primitive chondritic Allende meteorite has been detected, which is probably due to the presence of a dynamo, and therefore an iron core, in its parent-body (Carpözen et al. 2011). This discovery not only blurs one of the fundamental frontiers in the solar system, that between chondrites and differentiated bodies, but also forces us to question whether our knowledge of asteroid surfaces is representative of their bulk.

Probing for the first time the internal structure of several asteroids will represent a giant step forward and specifically allow:

- deciphering the early thermal evolution of primordial asteroids,
- demonstrating whether all large asteroids are differentiated,
- understanding how the internal structure varies as a function of the surface composition,
- completing the chronology of the time of formation of the different asteroid classes that currently relies on meteorite measurements only,
- characterizing the internal effects of collisions (as of today, we have only studied the surface effects of collisions),
- explaining the origin of the mysterious grooves and crater chains.

Extensive groove systems have been observed on the Martian moon Phobos, and on asteroids Lutetia and Vesta whereas crater chains have been observed on asteroid Steins. It is currently proposed that the presence of voids beneath the surface could explain their origin.

5.2 Which asteroid sizes should be targeted?

Similarly to other solar system objects (including the Earth), main belt asteroids have been heavily affected by collisions, evidenced by i) the identification of several dynamical families (~ 40 ; a family is produced when a large asteroid undergoes a catastrophic collision, leaving behind numerous fragments with similar proper orbital elements.), ii) the presence of several binary or even triple systems ($\sim 5\%$ of the total population), iii) the heavily cratered surfaces of the asteroids that have been visited by space missions, iv) the low densities of most asteroids, and v) the existence of many meteorites breccias in our collections (it is now well established that most recovered meteorites are asteroid fragments).

These numerous collisions explain the shape of the size frequency distribution (SFD) of the population which shows an excess near diameters of ~ 100 km. This bump suggests that most asteroids with diameters $< \sim 100$ km are collision-evolved fragments (larger versions of meteorites), whereas larger objects - ~ 150 objects with $D > 100$ km - are primordial planetesimals (possibly the smallest surviving protoplanets;

Morbidelli et al. 2009).

This implies that only the largest objects ($D > 100\text{km}$) offer the possibility to investigate and constrain the earliest stages of planetesimal formation, as the internal structure of rubble piles rather reflects the collisional evolution. **Whereas the surfaces of primordial ($D > 100\text{km}$) have suffered the effect of collisional erosion, their interiors still reflect their pristine formation either differentiated or not.**

5.3 Which asteroid types should be targeted?

To date, the interior of asteroids has never been the subject of in depth exploration via in situ measurements. This therefore leaves us with **complete flexibility** concerning the target selection. However, we discuss below the asteroid types that we think would be the most promising for this investigation.

i) Any C-complex object would be of interest. Their densities suggest that water ice could be present in their interior. Note that it would be particularly interesting to visit the parent body of CVs as these meteorites show a strong magnetic remanence whose origin is likely primordial, suggesting the presence of an early core dynamo – thus iron core - within its parent body.

ii) The Nice model predicts T/D type asteroids to be – as Jupiter Trojans – implanted TNOs (Levison et al. 2009). If this is true, we should find a considerable amount of water ice in their interior.

iii) As highlighted in section 4.2, the (porous) metallic asteroids (X-types) appear very intriguing and should be explored in situ.

iv) Asteroids that are currently unsampled by our meteorite collections (see 4.1) in order to collect a first order estimate of their time of formation.

v) S type asteroids should be avoided as they have already been visited several times by space missions. In addition, their meteoritic analogs (OCs) indicate that they did not differentiate but only underwent thermal metamorphism. They appear therefore far less interesting than the parent bodies of some other meteoritic candidates (e.g., CV).

6) The origin of water on Earth

6.1 Asteroid and comets as plausible sources

For decades, the source of Earth volatiles has been a matter of intense debate. This topic is not only important in order to understand the origin of life on our planet, but also because it holds crucial clues on the early evolution of the solar system.

If the solar system was dynamically quiet early on, both planets and small bodies that formed inside the snow line were likely born ‘dry’ (the Earth in particular) and volatiles would have been accreted at a later stage of Earth’s evolution through impacts of volatile-rich asteroids and/or comets (Owen & Bar-Nun 1995). [The snowline corresponds to the inner boundary of the water ice condensing region: beyond this limit, the solar nebula was cool enough so that volatiles condensed in icy grains, which were then accreted into planetesimals. This process occurs at a temperature in the range from 145 to 170 K depending on the partial pressure of water vapor. In the optically thin solar nebula, the snowline is estimated to be located at 2.7 AU]. On the contrary, if migration processes have been ubiquitous right after planetesimal formation (e.g., Grand Tack, Walsh et al. 2011), the Earth may have accreted ‘wet’. In the former case, the accretion of only a few C-type asteroids would have been sufficient to import the oceanic mass on Earth (Albarede 2009, Alexander et al. 2012). Whichever scenario is considered, the volatile composition of the Earth is governed by the volatile inventory of planetesimals, which were accreted by the planet.

When searching for the origin of Earth’s water, and thus looking at the distribution of water across the solar system, it appears that water is present among all classes of planetesimals. Water has been detected on Kuiper Belt Objects (KBOs), on the moons of all giant planets, and on comets (Takir & Emery 2012). Closer to the Sun, it has also been detected on main belt asteroids in various forms (Rivkin and Emery, 2010; Campins et al., 2010; Licandro et al., 2011). Specifically, hydrated minerals such as phyllosilicates have been identified both at the surface of asteroids, and in meteorites studied in the laboratory. These hydrated minerals can contain structurally bound OH or H₂O (some carbonaceous chondrites can contain up to 12% in water, see Takir & Emery 2012 for a review).

Water has also been detected on some asteroids as ice (see Fig 5). Themis and Cybele for example, exhibit a spectral feature around 3 microns, which has been attributed to the presence of water ice at their surface (Rivkin & Emery, 2010, Campins et al. 2010, Licandro et al. 2011). Jewitt & Guilbert-Lepoutre (2012) suggested a scenario in which repeated impacts could steadily bring buried ice at their surface (provided water ice could have survived up to now).

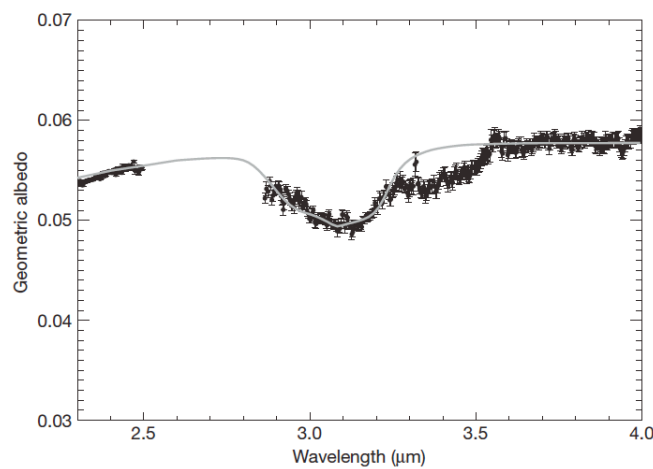


Figure 5: Detection of water ice on the surface of asteroid 24 Themis (Rivkin and Emery 2010). The grey curve corresponds to a model of water ice.

Since water is ubiquitous among all classes of planetesimals, its detection per se on a given object does not specify which population(s) of small bodies contributed to Earth’s water. An additional criterion is thus required to help discriminating among all plausible sources of water, namely the deuterium-to-hydrogen (D/H) ratio as discussed below.

6.2 From the D/H ratio to models of the solar system formation

Because of its potential for constraining the origin of volatiles on Earth, the measurement of the deuterium-to-hydrogen (D/H) ratio in small bodies and meteorites has been the subject of intense efforts. Deuterium was synthesized during the Big Bang (Wagoner et al 1967) and is essentially believed to be primordial, since there is no known mechanism that produces significant amounts of D in galaxies or stars thereafter (Epstein et al 1976). The D/H ratio in water is very sensitive to the conditions prevailing in the environment in which it is formed. Isotopic exchange reactions occurred with an efficiency that depends upon the turbulent mixing in the solar nebula, correlated with gas density and temperature. This ratio is thus predicted to vary with heliocentric distance (see Robert 2006 for a review) and/or the time of formation. Therefore, objects that formed in the same source regions and at similar times should have accreted ice with similar hydrogen isotopic compositions. This implies that a comparison of water D/H values in comets and other solar system small bodies is potentially a direct test of the predictions of the dynamical models (e.g.,

Nice model or Grand Tack model, see section 3).

Figure 6 summarizes the present state of knowledge of the D/H ratio among small bodies of the solar system. The marked difference between meteorites-asteroids and comets, that is between internal and external solar system objects, has long been assumed an essential property of the solar system, mainly based on the higher D/H ratio of comets (Robert 2006). This dichotomy has been questioned by the measurement of the Jupiter family comet Hartley 2 whose D/H is identical to the terrestrial value and therefore closer to that of carbonaceous chondrites (Hartog et al. 2011). This discovery strengthens the idea of a continuum between asteroids and comets (Gounelle et al. 2011).

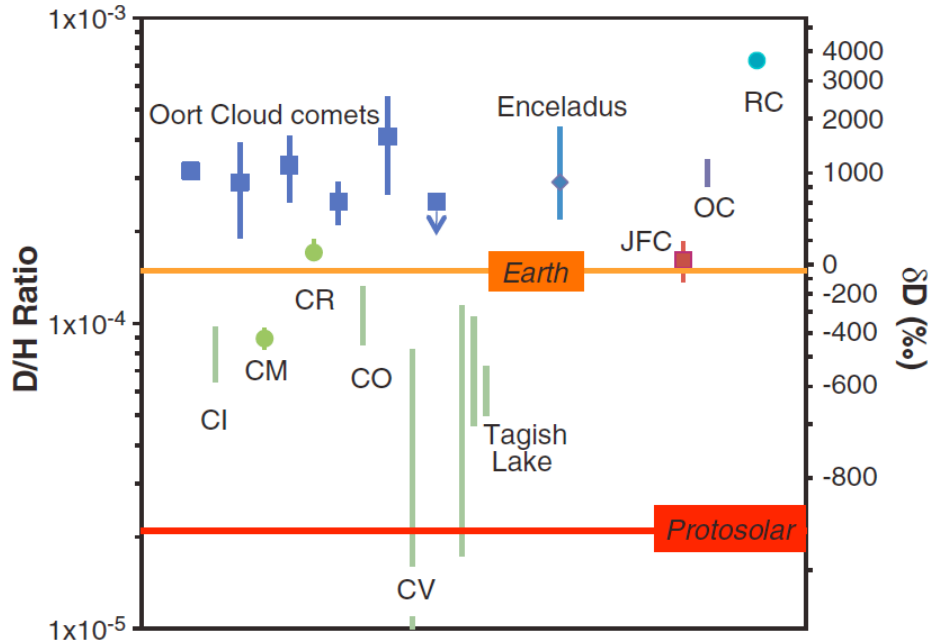


Figure 6: Comparison of the estimated D/H ratio in various chondrite groups with those measured in Oort cloud and Jupiter family comets (JFC), and Saturn's icy moon Enceladus (from Alexander et al., 2012).

Overall, figure 6 highlights that determining the origin of the water on Earth is a problem that is far more complex than solely constraining the D/H ratio for all classes of small bodies, although it is a necessary step. Indeed, the Earth cannot have formed solely by accretion of Jupiter family comets, as its bulk composition is not compatible with solely comets. It thus appears that constraining the origin of the water on Earth will also require constraining the isotopic composition for all classes of small bodies. Whereas this data partially exists via meteorites, it is obviously incomplete as certain asteroid classes are not present in our collections. **It is thus mandatory to extend the reconnaissance of both the D/H ratios and the isotopic composition for all classes of small bodies that are not sampled in our collections (see section 4.1). We should therefore avoid visiting the parent bodies of CI and CM meteorites (Ch and Cgh types) and primarily focus on water-rich B-, C-, Cb, Cg, T- or D-types.**

Ultimately, knowing both the D/H ratio and the isotopic composition of all classes of small bodies will allow reproducing the composition of the Earth by identifying the correct linear mixture of end members. At the same time, they will allow identifying the source of its water.

7) Mission profile and instruments

Our three top level science themes, diversity among asteroids, internal structure and origin of water require a mission profile allowing visiting a substantial number of asteroids satisfying the following general conditions. Large asteroids ($D > 100\text{km}$) are highly favored since they are expected to be primordial and possibly differentiated (as opposed to highly processed rubble pile asteroids). The targets must be of different spectral types and furthermore other than the S type since four of them have been visited by previous space missions. Multiple systems are also favored as they allow enhanced science at no extra cost (e.g., density of the primary and properties of the satellites). Orbiting each target is further required to allow detailed studies not possible during a fly by, with advanced instrumentation not flown so far on asteroid missions such as radar and magnetometer to address new science questions related in particular to the internal structure of the objects, along with more traditional instruments, such as cameras and spectrometers. Finally, a lander is needed for in-situ measurements on a water-rich asteroid not possible from an orbit. A strawman mission concept is described below starting first with the instrument payload of both the main spacecraft and of the lander, then the mission profile and finally the flight system.

7.1 Instrument payload description for Main spacecraft (MSC)

Visible imaging

Visible imaging will be performed by a complementary set of two cameras, narrow-angle (NAC) and wide-angle (WAC). The science objectives include the bulk characterization of the asteroids (size, shape, rotational properties), the surface geology via high resolution multispectral ($0.4 - 0.9\ \mu\text{m}$) imaging and topographic reconstruction, and identification of the landing site (via the construction of a DTM).

Near-infrared mapping spectrometer

The NIR mapping spectrometer will characterize the chemical and mineralogical compositions of the surface of the target asteroids then allowing inferring mineralogical processes and chemical evolution. The typical wavelength range of 0.75 to $4.4\ \mu\text{m}$ covers the signature of prominent mineralogical components.

Thermal emission spectrometer

The science objective of the MIR spectrometer is to produce mineral and thermal emission spectral maps of the surface of the target asteroids complementing the NIR spectrometer. The typical wavelength range is 6 to $30\ \mu\text{m}$.

Ultraviolet spectrometer

The science objective of the UV spectrograph is to examine the surface composition of the asteroids and search for outgassing products in the spectral passband $115\text{-}600\ \text{nm}$.

Gamma-ray spectrometer

The science objective of the gamma-ray spectrometer is to determine the atomic composition of the asteroid surface. It will provide whole asteroid maps of various elements such as H, C, O, Mg, Si, S, K, Ca, Fe, Ni, and Ti as well as measurements of any radioactive elements (K, Th, U).

Neutron spectrometer

The neutron spectrometer will determine the average atomic mass of the asteroids and is very effective in detecting ice (hydrogen) near the surface of the asteroids. It will in particular provide information about Fe, Ti, C, and possibly Gd and Sm through thermal neutrons, and average atomic mass from fast neutrons. Epithermal neutrons provide a nearly unique measure of H abundances, particularly so for low H abundances. Further, the combination of thermal, epithermal, and fast neutrons along with gamma-rays can provide constraints on the burial depth of a hydrated layer covered by a relative dry layer within the top $50\ \text{cm}$ of soil.

Laser altimeter

The laser altimeter will provide range information between the MSC and the surface of the targets and a mapping of the surface topography. Together with the radio science experiment (see below), it will give an estimate of their mass. It will also provide information on the gravity field and the position of the centre of mass.

Radio science experiment (RSE)

The INSIDER mission will rely on a dual X/Ka band telecommunication system for navigation, telemetry, and radio science. By using this system to measure the Doppler signal of the two downlink radio carrier frequencies resulting from the perturbation of the spacecraft motion during multiple orbits, mass, gravity field and possibly low order gravity harmonics of all targets will be determined to high precision.

Radar

The radar will identify and locally characterize subsurface structure (monolith/rubble pile) and stratigraphy by obtaining sounding profiles down to possibly a few kilometres at relatively high vertical resolution. Additional goals will be achieved by combining results from other instruments, in particular the laser altimeter.

Magnetometer

The magnetometer will detect and characterise any permanent intrinsic (fossil) magnetic field of the target asteroids with implications for their deep interior. It will measure the three-axis magnetic field components.

7.2 Lander

The Lander suite of instruments experiments will study the composition and structure of a water-rich asteroid and are designed to i) measure the elemental, molecular, mineralogical, and isotopic composition of the target's surface and subsurface material, ii) measure the near-surface strength, density, texture, porosity, ice phases and thermal properties, iii) image the target from panorama to microscopic surface texture of individual grains. The Lander will also carry a drilling device allowing collecting samples at various depths and distributing them for analysis. Typically, the Lander payload will include: an Alpha Proton X-ray Spectrometer (APXS) to determine the chemical composition, an ion laser mass analyser to perform molecular, isotopic and elemental analysis of the surface for geochemical characterization, one or several gas analysers to determine the elemental, molecular and isotopic composition of ices, a thermogravimeter to monitor the possible cometary activity and measure the volatile content in the regolith, a set of sensors to measure the mechanical, thermal, electrical and acoustic surface and subsurface properties, and a panoramic, close-up and microscopic imaging system. Additional lander payload could include a mid-infrared spectrometer, and a Raman microscope.

Considerable expertise and heritage exist within Europe for both the MSC and Lander instruments with only a few exceptions. In the case of the Lander, the proposed mission will capitalize and valorise the considerable investment put in the Philae Lander of the Rosetta mission. New developments to improve performances and miniaturization are expected in the coming years in the framework of new missions, in particular JUICE which shares many of the science objectives for the investigation of several satellites of Jupiter.

7.3 Mission

In addition to science requirements, operational constraints are clearly a critical aspect of the mission as multiple target rendez vous, orbit insertion and de-orbiting directly translate into a delta-V requirement. Selecting targets on close, low inclination orbits is therefore highly desirable. As a first exercise, we have identified four objects which nominally meet the above criteria, see Table below. A few comments are warranted: i) water has been detected on 24 Themis, hence a potential candidate for D/H in-situ measurement; ii) 10 Hygea is a good analog to Ceres, a probable dwarf planet; iii) 87 Sylvia is a triple system and considered to be an analog of Trojan asteroids (thus appears as a plausible implanted TNO); iv) 16 Psyche is the largest metallic asteroid.

Object	Type	Diameter (km)	Semi major axis	Inclination	Comment
24 Themis	B/C	200	3.14 AU	0.75°	Water
16 Psyche	X/Xk	250	2.92 AU	3.1°	Metal
10 Hygiea	C	400	3.14 AU	3.84°	Water
87 Sylvia	X/T	260	3.5 AU	11°	Triple

7.4 Flight system

The flight system for this mission will consist of a main spacecraft and a lander (possibly including a small rover). The spacecraft will incorporate all functions required to meet the science objectives, including communication functions with Earth and with the lander, manoeuvres, orbit insertion, lander delivery, stable pointing for the science payload, and powering of all systems. The spacecraft should accommodate a scientific payload of approximately 100/150 kg plus a 100kg-class Philae-type lander to be dropped on 24 Themis. The lander itself should accommodate a scientific payload of approximately 30 kg.

The spacecraft propulsion will ensure successive rendez vous, orbit insertion and de-orbiting for each target asteroid. Energetic considerations show that the trajectory, assuming a Dawn-like strategy with only a Mars Gravity Assist, requires about 8.2 km/s of delta-V in chemical propulsion and 13.5 km/s of delta-V if essentially achieved with electrical propulsion. Note that without Sylvia, those figures drop respectively down to 5.3 km/s and 8.5 km/s due to the large inclination of this object. In the case of a chemical-propelled spacecraft, its launch mass can be decreased by the addition of Earth and/or Venus gravity assists. In the case of electrical propulsion, a complementary chemical propulsion capacity would not only open the door to Earth gravity assist(s) so as to decrease the launched mass but would also allow for some impulsional capture into orbit around the different targets thus relaxing the need to slowly adjust the orbital parameters and reducing the waiting time in between the targets.

The presence of chemical propulsion on board is desirable also for the options that involve hovering on Themis before dropping the Lander. An additional 0.3 km/s of delta-V is then needed for descent close to the surface, hovering and subsequent return to orbit. If Themis' density is about 2.8 g/cm³, the gravity is about 0.008g thus low enough to allow for hovering without requiring detrimental propellant consumption, with a continuous thrust required to cancel Themis' gravity in the order of 80N per ton of spacecraft wet mass. ESA-led studies for the Phootprint Phobos Sample Return Mission have showed that a drop height larger than 20m will be associated to an acceptable contamination by the thrusters of the soil to be sampled. With a 20m drop height, under a 0.008g gravity, the touchdown conditions for the 100kg-class lander will be manageable as touchdown velocities should remain below 2m/s (similar to dropping it from less than 20cm under 1g). Landing shock absorbers will be desirable.

The propulsion type(s) and their sizing will have to be traded as a function of the launch dates.

Solar cells will be enough to power the spacecraft. The European industry has the capacity to produce large solar arrays in suitable configurations, while the ESA-led development of Low Illumination Low Temperature cells for the Jupiter environment of JUICE will cover the needs for the distances lower than 4AU of this asteroid belt mission. The most demanding case surface-wise for the Solar Array will be for the electrical propulsion option. Operating one Dawn-like engine at a time at about 3AU in addition to the rest of spacecraft consumption would require about 60 m² of Solar Array, well within the capacity of the European Industry.

The attitude control when in cruise or orbit and the RF communications do not present specificities with respect to other European interplanetary missions, either past or in development. In particular, the needs of JUICE will cover those of this mission in that respect.

The descent and hovering on Themis require some degree of autonomous control with vision-based/LIDAR-based navigation in a way similar to the Osiris-REX, Marco-Polo or Phootprint missions.

All in all, when comparing the required delta-V and operational functionalities to the ones already studied for other missions to the asteroid belt or to Phobos, the spacecraft class is expected to be JUICE-like but somewhat lighter, with a dry mass in the order of the ton (Lander not included) and propellant quantities that will depend upon the optimum mix of propulsion types and upon the number of gravity assists, bringing the wet mass within Ariane 5 capacities.

8) References

- Albarede. *Nature* 461, 1227-1233 (2009).
Alexander. *Science* 337, 721-723 (2012).
Alibert et al. *Astrophys. J* 417, L25–L28 (2004).
Alibert et al. *Astrophys. J.* 626, L57–L60 (2005).
Bottke et al. *Nature* 439, 821-824 (2006).
Bus & Binzel. *Icarus* 158, 146-177 (2002).
Campins et al. *Nature* 464, 1320- (2010).
Carporen et al. *PNAS* 108, 6386-6389 (2011).
Carry. *PSS* 73, 98-118 (2012).
Chambers. *EPSL* 223, 241-252 (2004).
Chambers. *Icarus* 180, 496–513 (2006).
Ciesla. *Science* 318, 613- (2007).
DeMeo. PHD thesis, Paris Observatory (2010).
Emery et al. *The Astronomical Journal* 141, id. 25 (2011).
Epstein et al. *Science* 263, 198- (1976).
Gomes et al. *Nature* 435, 466-469 (2005).
Gradie & Tedesco. *Science* 216, 1405-1407 (1982).
Hartogh et al. *Nature* 478, 218- (2011).
Jewitt & Guilbert-Lepoutre. *AJ* 143, 21- (2012).
Levison et al. *Nature* 460, 364-366 (2009).
Licandro et al. *A&A* 525, 34- (2011).
Lisse et al. *Science* 313, 635-640 (2006).
Malhotra. *Astronomical Journal* 110, 420 (1995).
Marchi et al. *Science* 336, 690-694 (2012).
Morbidelli et al. *Nature* 435, 462-465 (2005).
Morbidelli et al. *Icarus* 204, 558-573 (2009).
Nakamura et al. *Science* 321, 1664-1667 (2008).
Owen & Bar-Nun. *Icarus* 116, 215 (1995).
Poncy et al. *EGU General Assembly*, 6968 (2009).
Raymond et al. *Icarus* 183, 265-282 (2006).
Raymond et al. *Astrobiology* 7, 66-84 (2007).
Rivkin & Emery. *Nature* 464, 1322- (2010).
Robert. In *Meteorites and Early Solar System* (Lauretta & McSween Eds.), 341- (2006).
Takir & Emery. *Icarus* 219, 641- (2012).
Wagoner et al. *ApJ* 148, 3- (1967).
Walsh et al. *Nature* 475, 206-209 (2011).
Zolensky et al. *Science* 314, 1735-1739 (2006).

In situ Investigations of the Local Interstellar Medium

White Paper | May 24, 2013

Spokesperson:

Robert F. Wimmer-Schweingruber
Inst. f. Exp. & Appl. Physics
University of Kiel
24118 Kiel - Germany

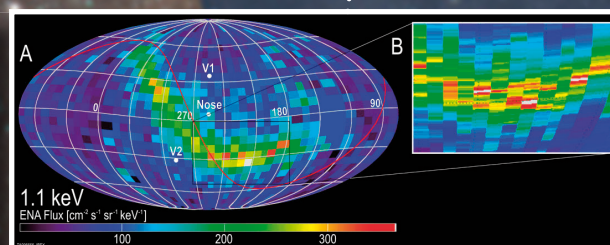
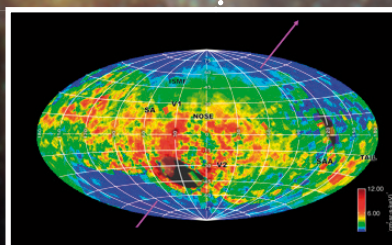
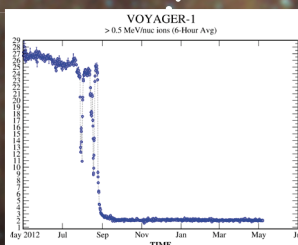
Email: wimmer@physik.uni-kiel.de
Phone: +49 431 880 3964
Fax: +49 431 880 3968

IP 200 AU

Voyager 124 AU

Cassini 9.6 AU

IBEX 1 AU



List of contributors and supporters

(Contributors are indicated in bold)

AUSTRALIA — Cairns, Iver (cairns@physics.usyd.edu.au)

GERMANY — Breitschwerdt, Dieter (breitschwerdt@astro.physik.tu-berlin.de); Burkert, Andreas (burkert@usm.lmu.de); **Dröge**, Wolfgang (Wolfgang.Droege@astro.uni-wuerzburg.de); Duschl, Wolfgang (duschl@astrophysik.uni-kiel.de); **Fichtner**, Horst (hf@tp4.rub.de); Heber, Bernd (heber@physik.uni-kiel.de); Klecker, Berndt (berndt.klecker@mpe.mpg.de); Leipold, Manfred (manfred.leipold@kayser-threde.com); Mall, Urs (mall@mps.mpg.de); Pohl, Martin (pohlmadq@gmail.com); Sasaki, Manami (sasaki@astro.uni-tuebingen.de); **Scherer**, Klaus (kls@tp4.rub.de); Semenov, Dimitry (semenov@mpia.de); Srama, Ralf (srama@irs.uni-stuttgart.de); **Steerken**, Veele (veerle.sterken@mpi-hd.mpg.de); **Wimmer-Schweingruber**, Robert F. (wimmer@physik.uni-kiel.de); Wolf, Sebastian (wolf@astrophysik.uni-kiel.de)

FINLAND — Janhunen, Pekka (pekka.janhunen@fmi.fi); Valtonen, Eino (eino.valtonen@utu.fi)

ITALY — Bruno, Roberto (roberto.bruno@ifsi.rm.cnr.it); Sorriso-Valvo, Luca (lucasorriso@gmail.com); Zimbardo, Gaetano (gaetano.zimbardo@fis.unical.it)

RUSSIA — **Baranov**, Vladimir B. (vladimir.b.baranov@gmail.com); **Chalov**, Sergey (chalov@ipmnet.ru); Izmodenov, Vladimir (vlad.izmodenov@gmail.com); Panasyuk, Mikhail (panasyuk@sinp.msu.ru)

SWITZERLAND — Wurz, Peter (peter.wurz@space.unibe.ch)

SWEDEN — Barabash, Stas (stas@irf.se)

TAIWAN — Ip, Wing-Huen (wingip@astro.ncu.edu.tw)

UK — Arridge, Chris (c.arridge@ucl.ac.uk); Bailey, Mandy (mandybailey22@gmail.com); Balikhin, Michael (m.balikhin@sheffield.ac.uk); **Barlow**, Mike (mjb@star.ucl.ac.uk); Barstow, Martin A. (mab@leicester.ac.uk);

Branduardi-Raymont, Graziella (g.branduardi-raymont@ucl.ac.uk); Clements, Dave (d.clements@imperial.ac.uk); Coates, Andrew (a.coates@ucl.ac.uk); Cowley FRS, Stanley W. H. (swhc1@ion.le.ac.uk); **Crawford**, Ian A. (i.crawford@bbk.ac.uk); Fletcher, Leigh N. (fletcher@atm.ox.ac.uk); Fossey, Steve (s.fossey@ucl.ac.uk); **Foullon**, Claire (c.foullon@exeter.ac.uk); Gao, Yang (yang.gao@surrey.ac.uk); Grande, Manuel (M.Grande@aber.ac.uk); **Horbury**, Timothy (t.horbury@imperial.ac.uk); Howarth, Ian (idh@star.ucl.ac.uk); Jackman, Caitriona (caitriona.jackman@ucl.ac.uk); Jones, Geraint H. (g.h.jones@ucl.ac.uk); Laitinen, Timo (timo.lm.laitinen@gmail.com); **Owens**, Matthew (m.j.owens@reading.ac.uk); Pope, Simon A. (s.a.pope@sheffield.ac.uk); **Quenby**, John J. (j.quenby@imperial.ac.uk); Serjeant, Stephen (Stephen.Serjeant@open.ac.uk); **Smith**, Keith Thomas (kts@ras.org.uk); van Loon, Jacco Th. (j.t.van.loon@keele.ac.uk); Ward-Thompson, Derek (dward-thompson@uclan.ac.uk)

USA — Allegrini, Frédéric (fallegrini@swri.edu); Bale, Stuart (bale@ssl.berkeley.edu); Bochslers, Peter (peter.bochsler@space.unibe.ch); Collinson, Glyn A. (glyn.a.collinson@nasa.gov); Desai, Mihir (mdesai@swri.edu); **Frisch**, Priscilla C. (frisch@oddjob.uchicago.edu); Gruntman, Mike (mikeg@usc.edu); Krimigis, Tom (Tom.Krimigis@jhuapl.edu); Mason, Glenn M. (Glenn.Mason@jhuapl.edu); McComas, David (DmcComas@swri.edu); **McNutt Jr.**, Ralph L. (ralph.mcnutt@jhuapl.edu); Mewaldt, Richard (rmewaldt@srl.caltech.edu); Möbius, Eberhard (Eberhard.Moebius@unh.edu); Opher, Merav (mopher@bu.edu); Provornikova, Elena (eprovn@bu.edu); Schwadron, Nathan A. (n.schwadron@unh.edu); Slavin, Jon (jslavin@cfa.harvard.edu); Trattner, Karlheinz (karlheinz.j.trattner.dr@lmco.com); Zank, Gary (garyp.zank@gmail.com); Zurbuchen, Thomas (thomasz@umich.edu)

SOUTH AFRICA — Burger, Adri (Adri.Burger@nwu.ac.za); Potgieter, Marius S. (Marius.Potgieter@nwu.ac.za)

Executive Summary

The discovery of a myriad of exoplanets in the past decade has revolutionized the understanding of our place in the universe. How different are they and do some of them harbor life, just like Earth? To do so, their parent star must drive a stellar wind and carve what we call an astrosphere into the surrounding interstellar medium (ISM). Astrospheres are ubiquitous in our immediate neighborhood [78] and show similar structure to our heliosphere. Voyager, IBEX, and Cassini have shown that the interaction between interstellar medium and solar wind is much more complex and involved than previously believed [53]. This stellar-interstellar interaction is key to understand the ubiquitous phenomenon of astrospheres and the shielding they provide to the planetary systems they harbor. It is only accessible to us on the outermost edges of our heliosphere where it must be probed to answer the following questions (see also Table 1):

- **H** – How do solar wind and interstellar medium interact to form the heliosphere and how does this relate to the universal phenomenon of the formation of astrospheres?
- **A** – What are the properties of the very local interstellar medium and how do they relate to the typical ISM and the material from which planetary systems are made?
- **F** – How do plasma, neutral gas, dust, waves, particles, fields, and radiation interact in extremely rarefied, turbulent, and incompletely ionized plasmas?

Scientific Relevance to Cosmic Vision 2015-2025

Thus, this science theme is highly relevant to the four Science Themes defined for the Cosmic Vision 2015-2025 programme and addresses all of them:

CV 2015-2025 Theme	Relevance
What are the conditions for planet formation and the emergence of life?	Shielding of GCR, dust, and neutrals: dust-plasma interactions
How does the solar system work?	Structure and dynamics of the heliosphere
What are the fundamental physical laws of the universe?	Fundamental plasma physics, extremely rarefied plasmas
How did the universe originate and what is it made of?	LISM composition and galactic chemical evolution

This science theme is also **timely** because the Sun is now transitioning from a Grand Solar Maximum which dominated the space age into a normal, less active state [1] with likely significant implications for the state of the heliosphere. It is time that humankind intentionally sends a probe to the stars.

Strawman Mission Concept

An interstellar probe has been studied by ESA [43] and NASA [e.g., 51 and references therein] and both agencies have shown it to be technologically feasible and challenging, and thus, to be an ideal candidate for a European-led L-class mission. The following two technological drivers would need to be addressed:

- **Propulsion:** Proposals have included solar sails, nuclear ion propulsion, electric sails, heavy launcher [see 51 for a summary].
- **Power:** Nuclear power would be unavoidable, payload power sharing strategies would be needed.

Solving both would significantly enhance European space-faring capabilities. Both are required for a reasonable mission duration.

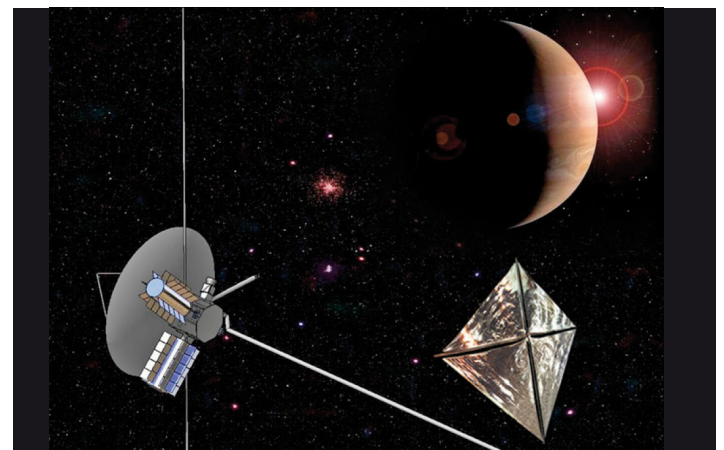


Fig. 1: Artists impression of an interstellar probe after having shed its solar sail.

Bonus Science Goals

On its way to the heliopause and beyond, the interstellar probe will allow the following bonus science goals from a variety of scientific disciplines:

- Measure extragalactic background light undisturbed by the solar system Zodiacal light.
- Determine the soft X-ray background in the heliosphere and solar-wind planet interactions.
- Constrain heliospheric dynamics by multispacecraft studies

Table 1:

Physics of the Local Interstellar Medium		
SCIENCE GOAL	SCIENCE QUESTION	REQUIRED MEASUREMENTS
Heliospheric Science (H)		
How do solar wind and interstellar medium interact to form the heliosphere and how does this relate to the universal phenomenon of the formation of astrospheres?	H1: How does the heliosphere shield against cosmic rays and neutral particles and what role does it play in the interstellar–terrestrial relations?	GCR, energetic particles, ENAs, plasma, B -field, waves
	H2: How do the magnetic field and its dynamics evolve in the outer solar system?	energetic particles, ENAs, plasma, B -field, waves
	H3: How do heliospheric structures respond to varying boundary conditions?	Plasma, B -field, ENAs, Ly-alpha
	H4: How do the boundary regions in the heliosphere modify the intensities of the various particle populations?	GCR, energetic particles, ENAs, plasma, B -field, waves, dust
	H5: How does the interstellar medium affect the outer solar system?	GCR, energetic particles, ENAs, plasma, B -field, waves, dust
Astronomy and Astrophysics (A)		
What are the properties of the very local interstellar medium and how do they relate to the typical ISM?	A1: What is state and origin of the local interstellar medium?	Charge-state and element composition, waves, B -field, Ly-alpha, ENAs
	A2: What is the composition of the local interstellar medium?	Composition
	A3: What is the interstellar spectrum of the GCR beyond the heliopause?	GCR
	A4: What are the properties of the interstellar magnetic field?	B -field, waves, plasma
	A5: What are the properties and dynamics of the interstellar neutral component?	ENAs, dust, plasma
	A6: What are the properties and dynamics of interstellar dust?	Dust, B -field, plasma
Fundamental Physics (F)		
How do plasma, neutral gas, dust, waves, particles, fields, and radiation interact in extremely rarefied, turbulent, and incompletely ionized plasmas?	F1: What is the nature of wave–particle interaction in the extremely rarefied heliospheric plasma?	Distribution functions, energetic particles
	F2: How do the multiple components contribute to the definition of the local plasma properties within the heliospheric boundary regions?	Plasma, ENAs, energetic particles, composition, waves, B -field
	F3: What processes determine the transport of charged energetic particles across a turbulent magnetic field?	Plasma, ENAs, energetic particles, composition, waves, B -field
Bonus (B)		
	<ul style="list-style-type: none"> - Extragalactic Background Light - Soft X-ray background - Multispacecraft studies 	IR/Vis wide-field imaging soft X-ray measurement time series

Introduction

After the exciting in-situ observations of the termination shock and the entry of the Voyager 1 spacecraft into the inner and possibly outer heliosheath (see Figs. 2 & 3), there is a growing awareness of the significance of the physics of the outer heliosphere. Its understanding helps to clarify the structure of our immediate interstellar neighborhood (e.g., [4]), contributes to the clarification of fundamental astrophysical processes like the acceleration of charged particles at a stellar wind termination shock (e.g., [17]) and beyond, and also sheds light on the question to what extent interstellar–terrestrial relations are important for the environment of and on the Earth [19,62] and exoplanets. In order to explore the boundary region of the heliosphere, it is necessary to send a spacecraft to perform advanced in-situ measurements particularly in the heliosheath, i.e. the region between the solar wind termination shock, and the heliopause, as well as in the (very) local interstellar medium (VLISM). Solar activity is decreasing to ‘normal values’ below those of the Grand Solar Maximum [1] which was typical of the space age so far (Fig. 8). This is likely to reduce the size of the heliosphere and allows us to study a ‘normal’ heliosphere by launching an Interstellar Probe (IP) which will also provide within a shorter time than previously believed the first comprehensive measurements of key parameters of the local interstellar environment such as its composition, state, and magnetic field. Together with an accurate determination of the state of the heliospheric plasma

across the heliosphere, these quantities are crucial to understanding how the heliosphere, and, much more generally, astrospheres, are formed and how they react to varying interstellar environments.

Our current understanding of the interstellar medium and heliosphere is undergoing dramatic changes. Today, we understand the interstellar medium as a turbulent environment with varying degrees of ionization, highly variable composition and dust-to-gas ratio interacting with a complex magnetized and highly ionized heliospheric plasma — all in a complex background field of UV, cosmic rays, and neutral particles which is modified by the interaction itself. Thus, the heliosphere and its boundary regions serve as the worlds largest laboratory for complex plasmas. This complex region strongly modulates the flux of galactic cosmic rays which account for one half of the natural background radiation that life is exposed to on Earth and shields Earth and solar system from highly reactive neutral hydrogen atoms, thus ensuring the habitability of Earth (and, in analogy, of potential life-supporting exoplanets). How does this shielding function depend on the strongly varying interstellar environment? How does this shielding depend on the solar activity-induced heliospheric structure (Fig. 3)? What is the role of (anomalous) cosmic rays in these interstellar–terrestrial relations?

The ongoing Voyager Interstellar Mission (VIM) and recent observations from the Interstellar Boundary Explorer (IBEX) [47] and Cassini missions [34] have revealed the interaction of the heliosphere with the VLISM to be much more complex than heretofore assumed. With new observations have come significant new puzzles for describing the physics of the interaction between solar (stellar) wind and the surrounding interstellar medium.

In-situ instruments on Voyager 1 and Voyager 2 up to very recently have revealed significant fluxes of energetic particles in the heliosheath, including a well-defined suprathermal ion “tail” in which the differential intensities fall off $\sim E^{-1.5}$ above ~ 30 keV [11]. At higher energies (~ 100 MeV), there is no “unfolding” of the energy spectrum of the anomalous cosmic rays (ACRs), thus pointing to a more remote location for the modulation region and source [60,64]. Most strikingly, direct measurements of the shocked solar wind flow speed obtained from Voyager 2 revealed

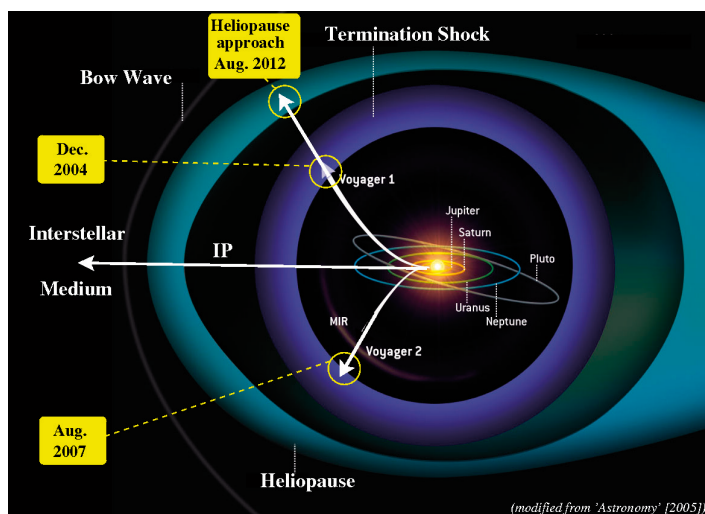


Fig 2: The positions of the Voyagers in the heliosphere. In August 2012 Voyager 2 entered a region likely to be associated with the heliopause. Neither Voyager will be able to probe the interstellar Medium, necessitating an Interstellar Probe.

that the flow remains supersonic in the heliosheath beyond the termination shock [61]. All of these particle observations, taken together, unambiguously imply that the bulk of the energy density in the plasma resides in a non-thermal component that extends to very high energies. Strong implications, both quantitative and qualitative, follow from this fact for the overall heliosheath structure. We have never encountered a large-scale plasma regime in which the non-thermal ion pressure dominates the thermal pressure and overwhelms the magnetic field stresses. The closest analog regime lies in localized regions of planetary magnetospheres during extremely disturbed conditions, but in the heliosheath these conditions always exist everywhere. Even sophisticated MHD models failed to predict anything like the striking new features that have been observed in the last few years.

There was however a foretelling of this recent revelation. Voyager 1&2 beginning in 1983 and continuing to the present had detected remarkable long-lasting radio emissions in the 1.6–3.4 kHz range that were identified with major disturbances in the heliosheath produced by giant coronal mass ejections (CMEs) [37]. The higher frequency emissions were localized, coming from an extended arc confined to the hemisphere toward the interstellar flow (i.e., the “nose” of the heliosheath), and lying close to, but not actually in the galactic plane [24]. These authors noted that the arc could perhaps be the curve on the heliopause (the boundary between shocked solar wind and interstellar plasma, see Fig. 2) where the interstellar magnetic field was normal to that surface ($\mathbf{B} \cdot \mathbf{n} = 0$), in accordance with the “hydrogen deflection plane” defined by the $\sim 4^\circ$ difference between the arrival directions of interstellar H atoms [39] (affected by charge exchange in a heliosheath deformed by the interstellar magnetic field) and the unaffected interstellar He atoms [76].

In 2009, remote sensing of the heliosheath proton population using images formed in energetic neutral atoms (ENAs) by IBEX and Cassini/INCA revealed stunningly unexpected structures on a variety of scales [46,43]. IBEX data show a relatively narrow “ribbon” of atomic hydrogen emission from ~ 200 eV to ~ 6 keV, roughly circular, but asymmetric in intensity, suggesting that it is ordered by the interstellar magnetic field (Fig. 4). It passes through, rather than being centered on, the “nose” at which the local, neutral interstellar plasma flow around the heliosphere stagnates. This suggests that the flow is not the primary driver of the system as has been thought, but rather it is the pressure of

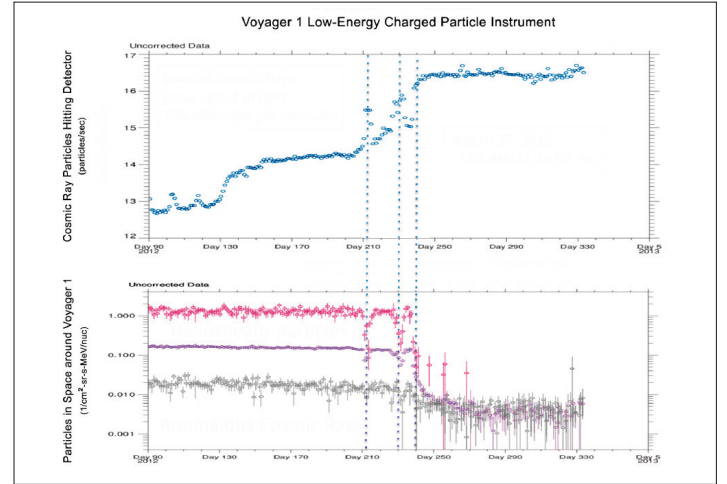


Fig. 3: Voyager-1 measurements of the flux of GCR (top panel) and low-energy particles (bottom panel) show dramatic changes in both particle populations around the end of August 2012, indicating that Voyager 1 may have detected the heliopause or its precursor.

the interstellar field that configures the heliosheath. The neutrals from both the glow and ribbon are also characterized by non-thermal distribution functions. The Ion and Neutral Camera (INCA) on Cassini sees at higher energies (10s of keV) a “belt” of emission in ENAs, broader than the ribbon and tilted significantly away from it and exhibiting a much steeper energy spectrum than observed in the IBEX energy range [35] (Fig. 5).

More recently, particle anisotropy measurements by the Low Energy Charged Particle (LECP) instrument on Voyager 1 suggested that the spacecraft had entered a heliosheath transition layer. The negligible flow velocity of the in situ particles suggested proximity to the heliopause [36]. In fact, very recent measurements

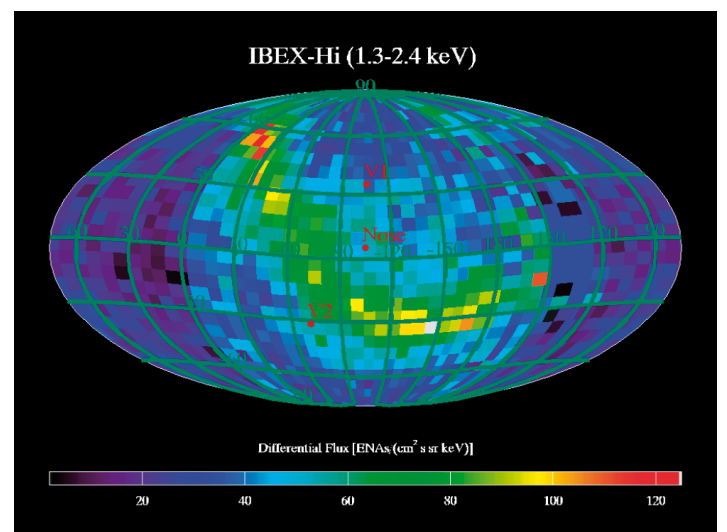


Fig. 4: IBEX map of energetic neutral hydrogen atoms (ENAs) from 1.3–2.4 keV shows the ‘ribbon’ and has the nose and Voyager 1 & 2 positions indicated.

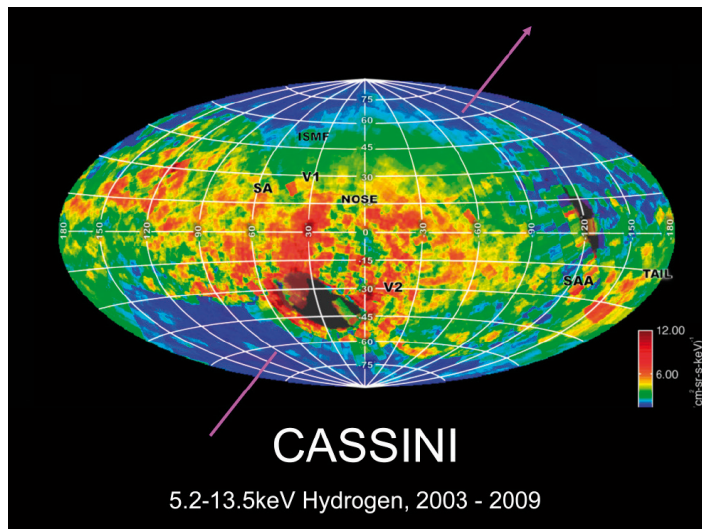


Fig. 5: ENA map from the INCA instrument on Cassini. The map is an equal-area projection that shows the emission “belt.” The nominal “nose” of the heliosphere from which there is a general flow of neutral atoms is indicated along with the outgoing asymptotic trajectories of Voyager 1 (V1) to the north and Voyager 2 (V2) to the south, respectively, of the plane of the ecliptic.

shown in Fig. 3 indicate that Voyager 1 has entered a new regime and may have detected or even passed the heliopause. The rise in GCR shown in the top panel illustrates the shielding provided by the heliosphere.

Attempts to explain consistently all the afore-mentioned fascinating observations are currently roiled in controversy, with no clear trend towards a consensus. All the diverse in situ and remote observations obtained to date only serve to emphasize the need for a new generation of the more comprehensive measurements that will be required to understand the global nature of our Sun’s interaction with the local galactic environment. Only an interstellar probe with

modern instruments and measurement requirements better defined by these recent observations can provide the new information required.

The interstellar medium is the primeval material which the Sun, the planets, and ultimately terrestrial life were made of some 4.6 billion years ago, just as many other stars and planetary systems were formed at other times in different places. Exploring our local interstellar neighborhood will vastly enhance our understanding of the origin, formation, and evolution of stars, their planetary systems, and possibly of life. Thus, this grand science theme addresses the following three core Science goals (see Tab. 1):

Heliospheric Science – H

- How do solar wind and interstellar medium interact to form the heliosphere and how does this relate to the universal phenomenon of the formation of astrospheres? (H1-H5)

Astronomy and Astrophysics – A

- What are the properties of the very local interstellar medium and how do they relate to the typical ISM and the material from which planetary systems are made? (A1-A6)

Fundamental Physics – F

- How do plasma, neutral gas, dust, waves, particles, fields, and radiation interact in extremely rarefied, turbulent, and incompletely ionized plasmas? (F1-F3)

The three Science Goals mentioned above can be broken down into more detailed questions which illustrate the breadth and importance of the overall Science Theme, as shown in Table 1.

Science Objectives

H – How do solar wind and interstellar medium interact to form the heliosphere and how does this relate to the universal phenomenon of the formation of astrospheres?

Remarkably, the better we understand the physical processes at work on our Sun, the more we view our Sun as a typical stellar object. The processes that give rise to our solar wind are clearly at work at other stars. We are beginning to understand not only how the Sun

heats its corona and powers the solar wind, but how these processes relate quite generally to stellar coronae and winds. The heliosphere which is inflated by the solar wind is the direct analog to astrospheres inflated by the stellar winds of other stars.

- H1:** How does the heliosphere shield against cosmic rays and neutral particles and what role does it play in the interstellar–terrestrial relations?

Cosmic rays are high-energy charged particles which bombard Earth from above the atmosphere. Several thousand pass through a person’s body every minute.

These can cause biological damage but also cause mutations which accelerate evolution. The majority of GCRs present in interstellar space are shielded out by the outer heliosphere (Fig. 6), presumably via a strong magnetic barrier that forms in the inner and outer heliosheath, where the solar wind slows down and is deflected by the interstellar flow (See, e.g., [18], but see discussion below of the relevant physics which was quite unexpected.). Figure 6 shows the differential flux of GCRs from beyond the heliosphere to inside the heliosphere at 1 AU. A small fraction of GCRs penetrate into the heliosphere and propagate toward the Sun and planets. These residual GCRs are modulated by the solar wind's magnetic field in the inner heliosphere.

What we know about the dominant shielding of GCRs in the inner heliosheath region is very limited and based mostly on models and theory. It is nonetheless clear that the solar wind must slow down prior to meeting the interstellar flow. This slowdown must result in a strong pile-up of magnetic field since the magnetic field is frozen in to the solar wind. This magnetic barrier is believed to be the primary shield against GCRs entering the inner heliosphere (e.g., [18]), although some additional modulation in the outer heliosheath appears to be needed [64]. Recent data from Voyager 1 show a dramatic drop in low-energy particles associated with a strong rise of GCR particles [73], providing further illustration of the complexity of the physics responsible for the shielding/modulation by the heliosphere.

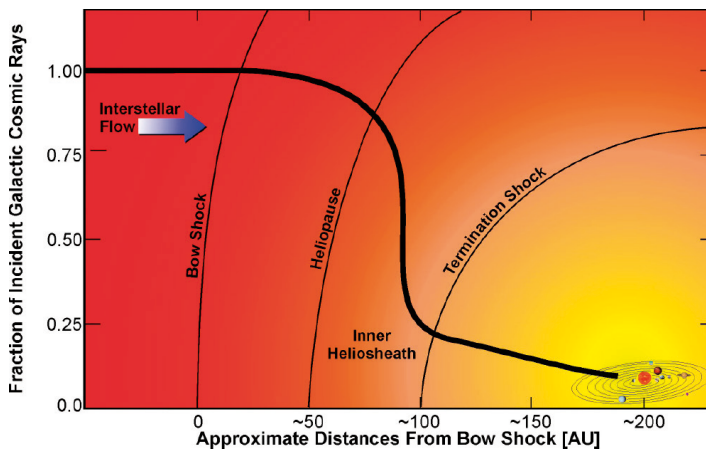


Fig. 6: The fraction of incident GCRs on the heliosphere is most strongly reduced in the inner heliosheath where the slowdown of solar wind creates a large magnetic barrier to GCRs; this barrier is the dominant shield against GCR radiation in the solar system.

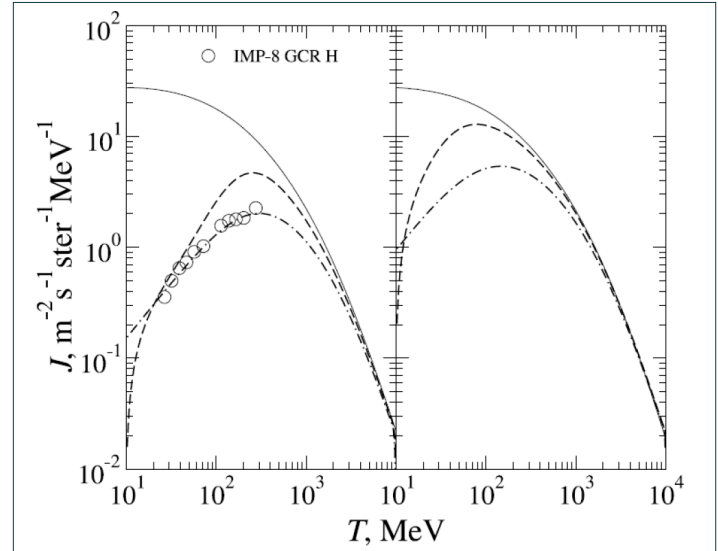


Fig. 7: Galactic cosmic rays differential intensities in the heliosphere during the present day, left, and a future or past period when the heliosphere was smaller, with termination shock near 20 AU, due to a larger ($\times 10$) density in the interstellar medium [18].

Large changes in the LISM have dramatic effects on the heliosphere and the radiation environment of the solar system. For example, a typical enhancement in the density of the local interstellar medium by a factor of 10 causes the entire heliosphere to shrink to about a quarter of its current size [81], and increases the fluxes of GCRs at Earth by a factor of 2–6 [63]. Such large changes in the LISM have certainly occurred in the past and will occur again in the future [81].

Figure 7 shows the differential intensity of GCR protons, on the left for the present day, and on the right for a period when the heliosphere was smaller due to a larger density ($\times 10$) of the local interstellar medium. Shown are external boundary conditions [49], conditions near the termination shock (dashed), and near Earth (dashed-dotted). Circles show IMP-8 data [28]. The large increase in the levels of GCR radiation (right panel) reveals the critical influence of local interstellar conditions on the radiation environment of the solar system. The estimations made in Fig. 7 are purely theoretical. We do not currently have the observational knowledge required to understand how the local interstellar medium interacts with the heliosphere; observations of that global interaction are essential for understanding the radiation environment experienced by astronauts on long missions to distant destinations, such as Mars.

On Earth, the radioisotope ^{10}Be provides a recent record of cosmic ray fluxes (Fig. 8). It is produced in Earth's upper atmosphere by spallation reactions of

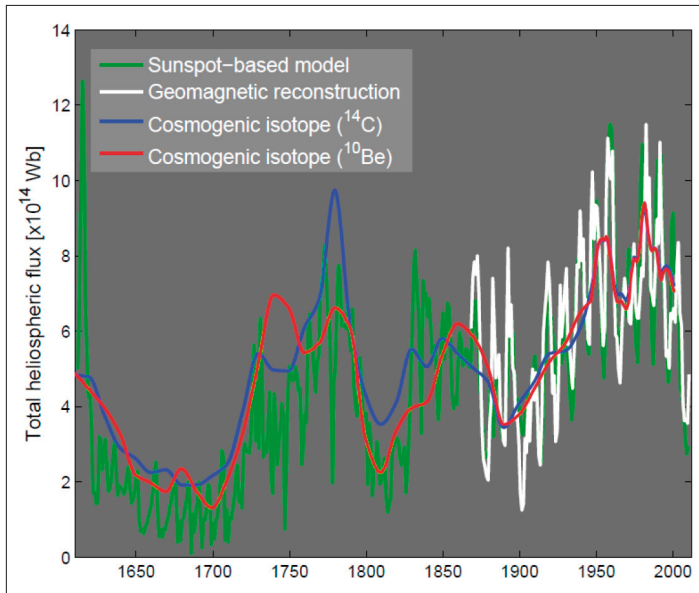


Figure 8: The total heliospheric magnetic flux over the last 400 years, as estimated by geomagnetic, cosmogenic isotope and sunspot number reconstructions. The space-age has been a period of anomalously high solar activity, which is currently drawing to a close. Adapted from [56].

cosmic rays (CR) protons ($E \geq 100$ MeV) and secondary neutrons with atmospheric nitrogen and oxygen. ^{10}Be records in Antarctic ice show two prominent peaks 35,000 and 60,000 years ago, when the radioisotope production rate was about twice the current value for about 1500 and 2000 years, respectively, which has been interpreted as due to supernovae in the vicinity of the heliosphere [59]. Could the GCR fluxes in the heliosphere change rapidly in the future due to changing conditions in the LISM? Again, the ^{10}Be record from ice cores can be used to show that at least in the past 300 years this has not been the case [2]. Nevertheless, because of the critical hazard posed by GCR radiation, future manned space travel will rely heavily on a better understand of the LISM's influence over the heliosphere, and the potential short and long-term changes to the radiation environment.

Figure 8 also shows a peculiarity which has only recently begun to be realized and appreciated, namely that the space age has been one of rather high solar activity [56]. How does the heliospheric modulation react to the changes in solar activity?

H2: How do the magnetic field and its dynamics evolve in the outer solar system?

Figure 9 shows Voyager measurements of the expected compression of magnetic field at the termination shock. However, no sector boundaries

were observed in the heliosheath during the first few months after the shock encounter, which could only be interpreted as due to a much lower convection speed (~ 17 km/s) of the local plasma relative to the spacecraft than expected. Further evidence for a significantly altered magnetic field in the downstream region comes from its fluctuations, which are much stronger in the heliosheath than in the heliosphere. Moreover, the statistical distribution of field magnitude changed [6] from lognormal (upstream) to Gaussian (downstream heliosheath), a transition that is not understood. This abrupt change in the nature of the magnetic field across the termination shock has important consequences for the acceleration of particles at the termination shock, as these are affected by turbulent motions of the surrounding plasma. The level of low-frequency turbulence in resonance with the high-energy particles accelerated at the termination shock is unknown but key to understanding the modulation of galactic cosmic rays and the acceleration of anomalous cosmic rays.

H3: How do heliospheric structures respond to varying boundary conditions?

Observations by SWAN on SOHO have shown that the magnetic field in the VLISM lies at a significant angle to the galactic plane (Fig. 10) [39], a result recently independently confirmed with Voyager radio data [55]. On the other hand, general considerations about a galactic dynamo suggest that it should lie in the galactic plane at least on large scales. Thus, the very local field lies at a significant angle to the large scale field which is interpreted as a consequence of turbulent motions in the local interstellar cloud. Furthermore, the overabundance of carbon (see Science Objective

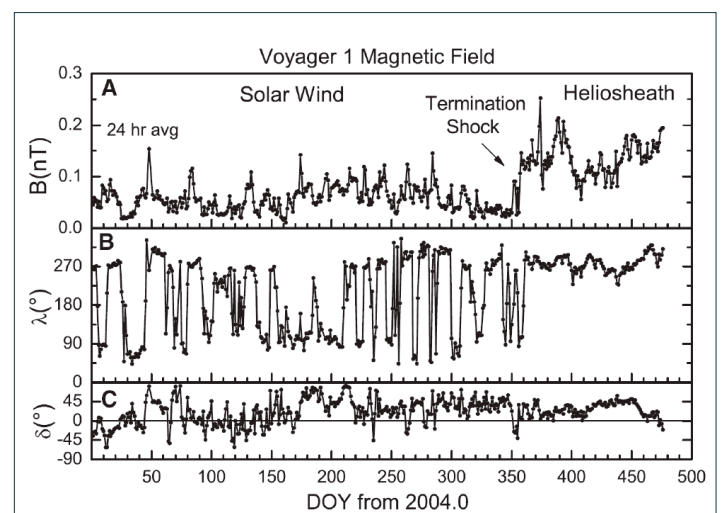


Fig. 9: Compression of the magnetic field across the termination shock [6].

A2) indicates an inhomogeneous local cloud. Together with observations of differences in flow angles, these observations imply an unexpected variability in the immediate interstellar vicinity of the heliosphere. Thus, we may expect that the heliosphere must react to these varying interstellar boundary conditions as well as to the solar-cycle variations at the inner boundary condition, the Sun (Fig. 8).

Based on modeling efforts we expect that several heliospheric structures will react quite sensitively to changes in the interstellar medium [13]. Density fluctuations in the hydrogen wall should propagate around the heliosphere and thus give us a record of past variations in the heliosphere's very local interstellar neighborhood. The three-dimensional structure of the hydrogen density surrounding the heliosphere can be measured, thus giving us access to this archive.

Because we do not know the strength of the interstellar magnetic field, we do not know whether the heliosphere has a bow shock, although there are strong indications from IBEX, that there is none [48]. The presence of a bow shock has important consequences for the turbulence in the outer heliosheath, i.e., between bow shock and the heliopause. The shock generates downstream turbulence that translates into locally decreased spatial diffusion of energetic particles, thus contributing to a shielding against galactic cosmic rays [64, who also show that there is substantial modulation beyond the heliopause.].

Furthermore, the trajectories of interstellar dust particles are altered by a bow shock. Thus the presence of a bow shock can be determined by a surprisingly simple measurement of the inflow direction of interstellar dust particles in a given mass range. Simulations [44] show that the flow direction of small particles is deflected by approximately 10° from the undisturbed direction when a sharp bow shock is present. Assuming that the inflow direction of gas and dust is the same, a measurement of the dust flow direction thus gives us the possibility to remotely detect the presence of a bow shock and, hence, indirectly determine a lower limit on the magnitude of the interstellar magnetic field.

Figure 8 illustrates the timeliness of investigating heliospheric response now. There are strong indications that the Grand Solar Maximum (GSM) is coming to an end and that the Sun is transitioning into an extended period of reduced activity [1]. This has two important implications. First, we expect to see a different heliosphere in the coming decades due to the changing inner boundary conditions. Second, reduced solar activity will likely result in a smaller heliosphere

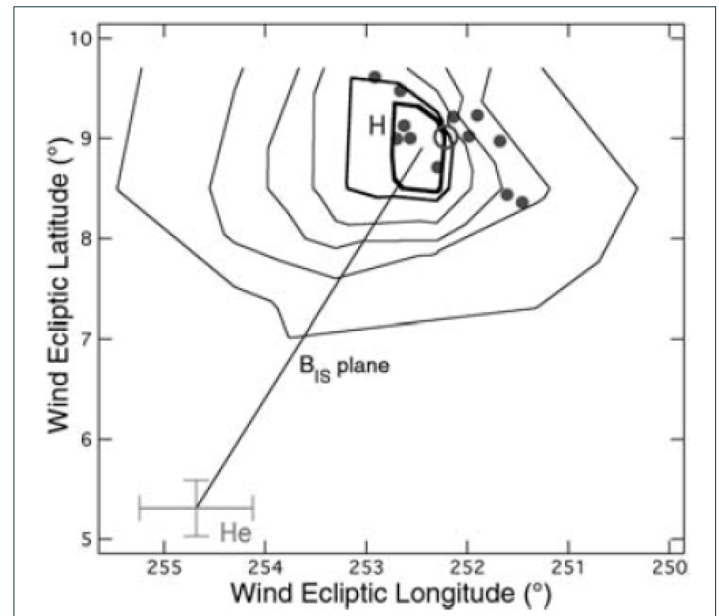


Fig. 10: Observations with SOHO/SWAN indicate that the direction of the very local magnetic field is deflected from the average galactic plane direction by turbulent motions in the local interstellar cloud. From [39]

which would allow an interstellar probe to reach the LISM sooner than previously believed.

H4: How do the boundary regions in the heliosphere modify the intensities of the various particle populations?

Early cosmic ray observers discovered an unusual subset of cosmic rays which consisted of singly ionized ions (instead of fully stripped nuclei) with energies of 1–50 MeV/nuc [54]. They were called Anomalous Cosmic Rays (ACRs). Most of the ACRs are species which have high ionization thresholds, such as He, N, O, Ne, and Ar. Until recently, ACRs were thought to arise only from neutral atoms in the interstellar medium [16] that drift freely into the heliosphere through a process that has four essential steps: first, the neutral particles stream into the heliosphere; second, they are converted into ions, called pickup ions since they are picked up and swept out by the solar wind; third, pickup ions are pre-accelerated by shocks and waves in the solar wind (see also [66]); and finally, they are accelerated to their final energies at the termination shock [57] or beyond it. Easily ionized elements such as C, Si, and Fe are expected to be strongly depleted in ACRs since such elements are not neutral in the interstellar medium and therefore cannot drift into the heliosphere.

Today, we are able to detect pickup ions directly, as well as unusual components of the ACR [e.g., 9].

There is a growing understanding that, in addition to the traditional interstellar source, grains produce pickup ions throughout the heliosphere: grains near the Sun produce an “inner source” of pickup ions, and grains from the Kuiper Belt provide an “outer” source of pickup ions and anomalous cosmic rays (see, e.g., [65], and references therein.).

Not only are recent observations calling into question the sources of ACRs, but also the very means by which they are accelerated. The prevailing theory until V1 crossed the TS was that pickup ions were energized at the TS to the 10–100 MeV energies observed [57]. However, when V1 crossed the TS, it did not see a peak in the ACR intensity as the aforementioned theory predicted [50, 71]. Instead the ACR intensities continued to increase in the heliosheath. Various suggestions have been proposed [e.g., 5, 6, 45, 50], but so far none has been able to explain all aspects of these puzzling observations. The dramatic fall in the intensity of the ACR shown in the lower panel of Fig. 3 promises valuable information on both their propagation and the boundary layer properties if investigated by a future probe.

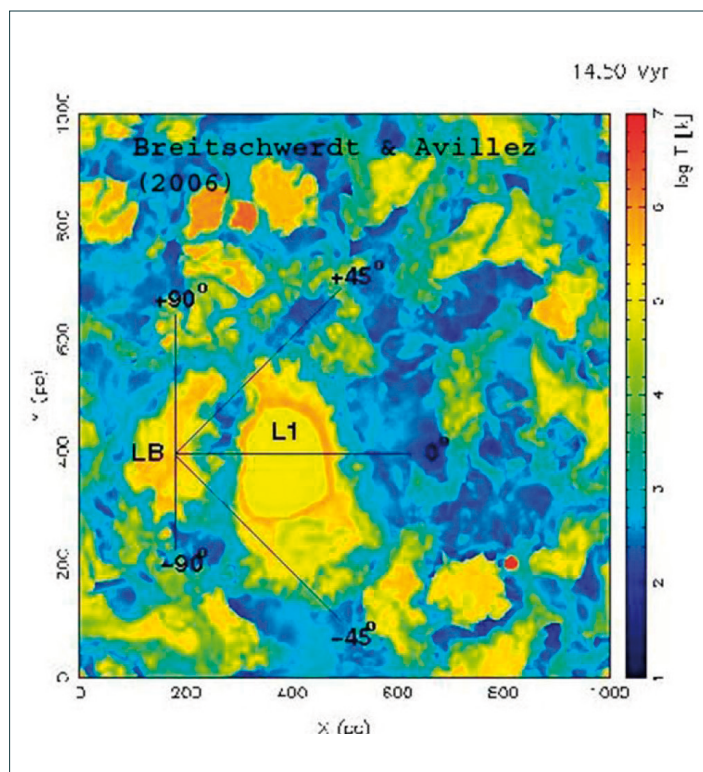


Fig. 11: The present day temperature distribution and extension of the Local Bubble (labeled LB) and the Loop 1 superbubble (L1) in a section through the galactic midplane about 14.5 million years after their origin. The solar system is located at the intersection of the various lines-of-sight (solid lines in the Figure) in the LB [4].

H5: How does the interstellar medium affect the outer solar system?

Interstellar dust entering the heliosphere interacts with the small planetary objects that are located beyond the orbits of the giant planets of our solar system. This region is believed to consist of remnants of planetesimals that were formed in the protoplanetary disk and studying the small objects in this trans-Neptunian regions is of basic interest for comparing the solar system to extra-solar planetary systems. The flux of the interstellar dust is considered as a source of dust production by impact erosion in this trans-Neptunian region [80, 62] and also limits the lifetime of the outer solar system dust cloud. Observations of the zodiacal dust itself can provide unique insights not only into the history and content of our solar system, but also provide a detailed template that can be used to understand the exo Zodiacal dust in other solar systems.

The Science Questions discussed so far show clearly that we need to understand better our immediate local interstellar neighborhood and naturally lead to the following Science Goal.

A – What are the properties of the very local interstellar medium and how do they relate to the typical ISM?

A1: What is state and origin of the local interstellar medium?

The Local Interstellar Cloud (LIC) belongs to a flow of low-density ISM embedded in the very low density and hot ($T \sim 10^6$ K) Local Bubble (LB, see Fig. 11). The bulk motion of this cluster of interstellar clouds points toward the center of the Loop 1 super-bubble (L1). Within this overall flow, distinct cloudlets have been identified with unique velocities. The motional direction of the cloud currently feeding interstellar gas into the heliosphere has been determined with the GAS experiment on Ulysses [75] and, interestingly, is not aligned with the overall motion—it appears to be 1.5 km/s slower than the observed ISM velocity towards α -Cen. This suggests that the heliosphere is at or close to the edge of the LIC and, thus, the material surrounding the heliosphere could change on time scales as short as the duration of IP (see, e.g., [19], for a review). Studies of the orientation of the local interstellar field also appear to indicate the importance of a highly turbulent interstellar flow

(see Science Objective A4). The associated timescales are comparable to the present duration of the space age and our understanding of the importance of the heliosphere in shielding us from the interstellar medium. For instance, neutron-monitors, first introduced in 1957 with the International Geophysical Year, have shown that the galactic cosmic ray intensity at Earth varies with solar activity. Galactic cosmic rays produce the important climate tracer ^{14}C by spallation of nitrogen in the Earth's atmosphere. We currently base much of our modeling efforts for climate physics on uncertain understanding of the relation between GCR-produced ^{14}C and solar activity based on historic records of sunspots (Fig. 8). Given that one of the time scales of the variability of the interstellar boundary conditions is roughly the same as the time scale as the neutron monitor data or maybe the sunspot record, the question naturally arises whether the naive assumption that the modulation of GCR by the heliosphere is only determined by solar activity may not be overly simplified. Heliospheric structure and modulation is determined by time-varying boundary conditions at the Sun and in the local interstellar medium.

A2: What is the composition of the local interstellar medium?

Compositional studies have established themselves as an extremely successful tool to understand the origin and evolution of astronomical and solar system bodies. Based on studies of the solar system, we believe that the central star and its planets are made of the same material with only small compositional gradients in similarly behaving elements across the planetary system (if any at all). The driving fractionation processes are condensation and heating. Similar studies of galactic composition and its evolution are hampered by these often neglected but important processes. Frequently, the composition of the ISM can only be determined in the gas phase using, e.g., absorption lines. The missing elements (relative to a “universal” galactic composition, derived from solar composition) are then thought to be locked into interstellar dust grains. The composition of dust is very hard to measure remotely, some progress has been made using measurements of extinction, polarization and emissivity over a wide range in wavelengths. However, the effects of space weathering on individual dust particles is hardly understood and accounted for. Thus, it is safe to say that the composition of the interstellar medium is only understood in a qualitative way. The only accessible

interstellar cloud is the local cloud, and, hence, we need to measure its composition in the dust, gas, and plasma phase. A key ingredient in this respect is the dust-to-gas mass ratio which is different when measured in the LIC and in-situ in the heliosphere. Radiation pressure, solar gravity, and Lorentz forces modify the flux of the dust into the solar system and the acting forces vary with the dust properties as well as with the plasma and magnetic field conditions (see, e.g., [20, 23, 67] for reviews). As a result, both the dust fluxes in the interstellar medium and in the outer solar system, and, hence, the corresponding dust-to-gas mass ratios, are estimated with great uncertainty. The small particles, which probably make up the majority of the dust number density, are deflected at the boundaries and inside the heliosphere [10,25,40,70]. Measuring the time dependence of their flux gives important information on the boundaries and on the properties of interstellar dust.

A key measurement is the abundance of certain abundant elements in the VLISM and comparison with measured abundances of interstellar ions (in the form of pickup ions) and atoms (in the form of neutral gas) within the heliosphere. Understanding the filtration effects on various elements will allow us to generalize them to other elements and thus to finally derive the elemental abundances in the very local interstellar medium from in-situ measurements within the heliosphere. The measurement of the abundances of elements in the LIC can only be done if we can measure the ionization state of hydrogen (or of oxygen (or N) because it readily charge exchanges with H). This is the most prominent hurdle in establishing the metallicity of the LIC (Fig. 12).

This becomes even more important if we want to compare the local interstellar composition with that of the solar system. Intriguingly, we observe that the Sun (and solar system) appear to be isotopically heavier than the interstellar medium at a similar galactocentric distance. This is currently the only indication that the solar system must have migrated several kiloparsec within its galactic environment. In other words, studying the differences between solar system and galactic abundances is the only opportunity we have to quantitatively assess the effects of galactic dynamics.

A further puzzle is the carbon abundance of the LIC. In interstellar space the C abundance is a factor of about 2.5 below solar abundances in the gas phase, and, as discussed above, the missing carbon is thought to be locked up in interstellar dust grains or giant molecules consisting of PAHs (Polycyclic

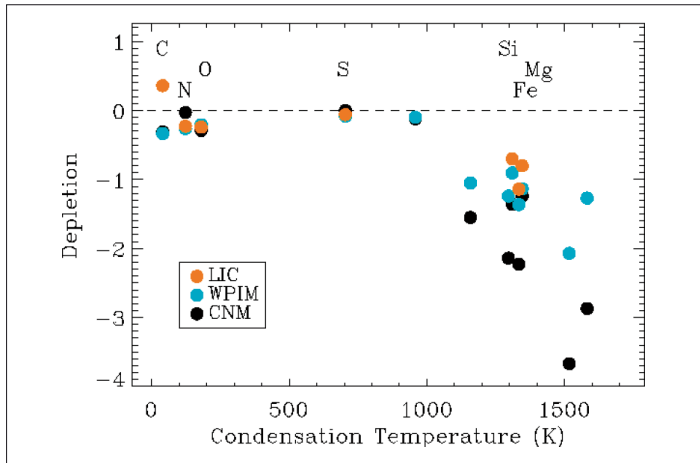


Fig. 12: Compositional patterns for the local interstellar environment. Orange symbols show LIC composition from model 26 of [69], cyan symbols show warm partially ionized matter, and black symbols show cold neutral material. The differences show the importance of measuring all the phases (plasma, gas, and dust) of the interstellar medium.

aromatic hydrocarbon). In the LIC, C appears to be significantly overabundant in the gas phase for reasons not understood [68]. This appears to indicate not only total destruction of carbonaceous dust grains locally, but also inhomogeneous mixing of gas and dust within the cloud, which in turn has consequences for the nature of turbulent mixing in the LISM. Moreover, as carbon is a direct pre-requisite for life as we know it, this intriguing puzzle deserves more attention. Direct measurements of singly-ionized and the small expected amount of doubly-ionized carbon, as well as the dust composition, will shed light on the life-cycle of carbon in the Milky Way.

A3: What is the interstellar spectrum of the GCR beyond the heliopause?

The GCR is believed to originate in particles accelerated at supernova-driven shock fronts. These shocks likely accelerate surrounding material, dust, gas, and plasma particles. Thus, GCRs offer a unique way to sample the composition of the galaxy and to understand the energetics of supernova shock expansion. Current modeling efforts show large variations in the possible interstellar spectrum [26,27,64,72]. One of the difficulties in these studies is the influence of the heliosphere which modifies the GCR spectrum as measured at the Earth. Tremendous gains in the understanding of the above topics could be made if we knew the undisturbed interstellar spectrum. This would allow us to understand and accurately model the filtering effect of the heliosphere and, hence, to much

more accurately interpret the information brought to us by galactic cosmic rays. IP will be able to address this question by measuring the unfolding of the GCR spectrum up to 100–300 MeV/nuc between the outer heliosphere and the ISM.

A4: What are the properties of the interstellar magnetic field?

Observations with SOHO/SWAN [39] as well as Voyager radio observations [55] indicate that the magnetic field (likely frozen into the interstellar medium as it also is in the solar wind) does not lie in the galactic plane as would be expected on large scales, but is distorted by the turbulence present in the LIC [24,55]. The direction, strength, and variability of the interstellar magnetic field are key to understanding the overall asymmetric structure of the heliosphere. Current modeling efforts are severely limited by the uncertain knowledge of the interstellar magnetic field and its influence on the heliosphere. The magnetic field strongly influences the flow of charged particles (and, through charge exchange, of neutral particles) and anisotropies of energetic particles and does so on time scales given by the level of interstellar turbulence. The latter is important for the propagation of galactic cosmic rays and for the properties of a number of astrophysical objects. Part of the variability may also be explained by reconnection of the heliospheric and interstellar magnetic field, a fundamental process in astrophysics. Thus, understanding and modeling of the heliosphere, its shielding effects, etc. remain severely limited because the strength of the local interstellar magnetic field is unknown.

A5: What are the properties and dynamics of the interstellar neutral component?

There is overwhelming evidence from the analysis of interstellar absorption lines for the existence of a hydrogen wall ahead of the heliosphere [22,41,78]. Such structures have been observed around other stars [29,78] as have been bow shocks, indicating that our heliosphere is not unique but rather a typical example of an astrosphere forming around wind-driving stars. However, we do not know the properties of the neutral component beyond the heliopause, yet alone understand sufficiently its dynamics in the hydrogen wall and interstellar medium.

So far, the aging Voyager spacecraft have provided some direct information on the plasma environment in

the outer heliosphere. However, Voyager 1 (at 124 AU) has now passed into a region of very low fluxes and can no longer provide this information. Moreover, the state of the neutral gas is unknown, and no observations will be available beyond ~ 140 AU, when the power supply on board the Voyagers will become insufficient. Ulysses has measured neutral interstellar gas directly out to ~ 5 AU, and first observations of energetic neutral atoms (ENAs) are confirming their likely production in the heliosheath [21,79], but, confusingly, also beyond [47]. These pioneering measurements are now routinely performed by the Interstellar Boundary Explorer, IBEX [47]. On the other hand, IBEX will not provide us with measurements beyond this region, especially within the hydrogen wall. These must be performed by IP, thus providing us with a detailed understanding of the interstellar neutral component.

A6: What are the properties and dynamics of interstellar dust?

Understanding the nature of the interstellar medium and its interaction with the solar system includes the dust properties in the outer solar system and in the interstellar medium. Moreover improving our knowledge of interstellar dust properties and quantifying the dust to gas ratio in the interstellar medium is of fundamental astrophysical interest, e.g., in star and planetary formation, and galactic evolution. The Ulysses data allowed constraining models of the local interstellar medium physics as well as of interstellar dust [20]. The measurements within the solar system provide valuable information, but they improve our understanding of the interstellar dust only within certain limits and important parameters like the size distribution of interstellar dust and the dust-to-gas mass ratio can not be measured within the solar system (see A2).

The Science Questions discussed above all point to some fundamental issues which affect the physics of the interstellar medium and lead to the questions discussed in the following paragraphs.

F – How do plasma, neutral gas, dust, waves, particles, fields, and radiation interact in extremely rarefied, turbulent, and incompletely ionized plasmas?

Our understanding of the physics of complex interstellar plasmas is extremely limited. At least

part of the problem lies in the multiple components constituting the interstellar medium which all contribute similarly to, e.g., the pressure in the LISM and the heliospheric boundary region.

F1: What is the nature of wave–particle interaction in the extremely rarefied heliospheric plasma?

As discussed in question H4, the pre-acceleration of the anomalous component is incompletely understood. Why do ACR not peak at the termination shock? Obviously, the magnetic structure in this interface region plays a major role, as does the detailed wave–particle interaction in this turbulent region. While the spectra at higher energies can be modeled fairly accurately with a combination of first order diffusive (shock) acceleration and second-order (stochastic) Fermi acceleration, together with limited adiabatic heating, the injection problem at lower energies still remains unsolved. Here, detailed measurements of magnetic field variations and distribution functions of suprathermal particles, especially below ~ 40 keV/nuc, are key to understanding this problem which, of course, is not limited to particle injection and acceleration in the heliosphere, but must occur at all astrophysical shocks.

F2: How do the multiple components contribute to the definition of the local plasma properties within the heliospheric boundary regions?

Several contributors are about equally important contributors to the pressure in the interstellar medium. GCR, thermal plasma, pickup-ions, magnetic field, but especially the non-thermal particle populations [11, 61] are key players in determining the complex properties of the heliosheath. The influence of the non-thermal population is occasionally observed during highly disturbed situations in planetary magnetospheres, but the other contributions are unique to the outer heliosphere and can only be measured in situ.

F3: What processes determine the transport of charged energetic particles across a turbulent magnetic field?

Impulsive solar particle events have long been observed at longitudes which appeared to be badly magnetically connected, indicating perpendicular

transport, implying action of coronal shocks, or a considerably more complicated magnetic configuration than generally assumed. Similarly, Ulysses observations of recurrent energetic particle enhancements at much higher latitudes than the accelerating corotating interaction regions (CIRs) appear to imply perpendicular transport or a more complicated heliospheric magnetic configuration that connects CIRs to high latitudes (see, e.g., [15], and references therein). Similarly, again, detailed observations of low-energy particle distribution functions in CIRs near Earth were best explained by substantially enhanced transport of particles perpendicular to the magnetic field [12]. Intriguingly, similar phenomena have proved to be extremely puzzling in the fusion community. Cross-field transport is a limiting factor in magnetically confined fusion. IP would strongly constrain models for perpendicular transport in the outer heliosphere and, for the first time, measure it in the ISM.

B – Bonus Science Goals

Moving from close to the Sun to far-flung regions in the solar system has the potential of enabling science otherwise not possible. We present three examples of bonus science goals which could be achieved by sending an Interstellar Probe far from the Sun.

B1: Soft-X-Ray Background

ROSAT observations of the soft X-ray (0.25 keV) background were initially interpreted as a signature of hot (10^6 K) plasma filling the Local Bubble [38, and references therein]. However, this was difficult to reconcile with the observed abundance of O VI ions [7]. The soft X-ray background is contaminated by foreground emission from solar-wind charge exchange (SWCX) reactions. When highly energetic metals in the solar wind encounter neutral hydrogen and helium (of either solar or interstellar origin) they can exchange an electron, leaving the metals in a highly excited state from which they relax by emitting X-ray photons. This process has been unambiguously observed in comets [42], and is thought to occur in the boundaries of the heliosphere. However, the location of the emission and the relative contribution to the background are currently unknown. Best current observations indicate that SWCX is responsible for anything between 20% and 100% of the soft X-ray background [8,33,38, and references therein]. Attempts to separate the contributions using spectroscopy have been

unsuccessful [8], and are unlikely to solve the problem. However, measurements of the X-ray background will allow a) an understanding of a fundamental physical process occurring in the heliosphere (SWCX); b) determination of the properties of the material filling the Local Bubble; c) measurement of the local X-ray ionization rate, which plays a crucial role the heating and chemistry of the ISM [77]. Carrying this type of instrument on an interstellar probe would also allow to study the interaction between solar wind and solar system bodies in X-rays.

B2: Extragalactic Background Light

The extragalactic background light (EBL) is made up of the redshifted emission from the first stars, protogalaxies and supermassive black holes to form in the universe. Accurate measurements of the flux and spectrum of the EBL can provide key constraints on the objects responsible for reionising the universe, as well as on models of galaxy and AGN evolution. However, sufficiently accurate measurements of the EBL in the optical/near-IR are not possible from the Earth or the inner solar system because they are dominated by the foreground emission from the Zodiacal Dust. A mission travelling beyond the orbit of Jupiter will escape this dominant foreground and can make observations of sufficient accuracy to provide these uniquely powerful cosmological observations. This study would require addition of a wide field optical/near-IR imager [3] and would make observations during the journey to the heliopause, once the spacecraft is beyond the zodiacal dust.

B3: Multi-spacecraft studies

Recurring alignments between Earth, other assets throughout the solar system, and an interstellar probe will allow us to constrain the dynamics of heliospheric phenomena using a combination of measurements and modeling. Potential planetary flybys, e.g., of Jupiter, at the time, e.g., of JUICE, will allow collaborative science and augment their science return.

Strawman Mission Concept

Top-Level Mission Requirements

- M1)** Spacecraft to arrive within a $\sim 25^\circ$ cone of the heliospheric 'nose' ($+7^\circ$, 252° Earth ecliptic coordinates) or a similarly interesting region (based on IBEX results). This aims at the scientifically most compelling region and also minimizes travel time.
- M2)** Provide data from 5 AU to at least 200 AU.
- M3)** Arrive at 200 AU 'as fast as possible', ideally within 25-30 years.

A large variety of solutions exist and have been demonstrated to be feasible to achieve these mission requirements, see, e.g., [53] and references therein. Possible mission designs always rely on nuclear power and some propulsion system to achieve high escape velocities of several AU per year. Not all can achieve the short travel times envisaged in M3, thus requiring that scientific and technical mission know-how be maintained over multiple decades. Ulysses, SOHO, and the Voyagers and Pioneers have shown that this is important and can be achieved by careful mission management. Requirement M2 is an important contributor to the scientific success of the mission because it keeps the community 'alive' and excited about heliospheric physics and science.

Mission design

Voyager 1 at $\sim 3.6 \text{ AU/yr}$ is currently the fastest object to ever leave the solar system. An interstellar probe should be at least twice as fast, resulting in a primary mission duration of ~ 28 years. While this places strong reliability requirements on spacecraft and payload, countless scientific missions have outlasted their design durations and shown that this can be done.

Propulsion

Getting to 200 AU within ideally 25 to 30 years requires very high speeds on the order of 10 AU per year (after a significant acceleration time). Several options have been proposed and studied, some of which are briefly summarized here to show that solutions do exist and are either ready to be implemented or close to being tested.

Solar Sails

ESA has studied a baseline mission in which a solar-sail spacecraft was launched from Earth with $C_3 \sim 0$. The spacecraft would approach the Sun to within 0.25 AU where solar sails are highly effective. Through two such 'photon assists' in the inner solar system, an escape velocity approaching 10-11 AU can be achieved. The large sail would be jettisoned at ~ 5 AU because no significant acceleration is obtained from it anymore. Thus, the science phase could begin after this initial acceleration phase after approximately 7 years. This mission design is attractive because it achieves very high speeds and an early beginning of the science phase. Because high speeds are already acquired very early, it requires no additional gravity assists at the outer planets and therefore has launch windows repeat every year. Thus, a solar sail implementation has many advantages. In addition, implementing such an approach would establish European leadership in this important and highly enabling propulsion technology. There is also significant know-how and interest in solar sails in European industry. A difficulty lies in the availability of ultra-thin solar sail material and deploying the large sail needed for this mission. Moreover, having to go close to the Sun for two photon assists is non trivial and adds considerable mass to the sailcraft for thermal control. While Helios, BepiColombo, and Solar Orbiter all show that going close to the Sun is achievable and that this tough task should not be underestimated.

Nuclear Electric Ion Propulsion

After the great success of ESA's Smart-1 mission, electric ion propulsion is also a good candidate for a long-duration space mission. While Smart-1 relied on solar generators, an interstellar mission would need to use nuclear power, e.g., radioisotope thermoelectric generators (RTGs) or some kind of 'next-generation' Stirling radioisotope generator (SRG). Both would need to provide a relatively high power output of at least 8W/kg (i.e., a specific mass of 125 kg/kW). This approach would also require a high excess escape energy, $C_3 \geq 100 \text{ km}^2/\text{s}^2$, followed by a long period of electric propulsion of at least 15 years. However, this solution would also be very flexible in allowing many possible gravity assists at the outer planets, especially

Jupiter [14]. An optimal gravity assist at Jupiter can result in a Δv of 28 km/s, other outer planets provide somewhat less, but a combination with Jupiter can result in similar values for an additional Δv . The orbital period of Jupiter means that there are such opportunities only every 12 years, but several less optimal ones repeat every 13 months around them. Fig. 8 in [14] shows five opportunities with trip times to 200 AU of less than 30 years for launches around 2014 (too early for this proposal, yes, but similar enough to be an allowable analogy). Jupiter gravity assists pose the difficulty of the intense radiation experienced by the spacecraft when inside Jupiter's radiation belts. This would require additional shielding and could add extra mass to the spacecraft. Nevertheless, with ESA now preparing the implementation of JUICE, this problem should be well understood and under control. In fact, a flyby of an interstellar precursor mission during the scientific phase of JUICE could potentially add to the scientific impact of both missions.

Heavy launcher

Most mission implementations would profit from a substantial excess escape energy ($C_3 \geq 100 \text{ km}^2/\text{s}^2$) which can only be achieved with a heavy launcher. While the solar sail study forfeited this advantage in view of the large acceleration offered by the photon assists, all other low-thrust implementations would have to rely on a substantial boost at the beginning of the mission. Heavy launchers include the Ariane 5, Atlas V 551, Ares V, or a Falcon 9 or Falcon Heavy, of which the last two are probably the most cost-effective (see Space-X's web site for a quoted price of 128M\$ for a 2012 launch of more than 6.4 tons to GTO (accessed May 10, 2013)). All would launch the probe and an upper 'kick-stage' to provide the extra excess energy. [51] show that a heavy launcher with nuclear electric ion propulsion (and possibly a Jupiter gravity assist) is probably the lowest risk option to get an interstellar probe out to 200 AU within 25–30 years.

Venus and Earth gravity assists

Venus and Earth gravity assists would certainly aid in achieving high escape speeds early on in the mission. However, these are generally not studied in more detail because they would add complexity to the mission. A Venus flyby would increase the mass of the thermal control system because the spacecraft would have to cope with about two solar constants there, but

has to be designed for large distances from the Sun. An Earth flyby could be even more effective, but the political uncertainties of using this technique with a nuclear powered spacecraft add risk.

Electric sails

Electric sail would use the pressure exerted by the solar wind on an electrically charged 'wire-sailcraft'. The penetration distance of the high voltage tether's electric field at 1 AU is about 10 times the Debye length of 10 m, i.e., about 100m, so wires or wire structures do not need to be space-filling to present the solar wind with a large cross section. Of course, the solar wind carries with it much less momentum than solar photons, but it is also much easier to deploy thin wires and charge them to high voltage [30,31,58]. A Cube-Sat demonstrator mission (ESTCube-1) is currently being undertaken by Estonia (Pekka Janhunen, PI) and will demonstrate opening a 10 m tether in orbit and measure the Coulomb drag force acting on it. ESTCube-1 was successfully launched from Kourou on May 6th, 2013. Tether deployment is expected this summer. Initial estimates scaling this concept to an interstellar probe show that it could reach 200 AU within 25 years [58].

Power

Because an interstellar probe necessarily needs to travel far from the Sun, only nuclear power is a realistic option. Several studies have already been performed on new, next-generation power systems, mainly in the US. ESA's study [32] and previous proposals have assumed a specific power of at least 8W/kg. This is not unrealistic and is considered the design minimum

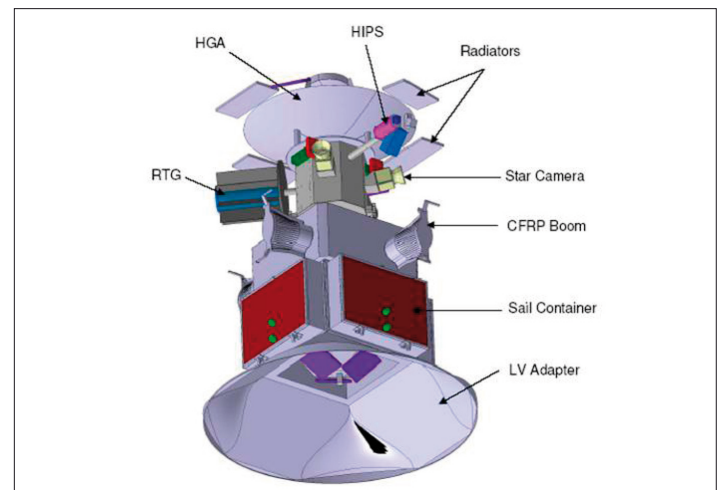


Fig. 13: Sailcraft in stowed configuration with launch adapter [32, 74].

for nuclear power sources under development [51]. A realistic nuclear electric ion propulsion system would require ~1 kW of electric power. Accounting for the ^{238}Pu half life of ~88 years, this leaves ample power

Item	Mass [kg]	Power [W]
platform	170	70
Propulsion system	200	70/900 (sail/electric)
S/C adaptor	45	
payload	25	40
margin	90	20
Launch	530	200/1030

for payload and telemetry during the science mission which would follow the cruise phase.

Spacecraft Bus

Depending on the exact mission implementation, especially on the choice of propulsion system, the spacecraft bus will have to be optimized for the specific requirements. Some key drivers may be the thermal control system, radiation shielding, etc. A recently proposed mission design came up with a spacecraft bus similar to the one given in the table below [e.g., 74], while electric propulsion would require higher mass [51]

mainly because of the substantial power requirements and the mass (>400 kg) of the Xenon fuel.

An example for a solar-sailcraft in its stowed configuration is shown in Fig. 13 [32, 74].

Payload

Most mission studies have considered the same baseline payload, similar to that summarized in Table 2. It is driven by the measurement requirements summarized in Table 1. Typical resource requirements are on the order of 25 kg and 25 W. The design of every interstellar spacecraft will be driven by the power and telemetry system, especially the large high-gain antenna (2–3m diameter).

Telemetry

Several options for telemetry have been studied, but unrealistic constraints on attitude control allow only classic radio communication. A constant data acquisition rate of up to 500 bps (compressed) would need to be relayed to Earth in typically two weekly passes (2x8 hours nominal), thus requiring a downlink capability of 5.8 kbps also at 200 AU. This could be achieved with 35m antennae in the initial mission years, but a 70m antenna would be needed once the spacecraft reaches larger distances.

Table 2: Strawman payload. Augmentation (bonus) payload is indicated in parentheses and not included in mass and power total. Initial studies have shown that such a payload complement could measure the expected small signals.

Acronym	Instrument	Mass [kg]	Power [W]	Telemetry [bps]	Volume [cm ³]	Measurements
MAG	Magnetometer	2.0	1.5	50	500	1 Hz magnetic fields
PA	Plasma Analyzer	3.5 (2)	3.5 (2)	60 (20)	2 x 25x25x25	Plasma composition
NA	Neutral Analyzer	2.5	3.5	50	25x25x25	Neutrals, limited composition
PW	Plasma Waves	5	4	30	25x25x25	Radio and Plasma waves
DA	Dust Analyzer	1	1	10	25x25x30	Dust mass, velocity, composition
EP	Energetic Particles	4.5 (2)	5 (2)	60	2 x 25x25x25	H: 4 keV – 300 MeV ions: 5 keV/n – 400 MeV/n e-: 2keV – 20 MeV
ENA	Energetic Neutrals	5	5	50	60x60x20	Hydrogen ENAs: 0.05 – 5 keV Key elemental composition
LA	Ly-alpha	1.2	1.5	50	tbd	Ly-alpha broad-band photometry
IRV	IR/VIS imager	(5)	(5)	(50)	tbd	Wide-field infrared and visible imaging
SXR	Soft-X-Ray	(5)	(5)	(50)	tbd	Soft X-ray background, solar wind – planet interactions
Total		24.7	25	360		

- [1] Barnard, L., et al., (2011), GRL, 38, L16103, doi:10.1029/2011GL048489
- [2] Beer, J., et al., (1990), Nature, 347, 164–166
- [3] Bock, J.J., (2011), Bull. AAS, 43, 320.06
- [4] Breitschwerdt, D., and de Avillez, M.A., (2006), A&A, 452, L1–L5
- [5] Burlaga, L.F., et al., (1994), JGR, 99, 21511–21524
- [6] Burlaga, L.F., et al., (2005) Science, 309, 2027–2029
- [7] Cox, (2005), ARA&A, 43, 337
- [8] Crowder et al., (2012), ApJ, 758, 143
- [9] Cummings, A.C., et al., (2002), ApJ, 578, 194–210
- [10] Czechowski, A., and Mann, I., (2003), JGR, 108, L13.1–L13.9
- [11] Decker, R.B., et al., (2010), AIP Conf. Proc. 1302, 51–57
- [12] Dwyer, J.R., et al., (1997), ApJ, 490, L115–L118
- [13] Ferreira, S.E.S., et al., (2007), JGR, 112A, 11101
- [14] Fiehler, D.J., and McNutt, R.L., (2006), J. Spacecraft & Rockets, 43, 1239–1247
- [15] Fisk, L.A., and Jokipii, J.R., (1999), Space Sci. Rev. 89, 115–124
- [16] Fisk, L.A., et al., (1974), ApJ, 190, L35–L37
- [17] Florinski, V., and Zank, G.P., (2006), GRL, 33, 15110.1–15110.5
- [18] Florinski, V., et al., (2003), J. Geophys. Res., 108, 1228
- [19] Frisch, P.C., and Slavin, J.D., Short-term variations in the galactic environment of the sun. ArXiv -Astrophysics e-prints
- [20] Frisch, P.C., et al., (1999), ApJ, 525, 492–516
- [21] Galli, A., et al., (2006), ApJ, 644, 1317–1325
- [22] Gayley, K.G., et al., (1997), ApJ, 487, 259 (1997)
- [23] Grün, E., et al., (2005), Icarus 174, 1–14
- [24] Gurnett, D.A., et al., (2006), AIP Conf. Proc., 858, 129–134
- [25] Gustafson, B.A.S. & Misconi, N.Y. (1979), Nature, 282, 276–278
- [26] Heber, B., et al., (2001) in Scherer, K., Fichtner, H., Fahr, H.J., Marsch, E. (eds.) The Outer Heliosphere: The Next Frontiers, 191
- [27] Herbst, K., et al., (2012), ApJ, 761, 17
- [28] Ip, W.-H. and Axford, W.I., (1985), Astron. Astrophys. 149, 7–10
- [29] Izmodenov, V.V., et al., (1999), JGR, 104, 4731–4742
- [30] Janhunen, P. A., et al., (2013), Geosci. Instrum. Method. Data Syst., 2, 85–95,
- [31] Janhunen, P., et al., (2010), Rev. Sci. Instrum., 81, 111301
- [32] Kayser-Threde (2004), Technical Report IHP-TN-KTH-001 1
- [33] Koutroumpa et al., (2009), ApJ, 697, 1214
- [34] Krimigis, S.M., et al., (2009), Science 326, 971–973
- [35] Krimigis, S.M., et al., (2010), AIP Conf. Proc. 1302, 79–85
- [36] Krimigis, S.M., et al., (2011), Nature, 474, 359–361
- [37] Kurth, W.S., and D.A. Gurnett, (2003), J. Geophys. Res. 108, 2–13
- [38] Lallement, R., et al., (2004), A&A 418, 143
- [39] Lallement, R., et al., (2005) Science 307 (2005) 1447–1449.
- [40] Landgraf, M., et al. (2000), J. Geophys. Res., 105, 10343–10352.
- [41] Linsky, J.L., and Wood, B.E., (1996), ApJ, 463, 254
- [42] Lisse et al, (1996), Science, 274, 205
- [43] Lyngvi, A.E., et al., (2007), ESA Study Report, 40
- [44] Mann, I., et al., (2004), Adv. Space Res. 34, 179
- [45] McComas, D.J. And Schwadron, N.A., (2006), GRL, 33, 4102
- [46] McComas, D.J., et al., (2009), Science 326, 959–962.
- [47] McComas, D.J., et al., (2012), ApJS, 203, 1–36
- [48] McComas, D.J., et al., (2012), Science, 336, 1291
- [49] McDonald, F.B., (1998), Space Sci. Rev. 83, 33–50
- [50] McDonald, F.B., et al., (2006), AIP Conf. Proc., 858, 79–85
- [51] McNutt, R.L., and Wimmer-Schweingruber, R.F., (2011), Acta Astronautica, 68, 790–801
- [52] McNutt, R.L., et al., (2005), AIAA, 2005–4272
- [53] McNutt, R.L., et al., (2011), Acta Astronautica, 69, 767–776
- [54] Mewaldt, R.A., et al., (1994), EOS Trans. 75, 185
- [55] Opher, M., et al., (2007), Science 316, 875–878
- [56] Owens, M. J., and Lockwood, M., (2012), JGR, 117, A04102, doi: 10.1029/2011JA017193
- [57] Pesses, M.E., et al., (1981), ApJ, 246, L85–L88
- [58] Quarta, A.A. and Mengali, G., (2010), J. Guid. Contr. Dyn., 33, 740–755
- [59] Raisbeck, G.M., et al., (1987), Nature, 326, 273–277
- [60] Richardson, J. D., and Stone, E.C., (2009), Space Sci. Rev. 143, 7–20
- [61] Richardson, J.D., et al., (2008), Nature 454, 63–66
- [62] Rowan-Robinson, M. and May, B., (2013), MNRAS, 429, 2894–2902
- [63] Scherer, K., et al., (2002), J. Atmos. Sol. Ter. Phys. 64, 795–804
- [64] Scherer, K., et al., (2011), ApJ, 735, 128
- [65] Schwadron, N.A., et al., (2002), GRL, 29, 1–4
- [66] Schwadron, N.A., et al., (1996), GRL, 23, 2871–2874
- [67] Slavin, J., et al., (2012), ApJ, 760, 46–61
- [68] Slavin, J.D., and Frisch, P.C., (2006), ApJ, 651, L37–L40
- [69] Slavin, J.D., and Frisch, P.C., (2007), Space Sci. Rev. 130, 409–414
- [70] Sterken, V.J., et al. (2012), A&A, 538, A102
- [71] Stone, E.C., et al., (2005), Science 309, 2017–2020
- [72] Strauss, R.D., et al., (2013), ApJ, 765, L18
- [73] Webber, W.R., and McDonald, F.B., (2012), GRL, 40, doi:10.1002/grl.50383.
- [74] Wimmer-Schweingruber, R.F., et al., (2009), Exp. Astron., 24, 9–46
- [75] Witte, M., (2004), A&A, 426, 835–844
- [76] Witte, M., et al., (1996), Space Sci. Rev 78, 289–296
- [77] Wolfire et al., (2003), ApJ, 587, 278
- [78] Wood, B.E., et al., (2005), ApJS, 159, 118–140
- [79] Wurz, P., et al., (2006) AIP Conf. Proc. 858, 269–275
- [80] Yamamoto, S., and Mukai, T., (1998), A&A, 329, 785–791
- [81] Zank, G.P., and Frisch, P.C., (1999), Astrophys. J. 518, 965–973

The Exploration of Titan with an Orbiter and a Lake-Probe

Giuseppe Mitri and 124 Colleagues

In response to

Call for White Papers for the definition of the L2 and L3 missions
in the ESA Science Program 5 March 2013

24 May 2013

Spokesperson:

Giuseppe Mitri

Istituto Nazionale di Astrofisica e Planetologia Spaziali

Istituto Nazionale di Astrofisica

Via del Fosso del Cavaliere 100, Rome, Italy

Email: Giuseppe.Mitri@iaps.inaf.it

Phone: +39.06.454.88504

Authors List: Mitri Giuseppe (*INAF-IAPS, Italy*), Orosei Roberto (*INAF-IAPS, Italy*), Hayes Alex (*Cornell, USA*), Coustenis Athena (*Observ. of Paris, France*), Fanchini Gilbert (*La Sapienza, Italy*), Khurana Krishan (*UCLA, USA*), Lebreton Jean-Pierre (*LPC2E-CNRS & LESIA-Obs. Paris, France*), Lopes Rosaly (*JPL, USA*), Lorenz Ralph (*JHU-APL, USA*), Iess Luciano (*Univ. La Sapienza, Italy*), Meriggiola Rachele (*Univ. La Sapienza, Italy*), Moriconi Maria Luisa (*ISAC-CNR, Italy*), Sotin Christophe (*JPL, USA*), Stofan Ellen (*Proxemy Research, USA*), Tokano Tetsuya (*Universität zu Köln, Germany*), Tosi Federico (*INAF-IAPS, Italy*)

Supporters List: Altieri Francesca (*INAF-IAPS, Italy*), Barabash Stas (*IRF-SISP, Sweden*), Behoukova Marie (*Charles Univ., Czech Rep.*), Bézard Bruno (*Obs. Paris, France*), Bills Bruce (*JPL, US*), Boudon Vincent (*Univ. Bourgogne, France*), Briois Christelle (*LPC2E-CNRS, France*), Bruzzone Lorenzo (*Univ. of Trento, Italy*), Cadek Ondrej (*Charles Univ., Czech Rep.*), Cairo Francesco (*ISAC-CNR, Italy*), Callegari Mattia (*EURAC, Italy*), Capria Maria Teresa (*INAF-IAPS, Italy*), Cardì Margherita (*La Sapienza, Italy*), Carrasco Nathalie (*LATMOS, France*), Carpy Sabrina (*Univ. of Nantes, France*), Carrozzo Giacomo (*INAF-IAPS, Italy*), Cartacci Marco (*INAF-IAPS, Italy*), Casarano Domenico, (*CNR-IRPI, Italy*), Castillo Julie (*JPL, USA*), Coates Andrew (*UCL, UK*), Coll Patrice (*LISA, France*), Conde Aitor (*La Sapienza, Italy*), Cordier Daniel (*Obs. of Besancon, France*), Cornet Thomas (*Univ. of Nantes, France*), Dehant Veronique (*OBS, Belgium*), D’Aversa Emiliano (*INAF-IAPS, Italy*), de Kok Remco (*SRON, Netherlands*), Dougherty Michele (*Imperial College, UK*), Encrenaz Pierre (*Obs. of Paris, France*), Esposito Francesca (*INAF, Italy*), Ferri Francesca (*CISAS, Italy*), Filacchione Gianrico (*INAF-IAPS, Italy*), Fletcher Leigh (*Oxford Univ., UK*), Fortes Andrew (*Univ. College London, UK*), Frampton Robert (*Boeing, USA*), Galand Marina (*Imperial College, UK*), Geppert Wolf (*Stockholm Univ., Sweden*), Grasset Olivier (*Univ. of Nantes, France*), Grodent Denis (*Univ. Liège, Belgium*), Gulino Marco (*La Sapienza, Italy*), Hartogh Paul (*MPS-MPG, Germany*), Hussmann Hauke (*DLR, Germany*), Irwin Pat (*Oxford Univ., UK*), Jackman Caitriona (*UCL, UK*), Leese Mark (*Open Univ., UK*), Karatekin Ozgur (*ORB, Belgium*), Komatsu Goro (*IRSPS, Italy*), Langevin Yves (*IAS, France*), Lebonnois Sébastien (*LMD, France*), Lebleu Denis (*Thales, France*), Le Gall Alice (*LATMOS, France*), Le Mouélic Stéphane (*Univ. of Nantes, France*), Lilensten Jean (*IPAG, France*), Livengood Tim (*NASA/Goddard, USA*), Longobardo Andrea (*INAF-IAPS, Italy*), Lucas Antoine (*IPGS, France*), Lunine Jonathan (*Cornell, USA*), Malaska Mike (*JPL, US*), Mandt Kathy (*SWRI, USA*), Mastrogiuseppe Marco (*Univ. La Sapienza, Italy*), Messina Angela (*Univ. La Sapienza, Italy*), Mitchell Karl L. (*JPL, US*), Mitchell Jonathan (*UCLA, US*), Morse Andrew (*Open Univ., UK*), Mousis Olivier (*Obs. of Besancon, France*), Mueller-Wodarg Ingo (*Imperial College, UK*), Nna Mvondo Delphine (*Univ. of Nantes, France*), Nimmo Francis (*UCSC, USA*), Nixon Conor (*GSFC, USA*), Notarnicola Claudia (*EURAC, Italy*), Ori Gian Gabriele (*IRSPS, Italy*), Norman Lucy (*Univ. College, UK*), Paillou Philippe (*Obs. Bordeaux, France*), Palumbo Pasquale (*Univ. Parthenope, Italy*), Pappalardo Robert (*JPL, USA*), Patel Manish (*Open Univ., UK*), Palomba Ernesto (*INAF-IAPS, Italy*), Patterson Wes (*JHU-APL, US*), Pettinelli Elena (*Univ. Roma Tre, Italy*), Pizzini Annalisa (*Univ. La Sapienza, Italy*), Plaut Jeffrey J. (*JPL, US*), Poggiali Valerio (*Univ. La Sapienza, Italy*), Poncy Joël (*Thales, France*), Postberg Frank (*Heidelberg/Stuttgart Univ., Germany*), Radebaugh Jani (*Brigham Young University, US*), Raulin Francois (*LISA, France*), Rengel Miriam (*Max Planck Institut, Germany*), Reh Kim (*JPL, USA*), Rodriguez Sébastien (*CEA, France*), Schubert Gerald (*UCLA, US*), Schulze-Makuch Dirk (*WSU, USA*), Scipioni Francesca (*INAF-IAPS, Italy*), Seu Roberto (*Univ. La Sapienza, Italy*), Sittler Ed (*NASA/Goddard, USA*), Soderblom Jason (*MIT, USA*), Sohl Frank (*DLR, Germany*), Spilker Linda (*JPL, USA*), Stephan Katrin (*DLR, Germany*), Teanby Nick (*Univ. Bristol, UK*), Thissen Roland (*IPAG, France*), Tobie Gabriel (*Univ. of Nantes, France*), Tortora Paolo (*Univ. of Bologna, Italy*), Turtle Elizabeth (*JHU-APL, USA*), Van Hoolst Tim (*Royal Obs. of Belgium, Belgium*), Vinatier Sandrine (*Obs. of Paris, France*), West Robert (*JPL, USA*), Westlake Joseph (*JHU-APL, USA*), Yelle Roger (*Univ. of Arizona, USA*), Yonggyu Gim (*JPL, US*)

1. Overview

Titan, Saturn's largest satellite, is a unique object in the Solar System presenting a substantial atmosphere, a surface with a complex interplay of geological processes and an outer ice shell overlying a subsurface ocean. The climate on this icy satellite boasts a multi-phase hydrological cycle where methane plays a role similar to water on Earth. The Cassini-Huygens mission to the Saturn system, which is still ongoing, accomplished a first in-depth exploration of Titan showing extended hydrocarbon lakes and seas in Titan's polar regions which, combined with extensive equatorial dune deposits, form the largest reservoir of organics in the solar system. The presence of standing bodies of liquid, dissected fluvial channels, tectonic features, vast dune fields and putative cryovolcanic flows express striking analogies with terrestrial geological activity. Titan is the only other place in the Solar System with the Earth to have lakes and seas.

In spite of all its advances, the Cassini-Huygens mission, due to its limited instrumentation and timeline among other, cannot address some fundamental questions such as "What are the complex organic molecules that form the dune material, and how have they been formed?" and "What is the composition of the lakes and seas". Cassini has demonstrated that there must be exchanges between the interior and the atmosphere: "What are the processes that form the atmosphere?" and "Is Titan endogenically active?" We discuss how the future exploration of Titan can benefit from a dedicated orbiter and an *in situ* lake-probe that would address central themes regarding the nature and evolution of this icy satellite and its organic-rich environment. Prior to an orbiter phase around Titan and an *in situ* exploration of its lakes, mission scenarios could also involve investigations of the planet Saturn and its other satellites including the geologically active icy moon Enceladus.

2. Scientific Objectives and Investigations, and Scientific Case for Titan's Exploration

Section 2.1 presents the scientific objectives for Titan's exploration and summarizes the scientific investigations. Section 2.2 discusses the scientific case for the exploration of Titan, and presents in more detail the scientific investigations for a future exploration of Titan from a dedicated orbiter and lake-probe. Section 2.2.1 discusses how the combined measurements of the gravity field, rotational dynamics and electromagnetic field can improve our understanding of the interior and evolution of Titan. Section 2.2.2 discusses how observations from an orbit-platform will improve our understanding of the geological processes and surface composition. Section 2.2.3 presents fundamental questions on Titan's atmosphere that can be resolved by a future mission offering a dedicated orbiter and *in situ* exploration of a lake. Finally in Section 2.2.4 we discuss how *in situ* exploration of Titan's lakes is an unprecedented opportunity to understand the hydrocarbon cycles, to investigate a natural laboratory for prebiotic chemistry and the limits of life, and study meteorological and marine processes in an exotic environment.

2.1 Scientific Objectives and Investigations

The scientific objectives for an exploration of Titan are:

G1. Determine how Titan was formed and evolved, and its internal structure;

G2. Determine the nature of the geological activity of Titan;

G3. Characterize the atmosphere of Titan and determine how the hydrological-like cycle works;

G4. Determine the marine processes and chemistry of Titan's lakes and seas.

These scientific objectives can be achieved by complementary investigations of a dedicated orbiter and a lake-probe. The *in situ* exploration of a sea will contribute to the achievement of all of the scientific objectives G1, G2, G3 and G4. [Table 1](#) summarizes the scientific investigations, measurements and examples of approaches considered for an exploration of Titan from an orbiter and lake probe. Table 1 also indicates what scientific investigation can be used to achieve which scientific objective G1, G2, G3, and G4.

An exploration of Titan from an orbiter and *in situ* exploration of Titan's seas will address fundamental questions involving the origin, evolution, and history of both Titan and the broader Saturnian system. Titan's ongoing organic chemistry has direct applicability to early prebiotic chemistry on Earth, allowing the investigation of reactions and timescales inaccessible to terrestrial labs. Titan also represents a potentially common climate system amongst extrasolar planets and a detailed understanding of its nature and evolution will help to interpret data from these distant objects. Finally, it is important to emphasize the educational potential of this mission concept. The novel idea of sending a "boat" to an extraterrestrial sea, which is both innovative and awe-inspiring, can be used as an inspirational tool at all levels of education and public outreach.

Table 1. Science Investigations, measurements and example of approaches.

Science Investigations	Measurements	Example of Approaches
Internal Structure	Determine the internal structure (G1)	Gravity field
	Obliquity, libration, occupancy of a Cassini state	S/C radio-tracking to measure static gravity field
		Imaging (e.g. SAR) (S/C) with spatial resolution ~30-50 m
	Gravity anomalies in the ice shell and deep interior (G1)	S/C radio-tracking to measure static gravity field
		High order gravity field components
	Topography	Altimetry / stereo imaging / interferometric topography (S/C) with vertical resolution of ~10 m
Subsurface ocean	Present tidal quality factor Q (G1)	Time varying of the degree-2 gravity field
	Hydrostatic compensation (G1)	S/C radio-tracking to measure time dependent gravity field
		Gravity field
	Extent of the subsurface ocean (G1)	Topography
		Altimetry / stereo imaging / interferometric topography (S/C) with vertical resolution ~10 m
		Time varying of the degree-2 gravity field (k_2)
		Detection of electromagnetic induction from a subsurface ocean
Ice shell and Geology	Thickness of the ice shell (G1)	Magnetometer measurements (S/C)
		Tidal deformation (h_2)
		Altimetry and/or interferometric topography (S/C) with ~1 m vert. res.
		Libration and Obliquity
	Characterize the surface and subsurface structures and the determine their correlation (G1, G2)	Imaging (e.g. SAR) (S/C) with 30-50 m resolution
		S/C radio-tracking to measure the static gravity field
		Time varying of the degree-2 gravity field (k_2)
		S/C radio-tracking to measure time dependent gravity field
	Correlate the surface composition and their distribution with surface structures (G2)	Tidal deformation (h_2)
		Altimetry and/or interferometric topography (S/C) with ~1 m vert. res.
		Libration
		SAR imaging (S/C)
	Global mapping of the surface	S/C tracking to measure the static gravity field
		Global topography
		Altimetry / stereo imaging / interferometric topography (S/C)
	Profiles of subsurface structures and thickness of the surface organic material layer	Radar sounding (S/C) with ~10 km penetration depth and ~30 m vertical resolution
		Global mapping of the surface
	Global mapping of the surface composition	SAR imaging (S/C) with ~30-50 m resolution
		Infrared spectral imaging (S/C) within the methane band windows

Atmosphere	Characterize the atmospheric circulation and the composition of the atmosphere from orbit and during probe decent (G3)	Determine the dynamics and thermal state of the atmosphere including the clouds	Infrared spectra imaging of the atmosphere (S/C)
			Lidar (S/C)
		Determine the composition	Infrared spectra imaging of the atmosphere (S/C)
			Mass spectrometry with direct sampling from orbit (S/C) and during decent (L/P)
	Characterize the lake / atmosphere interactions in situ (G3)	Determine temperature, pressure, composition, evaporation rate and physical properties that characterize lake and atmosphere interactions	Lidar (S/C)
			Mass spectrometry (L/P)
Lakes/Seas	Characterize lakes/seas and their composition / chemistry (astrobiological potential) (G4, G1, G2, G3)	Lake and sea composition, including low/high mass hydrocarbons, noble gases, and carbon isotopes	Physical properties package (L/P)
		Exchange processes at the sea-air interface to help constrain the methane cycle	Mass spectrometry / chemical analyzer (L/P)
			Atmosphere physical properties package (temperature sensor, barometer, anemometer)/ Meteorological package / Thermo-gravimetry (L/P)
		Presence and nature of waves and currents	Atmosphere physical properties package (temperature sensor, barometer, anemometer) / Meteorological package (L/P)
			Physical properties package (L/P)
		Properties of sea liquids including turbidity and dielectric constant	Surface Imaging (~250 μrad/pixel)
			Lake physical properties package (turbidity and dielectric constant measurements) (L/P)
		Sea depths to constrain basin shape and sea volume	Sonar (L/P)
		Shoreline characteristics, including evidence for past changes in sea level	Surface Imaging (~250 μrad/pixel)
Descent camera imaging / stereogrammetry (L/P)			

2.2 Scientific Case for the Exploration of Titan

2.2.1 Geophysics

The combination of measurements of the gravity field, rotational dynamics and electromagnetic field will contribute to the achievement of the scientific objective G1. Static and dynamic gravity field measurements (*Iess et al.* 2010, 2012), and shape models (*Zebker et al.*, 2009; *Mitri et al.*, 2013) from the Cassini-Huygens mission have greatly expanded our knowledge of the geophysical processes and structure of Titan's interior and as a consequence its formation and evolution have also become better understood. Data from the Cassini-Huygens mission suggest that either Titan has no induced magnetic field or else it is not yet detectable, while an intrinsic magnetic field is found to be absent (*Wei et al.*, 2010). While there is still not yet a unique internal structural model of Titan, constraints provided by gravity, topography and rotation data suggest that Titan's interior has some degree of differentiation and possesses a water-ice outer shell separated from its deep interior by a dense subsurface water ocean (*Mitri et al.*, 2013). Thermal models suggest the presence of a high-pressure polymorph ice layer between the deep interior and the subsurface ocean (*Tobie et al.*, 2005; *Mitri et al.*, 2010).

Radio-tracking data registered during multi-flybys of Titan has yielded gravity field measurements with a spherical harmonic expansion to degree three (*Iess et al.*, 2010, 2012). The quadrupole (J_2 and C_{22}) moments of the gravity field were used to constrain the normalized axial moment of inertia which value was inferred using the Radau-Darwin approximation as $MoI \cong 0.3414 \pm 0.0005$ (*Iess et al.*, 2010). The relatively high moment of inertia suggests that Titan's deep interior is formed of relatively low-density material ($\sim 2600 \text{ kg m}^{-3}$) that it is consistent with a deep interior differentiated and composed of silicate hydrates such as antigorite (*Castillo-Rogez and Lunine*, 2010; *Fortes et al.*, 2007) or only partially differentiated and

composed of a mixture of ice and rocks (*Iess et al.*, 2010). The state of the internal structure of Titan is still debated, and future investigations are therefore necessary to determine its internal structure. The interior structure of Titan can be inferred from a combined measurement of the gravity field, rotational dynamics and electromagnetic field.

Future gravity measurements from a spacecraft (S/C) in a polar orbit will provide a significant improvement to the gravity field determination with respect to Cassini's gravity measurements. We consider a nominal final orbit of the S/C around Titan as a 1500 km altitude circular near-polar orbit. Using such an orbit, the static gravity field of Titan can be resolved to degree 10 and possibly higher order. Also the gravity anomalies can be determined with a spatial resolution lower than 260 km allowing a better characterization of the density contrasts within the ice shell as well as within the deep interior.

Analysis based on Titan's shape suggests that Titan's outer ice shell is thicker at the equator and thinner at the poles (*Nimmo and Bills*, 2010; *Mitri et al.*, 2013). Estimates for the thickness of the ice shell vary from thin (50 km) to thick (200 km) (e.g., *Tobie et al.*, 2005, 2006; *Mitri et al.*, 2008), and determining which model is correct will require observations beyond the capability of Cassini. The thickness of Titan's ice shell has major implications for the thermal state of the crust. In fact, geological activities and processes on Titan could be related to the onset of thermal convection in the outer ice shell (*Tobie et al.*, 2005, 2006; *Mitri and Showman*, 2008). Analysis of Titan's topography suggests that the ice shell is currently in a conductive state (*Nimmo and Bills*, 2010; *Mitri et al.*, 2013); however during Titan's evolution it is possible that the ice shell may have transitioned between a convective to a conductive state (*Mitri et al.*, 2010).

The radio science experiment during Cassini mission allowed the measurement of the tidal Love number k_2 . The tidal Love number k_2 characterizes the tidal deformation of the interior. The relatively high measured tidal Love number k_2 (0.589 ± 0.150) of Titan indicates that its outer ice shell is decoupled from the deep interior by a global subsurface ocean (*Iess et al.*, 2012). The tidal Love number k_2 (*Iess et al.*, 2012) in conjunction with altimetric measurements taken during a future mission of the tidal deformation of the outer ice shell (Love number h_2 measurement) could be used to determine the thickness of the ice shell (see similar experiment proposed for Europa by *Wahr et al.*, 2006). Using a nominal polar circular orbit around Titan with a 1500 km altitude, the measurement of the tidal Love number k_2 can be performed with an expected uncertainty $\sigma_{k_2} \sim 5 \cdot 10^{-4}$ improving the actual accuracy on the determination of k_2 achieved during Cassini mission of two orders of magnitude. In addition, radar sounder observations with a penetration depth up to ~ 10 km with a vertical resolution of ~ 30 m from an orbital platform could directly determine the relict Brittle-Ductile transition of the ice shell revealing its thermal state.

Surface heights derived during the Cassini mission from the Cassini RADAR altimeter (*Elachi et al.*, 2004) and SAR-Topography technique (*Stiles et al.*, 2009) are sparsely distributed and cover only $\sim 1\%$ of Titan's surface. Using the available altimetric data, *Zebker et al.* (2009) and *Mitri et al.* (2013) determined the shape of Titan up to the seventh order of the spherical harmonic expansion. A quasi-global coverage of altimetric data with a ~ 10 m vertical resolution from a future mission is necessary to refine the shape model and gravity anomalies of Titan's interior and define the level of hydrostatic compensation of the shape and the outer ice shell, and a ~ 1 m vertical resolution to characterize the tidal deformation of the outer layer, determine the Love number h_2 . Radar altimetry is a suitable technique to determine the elevation of Titan's surface as demonstrated during the Cassini mission. Interferometric Synthetic Aperture Radar (InSAR) is also a possible technique to generate digital elevation model (DEM) of the surface and to produce maps of surface deformation in combination with the imaging of the surface.

The measurement of the imaginary part of the tidal Love number k_2 using gravity field measurements from the radio science experiment can be used to estimate the tidal quality factor $Q = |k_2|/\text{Im}(k_2)$. The estimation of Q will be used to constrain the present internal tidal dissipation of Titan providing information on the internal heat production as well as on the tidal

evolution of Titan. The estimation of Q will also provide information on the viscosity of the ice shell and indirectly on the thermal structure as well as on the thermal state (conductive or convective) of the ice shell.

The determination of rotational dynamics is crucial to understand the internal structure and geophysical processes on Titan. Evolution models of the orbital dynamics use the obliquity along with the eccentricity as a constraint. When the quadrupole gravity field is known, the obliquity may provide also the moment of inertia (*Bills and Nimmo, 2008*), a fundamental quantity to constrain the internal structure of Titan. The spin rate and the physical librations in longitude can provide essential information on the degree of internal differentiation and the ice shell thickness. Currently the only information about the rotational dynamics of Titan is derived from Cassini mission data (e.g., *Stiles et al., 2008*) and astronomical observations (*Seidelmann et al., 2006*).

During the Cassini observations the pole location was found to be not fully compatible with the occupancy of a Cassini state. Estimated obliquity (0.3°) is not in agreement with the predicted value (0.15°) inferred from the polar, dimensionless moment of inertia (0.34) derived from gravity field measurements (*Iess et al., 2010*). The interpretation of the disagreement between the measured and predicted values is controversial and it seems to point to a differentiated internal structure characterized by complex rotational dynamics. For this reason the pole location in terms of obliquity and the occupancy of a Cassini state should be investigated by a dedicated experiment. Physical librations in longitude are due to the eccentricity. Titan is in a 1:1 resonance and it is orbiting around Saturn at the same rate of its spin motion (*Meriggiola and Iess, 2012*). Since Titan's orbit is eccentric, the Saturn-satellite line and the long axis of the satellite are not aligned and they differ by a libration angle (g). The libration amplitude g_0 is inversely proportional to the moment of inertia of the body, and it is possible, if the quadrupole gravity field is known, to infer the thickness (h) of the outer icy shell from g_0 . The determination of g_0 can provide crucial information on the dimension as well as on the liquid/solid state of the shell and the deep interior. In the case of an 1-2 years orbital phase, the data availability could also allow to analyze and determine the influence of Titan's atmosphere on the spin motion, leading to a determination of the polar wandering, similar to previous experiments on Mars (*Folkner et al., 1997*). The rotational dynamics of Titan can be inferred from tracking surface landmarks along time using, for example, Synthetic Aperture Radar (SAR) images with a spatial resolution of ~ 30 -50 m, and/or from gravity field measurements (see *Cicalò and Milani, 2012*).

Observations of Cassini's magnetic field on Titan show that it currently does not have an intrinsic field generated by a dynamo (upper limit is 2 nT surface field at the equator). Even though the Cassini mission has not been successful thus far in inferring an electromagnetic induction response from a subsurface salty ocean as measured during the Galileo mission for Europa, Ganymede and Callisto, future, orbiting spacecraft can provide better information on the interior (subsurface ocean) from this technique. The discovery of periodic changes in the quadrupole gravity field of Titan at $\sim 4\%$ of the static value indicates the presence of an internal ocean (*Iess et al., 2012*). However, a low viscosity (but not liquid) layer could also equally explain these results; further independent confirmation of Titan's subsurface ocean is necessary.

Detection of electromagnetic induction from the interior of Titan induced by time-varying external magnetic fields could provide such a confirmation as has been successfully demonstrated in the discoveries of subsurface oceans of Europa, Callisto and Ganymede (*Khurana et al., 1998; Zimmer et al., 2000; Kivelson et al., 2002*) and the magma ocean of Io (*Khurana et al., 2011*). The basic principle of electromagnetic induction relies on the fact that subsurface conductors generate eddy currents on their surfaces in response to time varying electric fields imposed by the changing primary magnetic field. The secondary field generated by the eddy currents is dipolar in nature if the conductor has a spherical shape and can be easily characterized by a spacecraft bearing a magnetometer. Modeling of the secondary field provides information on the thickness and conductivity of the subsurface conductor. For the Galilean satellites, the time varying driving

primary field is provided by the rotating tilted magnetic field of Jupiter and its magnetosphere and has an amplitude ranging from 40 nT near Callisto to 800 nT at Io. However, Saturn's internal magnetic field is completely axisymmetric and therefore does not impose a time-varying field on its satellites. However a cyclical magnetic field of unknown origin has recently been detected; this field is supported by current systems in Saturn's magnetosphere (*Andrews et al.*, 2010). This magnetic field has an amplitude of 2-4 nT near Titan and has a period close to 10.8 hr which is an ideal sounding frequency. Assuming that the ice crust has a thickness of 100 km or less, a highly conducting ocean would then create a secondary induced field with a magnitude of ~ 0.6 -1.3 nT just above Titan's atmosphere (assuming an R^{-3} falloff for the dipolar response) which would be detectable by a magnetometer. Another inducing field with an amplitude exceeding 10 nT is imposed by the periodic exits of Titan from Saturn's magnetosphere. The several nT induction response to this signal would not be sinusoidal but would be easily characterized by an orbiting spacecraft. Unfortunately, the plasma interaction of Titan with the magnetosphere of Saturn creates magnetic perturbations with an amplitude of 10 nT or more (*Wei et al.*, 2010), therefore, it has so far not been possible to confirm the presence of an electromagnetic induction signal from Titan. This situation can be remedied if continuous observations from an orbiter were available. The spatial forms and frequency contents of magnetic fields generated by plasma interaction and by electromagnetic induction are distinct and can be easily separated by harmonic analysis if continuous observations from an orbiter were available. Such an experiment would be performed by the JUICE spacecraft mission at Ganymede where researchers hope to disentangle the inducing field from its ocean with an amplitude of ~ 60 nT from the much stronger permanent internal magnetic field with a magnitude of 1500 nT.

2.2.2 Surface Processes and Composition

A combination of measurements and observations from surface imaging, altimetric data, spectral data and radar sounding will allow the achievement of the scientific objective G2. The Cassini mission has revealed for the first time the complexity of the surface processes and composition of Titan. However, it is still not understood if Titan is endogenically active and if cryovolcanic activity is present. Other unsolved questions include determining the role of the lakes, channel network and dune field in the hydrological cycle. Sections 2.2.2.1 and 2.2.2.2 discuss how observations from the orbital-platform will improve our understanding of the geological processes and surface composition, respectively.

2.2.2.1 Geology

Our current knowledge of Titan's geology comes from Cassini observations, which show a geologically complex surface strikingly similar to Earth's. Titan has been called the Earth of the outer solar system. Its atmosphere is the second densest among the terrestrial planets and the present-day surface-atmospheric interactions make aeolian, fluvial, pluvial and lacustrine processes important on a scale previously seen only on Earth, despite vast differences in material properties and ambient conditions. These atmospheric and surface processes interact to create similar geologic features to Earth's: lakes filled with liquid, rivers caused by rainfall, and vast fields of dunes caused by wind (e.g. *Lorenz et al.*, 2006; *Hayes et al.*, 2008; *Radebaugh et al.*, 2008). The scarcity of impact craters indicates that Titan's surface is relatively young, with estimates ranging from 200 Ma to 1 Ga (*Lorenz et al.*, 2007; *Wood et al.*, 2010). Mountains indicate that tectonic activity has taken place (*Radebaugh et al.*, 2007; *Mitri et al.*, 2010) although it may not necessarily be endogenic in nature (*Moore and Pappalardo*, 2011). Several putative cryovolcanic features have been identified (e.g. *Lopes et al.*, 2007; *Soderblom et al.* 2009; *Lopes et al.*, 2013) and the existence of some ongoing volcanic activity has been proposed (*Nelson et al.*, 2009a,b), though the cryovolcanism interpretation has been disputed (*Moore and Pappalardo*, 2011). The variety of geologic processes on Titan and their relationship to the methane cycle make it particularly significant in solar system studies.

The Cassini's SAR swaths have revealed a wide variety of geologic features, examples of which are shown in [Figure 1](#). However, interpretations of Titan's surface properties and geologic history have been hindered by observational challenges primarily relating to its dense obfuscating atmosphere. Currently, the highest resolution observations of Titan's surface come from Synthetic Aperture Radar (SAR) images with a pixel-scale of ~300 m. At this resolution, Cassini has imaged ~35% of Titan's surface. Including lower resolution HiSAR data (~few km pixel scale), obtained at further distances from Titan than standard SAR, Cassini's surface coverage increases to ~50% (*Hayes et al.*, 2011).

A follow-up mission to Titan should image the surface at a pixel-scale of 30 m or better with an almost complete surface coverage using Synthetic Aperture Radar (SAR) technique. As has been shown in both the terrestrial and Martian communities, resolution increases of an order of magnitude or better results in fundamental advances in the interpretability of observed morphologies. Altimetric data with a vertical resolution of ~10 m is necessary for the geological interpretation of surface features. Both radar altimetry and Interferometric Synthetic Aperture Radar (InSAR) are suitable techniques to determine the elevation of Titan's surface (see Section 2.1). Radar sounder measurements with a penetration depth up to few km and a vertical resolution of 10-30 m are capable of characterizing the subsurface structures and determining their correlation with surface features allowing the determination of the internal processes that form the surface geological features. In addition, radar sounder measurements will determine the thickness and structure of surface organic-material deposits, which can be used to ascertain the global volume of organic material on Titan.

The major geologic unit types and their possible genesis have been recently reviewed by *Lopes et al.* (2010). Below, we briefly describe each of the major geomorphic units that have been identified and, where appropriate, state how higher resolution data could provide fundamental advancements in their interpretability.

Hummocky and mountainous terrain: This unit consists of numerous patches of radar-bright, textured terrains, some of which have been interpreted to be of tectonic origin (*Radebaugh et al.*, 2007). Among these features are long mountain chains, ridge-like features, elevated blocks which stand generally isolated, and bright terrains of a hilly or hummocky appearance. This terrain is mostly found as isolated patches or long mountain chains, small in areal extent. The exception is Xanadu, a large area about 4500 km across. The *hummocky and mountainous terrain unit*, including Xanadu, is thought to be the oldest geologic unit exposed on Titan's surface (*Lopes et al.*, 2010). Deciphering the origin of this feature class awaits high resolution imagery that can resolve the complex lineations and brightness variations that make up the hummocky appearance; altimetric data can be used to determine the elevation of this terrain.

Impact craters: Few well-preserved impact craters have been identified on Titan. However, there are numerous structures on Titan that may have been caused by impact. Radar-bright complete or incomplete circles are seen all over the surface. *Wood et al.* (2010) identified 44 possible impact craters with a wide variety of morphologies and, since then, others have been recognized and mapped. *Neish et al.* (2013) found Titan's craters are shallower than those on Ganymede, probably due to aeolian infilling, with possible contributions from viscous relaxation and fluvial modification. Radar sounding can be used to determine the structure of the craters. Detailed morphodynamic analysis of Titan's craters, however, require high resolution stereo models that allow quantitative comparison with numerical simulations.

Dunes: Aeolian transportation and deposition is a major and widespread process on Titan. Features resembling linear dunes on Earth have been identified by SAR, covering regions hundreds, sometimes thousands, of kilometers in extent (*Lorenz et al.*, 2006). Dunes are dark to RADAR and ISS and correlate well with dark surface units identified by VIMS. The dunes are thought to be made of organic particles (*Soderblom et al.*, 2007), which may have originated as photochemical debris rained out from the atmosphere, deposited and perhaps later eroded by fluvial processes into sand-sized particles suitable for dune formation (*Atreya et al.*, 2006;

Soderblom et al., 2007). Titan's sand dunes represent a detailed geomorphic record of both current and past wind regimes that can reveal global changes in climate and surface evolution (e.g. *Radebaugh et al.*, 2008). Studies of Titan's dune morphologies, how the morphologies spatially vary and how the dunes interact with local topography provide a basis for interpreting wind direction, sediment supply, and atmosphere conditions such as boundary layer depth. Thus dunes represent a morphologic unit where even slight increases in image resolution can result reveal patterns that have significant interpretive implications.

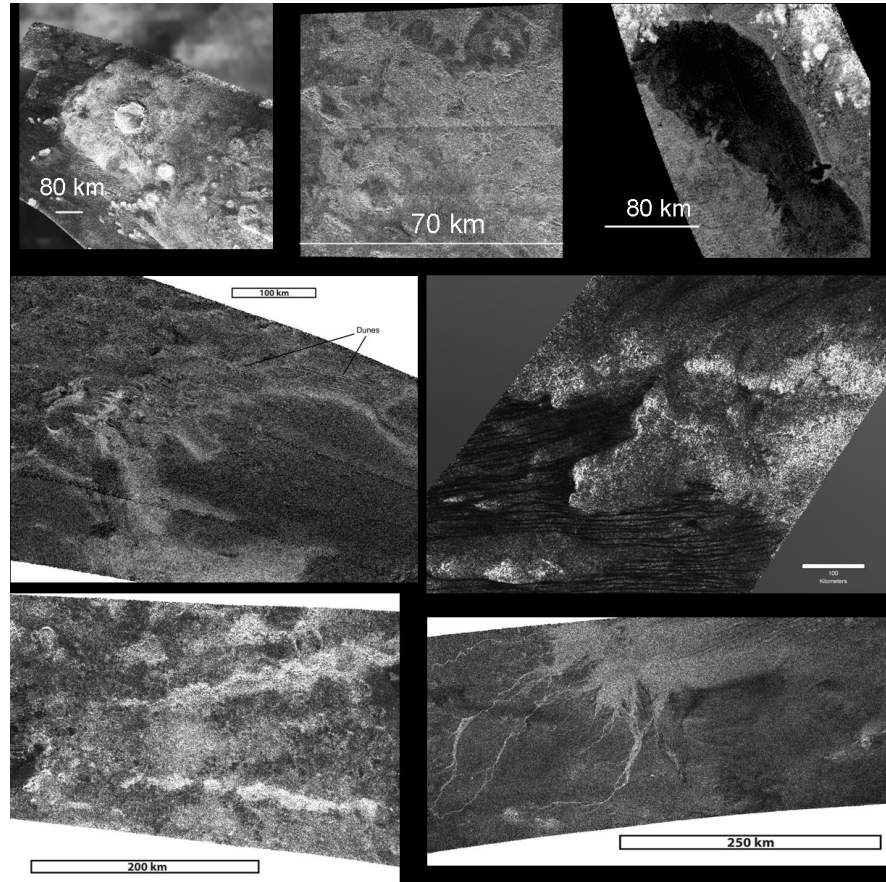


Figure 1. Examples of geologic features on Titan: Top row from left: Sinlap crater with its well defined ejecta blanket. Middle: Crateriform structure (suspected impact crater) on Xanadu. Right: Ontario Lacus near Titan's south pole. Middle row, left: Putative cryovolcanic flow Winia Fluctus (lobate, radar bright), note dunes overlaying part of flow. The flow overlays radar-dark undifferentiated plains. Right: Dunes abutting against hummocky and mountainous terrain. Bottom left: mountains (radar bright) in the equatorial region. Right: Elivagar Flumina, interpreted as a large fluvial deposit, showing braided radar-bright channels. Deposits overlay radar-dark undifferentiated plains.

Channels: Many radar-bright and radar-dark channels are seen in SAR images, widely distributed in latitude and longitude. It is likely that many more channels exist on Titan, but are too small for the Cassini spacecraft SAR data to resolve. Channels are seen connecting some lakes at high latitudes and cutting the hummocky, mountainous terrain at lower latitudes, possibly indicating orographic rainfall. Channels, lakes and dunes are thought to be the youngest geologic units on Titan (*Lopes et al.*, 2010). The *Labyrinthic terrains unit*, identified by *Malaska and Radebaugh* (2010) is a probable end-member of the different types of channel and valley network morphologies identified on Titan. There are few areas of this type of terrain exposed, mostly at

high latitudes. Higher resolution images of Titan's channel networks can allow pattern analysis, such as branching angle distribution, that will address the strength of the bedrock and origin (e.g., overland flow vs. subsurface drainage) of the network (e.g., *Burr et al.*, 2013).

Lakes: Hydrocarbon lakes are found in the polar regions (*Stofan et al.*, 2007). While Section 2.4 discusses the benefits of *in situ* lake exploration, there is also much to learn from studying the morphology and distribution of lacustrine features on Titan. Titan's lakes and seas are found both connected to and independent of regional and local drainage networks. They have shoreline transitions varying from sharp to diffuse and include a variety of plan-form shapes ranging from near-circular to highly-irregular. Exposed surface areas span from the limits of detection ($\sim 1 \text{ km}^2$) to more than 10^5 km^2 (*Hayes et al.*, 2008). Along with the surrounding landscape topography, these lacustrine features express a range of morphologic characteristics that constrain basin formation and evolution. For instance, the largest lakes, or seas, in the north, consisting of Ligeia, Kraken and Punga Mare, have shorelines that include shallow bays with well-developed drowned river valleys (*Stofan et al.*, 2007). Such drowned valleys indicate that once well-drained upland landscapes have become swamped by rising fluid levels where sedimentation is not keeping pace with a rising base level. These features also imply that liquid levels were previously lower and stable long enough to allow the currently drowned networks to be incised into the terrain. In contrast the largest southern lake, Ontario Lacus, expresses depositional morphologies along its western shoreline that include lobate structures interpreted as abandoned deltas (*Wall et al.*, 2010). Obtaining higher resolution images and topography data of Titan's lakes will allow quantitative morphology analysis that will lead to new hypotheses for landscape formation and evolution.

Candidate cryovolcanic terrain: The SAR swaths also revealed several lobate, flow-like features and constructs interpreted as resulting from cryovolcanism (*Lopes et al.*, 2007; *Wall et al.*, 2009) and VIMS data also identified some features that were thought to be cryovolcanic in origin (*Sotin et al.*, 2005; *Barnes et al.*, 2006; *Nelson et al.*, 2009a,b). As more data were acquired, particularly overlapping SAR coverage from which stereo could be derived, several putative cryovolcanic units were re-assessed and thought to be of a different origin (*Lopes et al.*, 2010) leading *Moore and Pappalardo* (2011) to argue that there may be no cryovolcanic units on Titan at all. However, radar stereogrammetry has revealed a tall mountain (Doom Mons) next to a deep elongated pit (depth $> 1 \text{ km}$), adjacent to several flow-like features. This is considered the strongest evidence for a cryovolcanic region on Titan (*Lopes et al.*, 2013). Higher resolution images and topography are needed in order to further scrutinize potential cryovolcanic features. Radar sounder measurements can distinguish cryovolcanic features such as calderas from other features that have similar morphology (e.g. impact craters) imaging the subsurface structures.

Undifferentiated plains: There are vast expanses of plains, mostly at mid-latitudes, that appear relatively homogeneous and radar-dark, and are classified as plains because they are really extensive, relatively featureless, and generally appear to be of low relief. The units may be sedimentary in origin, resulting from fluvial or lacustrine deposition, or they may be cryovolcanic, but their relatively featureless nature and gradational boundaries prevent interpretation using SAR data alone. Radar sounder measurements will determine the layering of the undifferentiated plains to determine if they are sedimentary or cryovolcanic in origin.

Mottled plains: The mottled plains are irregularly shaped terrains having dominantly intermediate backscatter that appear to have relatively small topographic variations. Mottled terrains may be erosional and a mix of several other units, for example, some may be undifferentiated plains covered over in small areas by isolated dunes.

2.2.2.2 Surface Composition

Observations of Titan's surface are severely hindered by the presence of an optically thick, scattering and absorbing atmosphere, allowing direct observation of the surface within only a few spectral windows in the near-infrared. Based on the $1.29/1.08 \text{ }\mu\text{m}$, $2.03/1.27 \text{ }\mu\text{m}$ and $1.59/1.27$

μm band ratios measured by Cassini's Visual and Infrared Mapping Spectrometer (VIMS), three main spectral units can be distinguished on the overall surface of Titan: whitish material mainly distributed in the topographically high areas; bluish material adjacent to the bright-to-dark boundaries; and brownish material that correlates with RADAR-dark dune fields (*Barnes et al.*, 2007; *Soderblom et al.*, 2007; *Jaumann et al.*, 2008). Even though the spectral units are distinct, their actual composition remains elusive. Bright materials may consist of precipitated aerosol dust composed of methane-derived organics (*Soderblom et al.*, 2007) superimposed on water-ice bedrock. The bluish component might contain some water ice as its defining feature (*ibid*), even though organics cannot be ruled out. Finally, the brownish unit reveals lower relative water-ice content than Titan's average that is compatible with an enhancement in complex organic material and/or nitriles clumping together after raining onto the surface (*Barnes et al.*, 2008).

Including water ice, a number of chemical species have been proposed exist on the surface of Titan including: carbon dioxide, ammonia, ammonia hydrate, cyanoacetylene, benzene, hydrogen cyanide, acetylene, acetonitrile, and liquid alkanes. But only a few absorptions have been unambiguously detected in Titan's methane windows from remote sensing observations carried out in Cassini fly-bys. Evidence from VIMS infrared spectra suggests CO_2 frost, which is probably condensed from the atmosphere (*McCord et al.*, 2008). Liquid ethane has been identified in Ontario Lacus (*Brown et al.*, 2008) and, although the exact blend of hydrocarbons in polar lakes and seas is unknown, liquid methane is almost certainly present. Finally, the correlation between the 5- μm -bright materials and RADAR-empty lakes suggests the presence of sedimentary or organic-rich evaporitic deposits in dry polar lakebeds (*Barnes et al.*, 2011).

Spectral mapping of Titan's diverse surface at both high spatial and high spectral resolution will be a major gap in knowledge after Cassini. Cassini's VIMS instrument is expected to cover only a few percent of Titan's surface at kilometer-or-better resolution, even after the extended mission. Thus a prime element in the Orbiter payload is a spectral mapper. In essence this instrument addresses two main goals: high-resolution near-IR imaging and spectroscopy of the surface. Other spectral end-members will likely appear as a dedicated Orbiter mission improves the spatial resolution and coverage by orders of magnitude. An important augmentation beyond Cassini is that the spectral mapper should cover the wavelength range beyond 5 μm (ideally up to 6 μm). This spectral region (in a methane window, thus the atmosphere is quite transparent) is highly diagnostic of several organic compounds expected on Titan's surface. Cassini VIMS data already showed some tentative signatures around 5 μm (possibly associated with aromatic and/or polycyclic compounds like phenanthrene, indole, etc.), but robust identification is impossible due to inadequate spectral resolution. VIMS/Cassini showed that Titan's surface can be observed and mapped at 5 microns (*Sotin et al.*, 2012) despite the low S/N of the VIMS instrument at this wavelength, the small amount of solar flux and the small surface albedo. With the current technologies, detectors in this wavelength domain would provide high resolution mapping as well as diagnostic signatures of several organic compounds likely to be present on Titan's surface either in the dune fields or on the shores of lakes (*Stofan et al.*, 2007; *Barnes et al.*, 2011).

2.2.3 Atmosphere

An instrument suite, enabling both remote and *in situ* measurements and observations of Titan's atmosphere including, for example, lidar and near infrared spectrometers on an orbiter platform and mass spectrometers, physical property (temperature, pressure, humidity, wind speed) sensors and an imaging camera onboard a lake-probe could be used to achieve scientific objective G3. Titan has a dense nitrogen-dominated atmosphere, which is unique among satellite atmospheres with regard to atmospheric dynamics, atmospheric chemistry and hydrological-like cycle. Unlike the other moons in the solar system, Titan has a substantial atmosphere, consisting of 98.4% nitrogen and 1.5% methane as well as trace amounts of hydrocarbons such as ethane, acetylene, and diacetylene, benzene and nitriles, such as hydrogen cyanide (HCN), cyanogen (C_2N_2) and cyanoacetylene (HCCCN). Somewhat more complex molecules such as propane, butane,

polyacetylenes, and complex polymeric materials follow from these simpler units. Cassini/Huygens mission detected and measured many of the constituents of Titan's atmosphere. Cassini mission has revealed the presence of large organic molecules with high molecular masses above 100 amu in the thermosphere and ionosphere. This is an indication of the photochemical production of Titan's aerosols and their upper atmosphere origin. While the identities of these molecules are still unknown, their presence suggests a complex atmosphere that could hold the precursors for biological molecules such as those found on Earth.

The thermal structure of Titan's atmosphere is similar to the Earth's, with a troposphere, a stratosphere, a mesosphere and a thermosphere; however these regions develop on a larger scale height due to Titan's lower gravity and lower temperatures. A detached layer of haze, the result of photochemical reactions in the upper atmosphere, is visible at altitudes greater than ~520 km (*Porco et al.*, 2005). The Voyager spacecraft showed the vertical structure of the haze to consist of a main layer, located at an altitude below 200 km, separated from a detached layer of haze at an altitude of ~500 km by a gap of approximately 300 km; these two layers merge near the North pole (*Porco et al.*, 2005). The extended haze layer has been found to have reduced its altitude from 500 km to ~380 km between 2007 and 2010 (*West et al.*, 2011). Cassini's images also confirmed the existence of bright clouds in Titan's atmosphere, which were already inferred from ground-based observations (*Brown et al.*, 2002). Titan's clouds are probably composed of methane, ethane, or other simple organics. Their presence seems scattered and variable, interspersing the overall haze.

A striking feature of Titan's atmospheric dynamics is the strong stratospheric super-rotation in either hemisphere (*Achterberg et al.*, 2011). The global profile of zonal winds between 100 and 500 km altitude can be reconstructed from the temperature data retrieved from Cassini CIRS (Composite Infrared Spectrometer), except very close to the equator (*Achterberg et al.*, 2011). On the other hand, the wind profile below 100 km altitude is known only at the entry site of the Huygens Probe. Huygens' Doppler Wind Experiment (*Bird et al.*, 2005) found that the superrotation gradually weakens below 150 km altitude, but it also detected an unexpected layer of extremely weak winds near 80 km altitude and a complex wind pattern near the surface. This demonstrates that *in situ* measurements can detect wind patterns especially in the troposphere that can even not be imagined from remote measurements in other parts of the atmosphere. The wind speed and direction in selected areas of the troposphere can be constrained by cloud tracking (*Turtle et al.*, 2011a) but only where clouds exist. The observed orientation and appearance of dunes have been used to constrain the surface wind, but this led to ambiguous interpretations on the prevailing surface wind direction (*Lorenz and Radebaugh*, 2009; *Rubin and Hesp*, 2009; *Tokano*, 2010). The wind direction near the surface and its possible seasonal variation are crucial measurements for the angular momentum budget of the atmosphere as well as length-of-day variation of the surface (*Tokano and Neubauer*, 2005). It is unknown from atmospheric observations whether there is an Earth-like large angular momentum exchange between the atmosphere and surface, which is equilibrated on an annual average, or there is almost no angular momentum exchange. Geodetic observations (Section 2.1) should ideally be accompanied with wind observations to establish a link between possible atmospheric forcing and response of the surface. Near-surface wind measurements by a lake lander will be meaningful in this regard since the wind at high latitudes is more indicative of seasonal wind fluctuations than equatorial winds and there are no dunes at high latitudes that could be used to guess the wind direction.

Another major characteristic of Titan's meteorology is the methane weather, which most impressively manifests itself in the visible clouds (*Roe*, 2012). During southern summer most convective clouds were seen at high southern latitudes but the main cloud activity moved northward around the recent equinox (*Schaller et al.*, 2009; *Turtle et al.*, 2011a). More persistent clouds at higher altitudes were seen near the north pole (*Rodriguez et al.*, 2009). For many years, the equatorial region was devoid of visible clouds and thus was expected to be dry. However, the vertical profile of temperature and pressure (*Fulchignoni et al.*, 2005) and methane mole fraction

(Niemann *et al.*, 2005) measured by Huygens found that methane condensation and weak precipitation in the form of drizzles occur even near the equator where no visible cloud was present (Tokano *et al.*, 2006). On the other hand, occurrence of convective clouds strongly depends on the global circulation pattern, static stability and methane humidity in the lower troposphere (Mitchell *et al.*, 2011; Schneider *et al.*, 2012). While the vertical temperature profile down to the surface can be sounded by radio occultation (Schinder *et al.*, 2012), the methane humidity is difficult to measure remotely. *In situ* measurement of the near-surface methane humidity in the polar region would therefore provide valuable insight into the mechanism behind the seasonality of Titan's methane weather. There is also evidence of substantial methane rainfall on Titan, but the evidence is indirect in that it came from temporary surface darkening and temporal correlation with a cloud system (Turtle *et al.*, 2011b). The precipitation amount is unknown despite its huge role it plays for Titan's geomorphology. Only direct measurements of rainfall on the surface would unambiguously show in which seasons it does or does not rain in the polar region, how much it rains and the chemical composition of the rains.

The meridional circulation pattern in the troposphere predicted by general circulation models (GCMs) differs from model to model, particularly at high latitudes. To some extent the global statistics of clouds constrain the meridional circulation pattern (Rodriguez *et al.*, 2009). However, clouds do not develop in the upwelling branch of the circulation if the near-surface air is too dry and not every condensation causes visible clouds. Therefore, a more certain quantity to verify the meridional circulation at high latitudes is the surface pressure, which can be compared to the *in situ* surface pressure data at the Huygens landing site (Fulchignoni *et al.*, 2005) after a sea-level correction. A surface pressure higher and lower than the Huygens pressure would indicate downwelling and upwelling, respectively, in the season of observation.

Theoretical studies predict some significant changes when the north polar lake area approaches summer during the next few years. Seasonal change in the global circulation pattern may cause substantial precipitation in the north polar region (Schneider *et al.*, 2012), warming of the sea surface may cause a sea breeze/land breeze depending on the sea composition (Tokano, 2009), freshening of the near-surface wind may cause previously undetected waves on the seas (Hayes *et al.*, 2013) and evaporation of a large amount of methane from the seas might even cause tropical cyclones (Tokano, 2013). These meteorological features cause substantial surface pressure variations and thus can be detected *in situ* by continuous surface pressure measurements and optical monitoring of the environment. The unknown exact chemical composition in the sea as well as in the air over sea is important for many questions related to the polar region. Evaporation of methane from a sea itself is difficult to observe *in situ* unless accompanied with fog, but it can be quantified from the methane vapor pressure and methane partial pressure. The mole fractions of other species in the sea and air over sea would indicate whether the lake and atmosphere are in short-term or long-term chemical equilibrium, if any.

GCMs predict gravitational tides and various waves in the troposphere, which affect the wind, pressure and temperature near the surface (Tokano and Neubauer, 2002; Mitchell *et al.*, 2011; Lebonnois *et al.*, 2012). Some of the waves may become visible as clouds but most other waves may not. The most reliable data to verify these waves are *in situ* time series of surface wind and atmospheric pressure over several days. The various waves have different frequencies and amplitudes and may thus be recognized in the power spectra of the time series.

2.2.4 In-Situ Exploration of a Titan Sea

The Cassini spacecraft has unveiled a world that is both strange and familiar to our own, with vast equatorial dune fields (e.g., Lorenz *et al.*, 2006), well-organized channel networks that drain from mountains into large basins (e.g., Burr *et al.*, 2013) and, perhaps most astonishingly, lakes and seas (e.g., Stofan *et al.*, 2007) filled with liquid hydrocarbons (Brown *et al.*, 2008). Titan is the only extraterrestrial body currently known to support standing bodies of liquid on its surface and, along with Earth and Mars, is one of only three places in the solar system that we know to

posses or have possessed an active hydrologic system (e.g., *Atreya et al.*, 2006). Just as Earth's history is tied to its oceans, Titan's origin and evolution is chronicled within the nature of its lakes and seas and their interactions with the atmosphere and surface. While Cassini has provided a wealth of information regarding the distribution of liquid deposits on Titan (e.g., *Hayes et al.*, 2008; *Turtle et al.*, 2009; *Aharonson et al.*, 2009), it has provided only the most basic information regarding their composition (e.g., *Brown et al.*, 2008) and role in Titan's volatile cycles (e.g., *Lunine and Atreya*, 2008). In order to address these fundamental questions, which link the lakes to Titan's origin and evolution, we must visit them in-situ. In-situ exploration of Titan's lakes and seas presents an unprecedented opportunity to **understand its hydrocarbon cycles** through the abundance and variability of the liquid's compositional components (methane, ethane, propane, etc.), **glimpse into the history of the moon's evolution** as chronicled by noble gas abundance (argon, krypton, xenon) and isotopic composition (e.g., C^{12} vs. C^{13}), **investigate a natural laboratory for prebiotic chemistry and the limits of life** through an examination of trace organics, and **study meteorological and marine processes in an exotic environment** that is drastically distinct from terrestrial experience (Scientific Objective G1, G3, and G4).

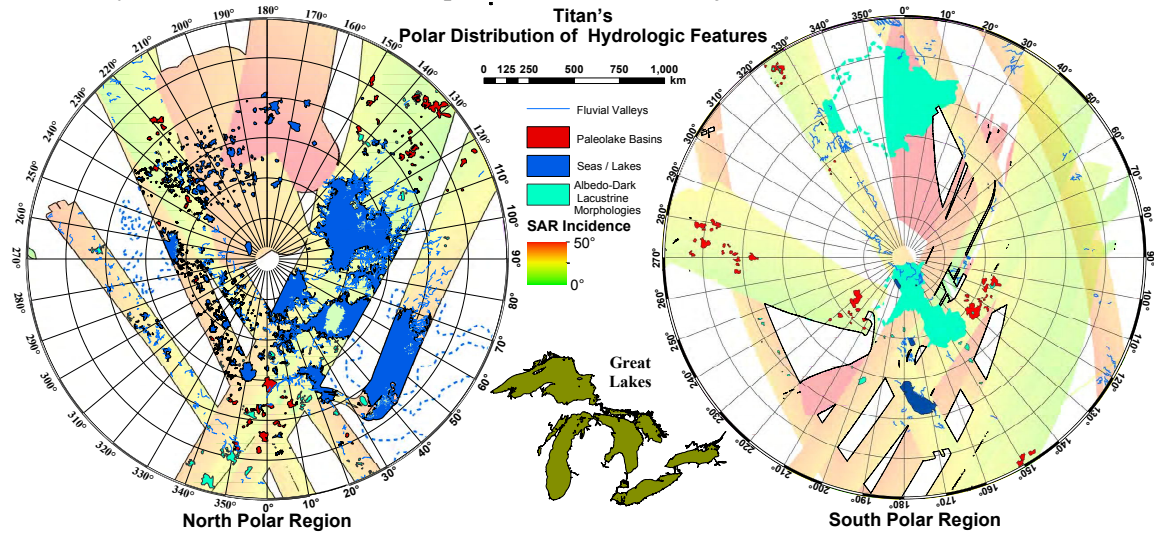


Figure 2. Distribution of lacustrine and fluvial morphologies in Titan's polar regions. Blue polygons depict modern lakes and seas filled with liquid hydrocarbons. The red polygons show the locations of radar-bright basins interpreted as paleolakes. Cyan polygons depict the distribution of albedo-dark features that are above the radar noise floor and express lacustrine-like morphologies. These features are believed to represent anything from mud flats to shallow liquid deposits. Fluvial valleys are represented by blue lines. Dashed lines were drawn on ISS images while the remaining polygons were created using Synthetic Aperture Radar (SAR) images generated from Cassini RADAR observations. SAR incidence angle and swath coverage is shown.

Titan's lakes are found poleward of 55° latitude and encompass 1.2% of the surface that has been observed ($\sim 50\%$) by Cassini's instruments (*Hayes et al.*, 2008, 2011). Lacustrine features are predominantly found in the North, where 11% of the observed area poleward of $55^\circ N$ ($\sim 60\%$) is covered by radar-dark lakes. Poleward of $55^\circ S$, only 0.5% of the observed area ($\sim 60\%$) has been interpreted as potentially liquid-filled. This dichotomy has been attributed to Saturn's current orbital configuration, in which Titan's southern summer solstice occurs nearly coincident with perihelion and results in a 25% higher peak flux than what is encountered during northern summer (*Aharonson et al.*, 2009). Over many seasonal cycles, this flux asymmetry can lead to net transport of volatiles (methane/ethane) from the south to the north. As the orbital parameters shift, however, the net flux of northward-bound volatiles is expect to slow and eventually reverse, predicting a liquid distribution opposite of today's in ~ 35 kyr (i.e., more liquid in the south than

the north). If this hypothesis is correct, the distribution of liquid deposits on Titan is expected to move between the poles with a period of ~ 50 kyr in a process analogous to Croll-Milankovich cycles on Earth. In-situ measurement and comparison between the relative abundance of volatiles that are mobile over these timescales (e.g., methane, ethane) versus those that are involatile (e.g., propane, benzene), can be used to test this hypothesis and understand volatile transport on thousand year timescales. Volatile transport over shorter timescales (diurnal, tidal, and seasonal) can be investigated via in-situ measurements of the methane evaporation rate and associated meteorological conditions (e.g., wind speed, temperature, humidity). These measurements can be used to ground-truth methane transport predictions from global climate models (e.g., *Mitchell, 2008; Tokano, 2009; Schneider et al., 2012*).

In the north, 87% of the area of observed lacustrine deposits are contained within the three largest lakes, Ligeia, Kraken, and Punga Mare, which are similar in size to the Great Lakes of the continental United States (**Figure 2**). The remainder of north polar lakes approximately follow a lognormal distribution with a mean diameter of ~ 10 km (*Hayes et al., 2008*). While Cassini has provided a wide range of evidence that the largest of Titan's lakes are indeed liquid-filled basins (e.g., *Brown et al. 2008; Stephan et al., 2010*), the returned data have only been able to provide lower limits on their depth that are highly dependent on the unknown dielectric properties of the both the liquid and the lakebed (e.g., *Paillou et al., 2008; Hayes et al., 2010, 2011*). In the absence of a substantial subsurface reservoir, these Mare represent the dominant liquid hydrocarbon inventory on Titan. Prior to Cassini, a global ocean of liquid hydrocarbon was predicted to cover the surface and act as both a source of atmospheric methane and sink for photolysis products such as ethane and higher order hydrocarbon/nitriles (*Lunine et al., 1983*). The absence of this global reservoir questions the long-term stability of Titan's atmosphere, which would be depleted in $\sim 10^7$ yrs in the absence of a crustal source or methane outgassing from the interior (e.g., *Tobie et al., 2006*). Determining the bathymetry of one of Titan's Mare, which can only be reliably accomplished in-situ, is essential to understanding the total volume of liquid available for interaction with the atmosphere. The inventory of methane in Titan's Mare, which requires knowledge of both depth and composition, will provide a lower limit on the length of time that the lakes can sustain methane in Titan's atmosphere (*Mitri et al., 2007*) and help to quantify the required rate of methane resupply from the interior and/or crust. Similarly, the absolute abundance of methane photolysis products (e.g., ethane, propane) will determine a lower limit for the length of time that methane has been abundant enough to drive photolysis in the upper atmosphere and deposit its products onto the surface and, ultimately, into the lakes. Furthermore, a comparison amongst the relative abundance of photolysis products could reveal fractionation processes such as crustal sequestration, as would happen if one constituent (e.g., propane) were preferentially introduced into a repository such as crustal clathrate hydrate.

Similar to the Earth's oceans, Titan's seas record a history of their parent body's origin and evolution. Specifically, the noble gas and isotopic composition of the sea can provide information regarding the origin of Titan's atmosphere, reveal the extent of communication with the interior, potentially constrain the conditions in the Saturn system during formation, and refine estimates of the methane outgassing history. The Huygens probe detected a small argon component in Titan's atmosphere, but was unable to detect the corresponding abundance of xenon and krypton (*Niemann et al., 2005*). The non-detection of xenon and krypton supports the idea that Titan's methane was generated by serpentinization of primordial carbon monoxide and carbon dioxide delivered by volatile depleted planetesimals originating from within Saturn's subnebula (e.g., *Atreya, 2006*). A competing model of methane origination, however, suggests that methane is primordial and was sourced from planetesimals originating from the outer solar nebula (*Mousis et al., 2009*). This competing model was originally suggested as an attempt to explain the measured ratio of deuterium/hydrogen (D/H) (1.32×10^{-4}), which *Mousis et al. (2009)* suggest would require a D/H ratio in Titan's water ice that does match that of Enceladus. In order to support a primordial methane source, xenon and krypton would both have to be sequestered from the

atmosphere. While xenon is soluble in liquid hydrocarbon (solubility of 10^{-3} at 95 K) and could potentially be sequestered into liquid reservoirs, argon and krypton cannot. Therefore, the absence of measureable atmospheric krypton requires either sequestration into non-liquid surface deposits, such as clathrates (Mousis *et al.*, 2011), or depletion in the noble gas concentration of the planetesimals that originally delivered Titan's atmosphere (Owen and Niemann, 2009). The low, albeit measureable, argon-to-nitrogen ratio has been used to suggest that Titan's atmospheric nitrogen was sourced from ammonia, as ammonia is significantly less volatile than both molecular nitrogen and argon (Niemann *et al.*, 2005), which have similar volatility. The presence of xenon in Titan's seas would indicate that the absence of atmospheric krypton is a result of sequestration into crustal sources (Mousis *et al.*, 2011) as opposed to the result of carbon delivery via depleted planetesimals (Owen and Niemann, 2009). In addition to noble gas concentration, isotopic ratios can also be used to decipher the history of Titan's atmosphere. For example, the $^{13}\text{C}/^{12}\text{C}$ ratio of CH_4 was used by Mandt *et al.* (2009) to conclude that methane last outgassed from the interior $\sim 10^7$ years ago. However, this calculation assumes that the exposed methane reservoir has an isotopic composition that is in equilibrium with the atmosphere. If the carbon isotope ratio of hydrocarbons in Titan's lakes/seas were found to be different than in the atmosphere, it would imply alteration of the isotopic composition by evaporation and condensation and indicate a different timescale for the history of methane-outgassing.

Titan's lakes and seas collect organic material both directly, through atmospheric precipitation of photolysis products, and indirectly, through aeolian / fluvial transport of surface materials (e.g., river systems flowing into the Mare). As a result, the lakes and seas represent the most complete record of Titan's organic complexity and present a natural laboratory for studying prebiotic organic chemistry. Titan's environment is similar to conditions on Earth four billion years ago and presents an opportunity to study active systems involving several key compounds of prebiotic chemistry (Raulin, 2008). For instance, the complex aerosols produced in Titan's atmosphere via photolysis, which are presumably transported into the lake via pluvial, fluvial, and aeolian processes, are similar to the initial molecules implicated in the origin of life on Earth. Other organics can act as tracers for specific environmental conditions. For example, the presence of oxygen-bearing organic species would suggest that redox reactions have taken place and that the organic deposits have, at some point, been in at least transient contact with liquid water. In general, simple abiotic organic systems exhibit monotonically decreasing abundance with increasing chain complexity. Distinctive deviations from such a distribution can mark the signs of prebiotic chemistry and potentially represent the initial markers on a pathway toward self-replication and a minimally living system.

Finally, lacustrine settings on Titan represent exotic environments that are being sculpted by many of the same meteorological and oceanographic processes that effect lake and ocean settings on Earth. *In situ* exploration of such an environment provides an opportunity to study these processes under vastly different environmental conditions (Scientific Objectives G2 and G4). While the Huygens probe provided atmospheric data over equatorial latitudes around winter solstice, models of Titan's climate predict that polar weather could be quite different from the equatorial environment investigated by Huygens. In-situ measurements of such an environment would provide invaluable ground-truth for Global Climate Models (GCM) of Titan's surface/atmosphere interactions. Ultimately, this could lead to a better understanding of how to accurately predict the evolution of our own climate system and even how to model extreme climates on extrasolar planets, as Titan may represent a common climate system amongst solid bodies in Earth-like orbits around M-dwarf stars.

3. Mission Concept

Key questions to be resolved during future exploration of Titan are as follows: 1) "What are the complex organic molecules that form the dune material, and how have they been formed?" 2) "What is the composition of the lakes and seas?" 3) "What are the processes that form the

atmosphere?” and 4) “Is Titan endogenically active?” In spite of all its advances, the Cassini-Huygens cannot however address these fundamental questions due to its limited instrumentation and timeline, among others. In fact, these central themes can be investigated by a future mission using an orbiter and a lake-probe.

The proposed mission concept consists of two elements, the spacecraft (S/C), and a lake-probe (lander). The S/C will carry a scientific payload consisting of remote sensing experiments and an Entry, Descent and Landing (EDL) module containing a lake lander equipped with an instrument suite capable of carrying out *in situ* measurements of Titan’s polar lakes. During the descent the probe will make *in situ* measurements of the atmosphere. Once a successful splashdown has been achieved, the lake lander will take measurements, including the sampling of both the liquid of the lake and the low-atmosphere. A valuable augmentation may be the capability to sample solid material on a beach or on the shallow sea-bed. The possible targets of the lake probe are Ligeia Mare (78°N, 250°W) and Kraken Mare (68°N, 310°W). Previous mission analysis devoted to the exploration of Titan using an orbiter and a lake-probe demonstrated the feasibility of such a mission. The concept for a mission capable of implementing the science described in previous sections is in part based on the Titan and Enceladus Mission TandEM (Coustenis *et al.*, 2009), Titan Saturn System Mission Study TSSM (Coustenis *et al.*, 2010) and the Titan Mare Explorer (TiME) (Stofan *et al.*, 2010). Some additional considerations on science from a Titan orbiter were explored in the detailed 2007 NASA Titan Explorer flagship mission study (Leary *et al.*, 2007).

ESA has indicated that the L2 and L3 missions have tentative launch opportunities respectively in 2028 and 2034. Our proposed mission profile is compatible with those launch opportunities, according to preliminary analysis. The duration of the cruise from Earth to Saturn is estimated to be less than 10 years. Following orbit insertion at Saturn, the S/C will conduct a tour of the satellite system, with the possibility of several flybys at Enceladus, while flybys at Titan would be used both as a scientific opportunity and as a way to modify the trajectory in order to finally enter orbit around Titan. A different mission scenario that is under analysis proposes that the orbit around Titan will be achieved through an aero-capture maneuver, followed by other maneuvers to refine the spacecraft’s final 1500 km circular high-inclination orbit. Release and entry of the lake-probe, and relay of its data to Earth will take place during 1) Titan flybys phase, or 2) Titan orbital phase. The architecture of the mission will be defined in future studies. We considered as a possible nominal final orbit of the S/C (orbiter) around Titan a 1500 km altitude circular near-polar orbit. A polar orbit optimizes the gravity field measurements and maximizes the imaging coverage of the surface. A 1500 km altitude orbit has negligible drag effect on the S/C; however lower altitude orbits are also possible increasing the orbit maintenance. After orbit insertion around Titan, mapping of Titan with remote sensing experiments will begin, lasting several months to a few years. The lake lander will relay data to the orbiter, which will serve as the communications link between the probe and Earth. Direct to Earth (DTE) communication from the lake lander (feasible with Earth view from northern polar seas starting from 2036 to 2050s and later) allows augmented communication and tracking. Scenarios for the lake probe are 1) The lake lander will have propulsion capabilities and mobility; 2) The lake lander will not have propulsion capabilities rather it will sail around the lake driven by winds and possible tides.

The power source both for the S/C and the lake probe is critical for this ESA-led mission on Titan and the Saturn system. Preliminary analysis indicates that, to meet power requirements through solar power, the orbiter would need large solar arrays (total surface 130 – 200 m²). Given the opacity of Titan’s atmosphere and the low sun elevation from the polar seas, the use of a solar powered generator for the lake-probe is unfeasible. The use of batteries for the lake probe will only provide power to the lander on the order of hours. The development of radioisotope space power systems which minimize or replace ²³⁸Pu fuel within the European Space Agency (ESA) funded programs includes the development of ²⁴¹Am fueled Radioisotope Thermoelectric Generator (RTG) (e.g. Ambrosi *et al.*, 2012) that should be considered as a possible power source for a ESA-led Titan mission.

References

- Achterberg, R.K. et al. 2011. *Icarus* 211, 686-698.
- Aharonson, O. et al. 2009. *Nature Geo.* 2, 851-854.
- Ambrosi R.M. et al. 2012. *Nuclear and Emerging Technologies for Space*, 3043.
- Andrews, D.J. et al. 2010. *J. Geophys. Res.* 115, A04212.
- Atreya, S.K. et al. 2006. *Planet. and Space Science* 54, 1177-87.
- Barnes, J.W. et al. 2006. *Geophys. Res. Lett.* 33, L16204.
- Barnes, J.W. et al. 2007. *Icarus* 186, 242-258.
- Barnes, J.W. et al. 2008. *Icarus* 195, 400-414.
- Barnes, J.W. et al. 2011. *Icarus* 216, 136-140.
- Bills, B.G., Nimmo, F. 2008. *Icarus* 196, 293-296.
- Bird, M. K. et al. 2005. *Nature* 438, 800-802.
- Brown, R.H. et al. 2008. *Nature* 454, 607-610.
- Brown, M.E. et al. 2002. *Nature* 420, 795-797.
- Burr D. et al. 2013. *GSA Bulletin* 125, 299-321.
- Castillo-Rogez, J., Lunine, J.I. 2010. *Geophys. Res. Lett.* 37, L20205.
- Cicalò, S., Milani, A. 2012. *Royal Astron. Society* 427, 468-482.
- Coustenis, A. et al. 2009. *Experimental Astronomy* 23, 893-946
- Coustenis, A. et al. 2010. *Lunar and Planetary Science Conf.* 1060.
- Elachi, C. et al. 2004. *Space Science Reviews* 115, 1, 71-110.
- Folkner, W.M. et al. 1997. *Science* 278, 5344, 1749-1752.
- Fortes, A. D. et al. 2007. *Icarus* 188, 139-153.
- Fulchignoni, M. et al. 2005. *Nature* 438, 785-791.
- Hayes, A.G. et al. 2008. *Geophys. Res. Lett.* 35, L09204.
- Hayes, A.G. et al. 2010. *J. Geophys. Res.* 115(E9), E09009.
- Hayes, A.G. et al. 2011. *Icarus* 211, 655-671.
- Hayes, A. G., et al. 2013. *Icarus* in press.
- Iess, L. et al. 2010. *Science* 327, 1367.
- Iess, L. et al. 2012. *Science* 337, 457.
- Jaumann, R. et al. 2008. *Icarus* 197, 526-538.
- Khurana, K.K. et al. 1998. *Nature* 395, 749.
- Khurana, K.K. et al. 2011. *Science* 332, 1186-1189.
- Kivelson, M.G. et al. 2002. *Icarus* 157, 507.
- Leary, J. et al. 2007. *AGU Fall Meeting* P53A-1007
- Lebonnois, S. et al. B. 2012. *Icarus* 218, 707-722.
- Lopes, R.M.C. et al. 2007. *Icarus* 186, 395-412.
- Lopes, R.M.C. et al. 2010. *Icarus* 205, 540-588.
- Lopes, R.M.C. et al. 2013. *J. Geophys. Res.* 118, 1-20.
- Lorenz, R.D. et al. 2006. *Science* 312, 724-727.
- Lorenz, R.D. et al. 2007. *Geophys. Res. Lett.* 34, L07204.
- Lorenz, R.D., Radebaugh, J. 2009. *Geophys. Res. Lett.* 36, L03202.
- Lunine, J.I. et al. 1983. *Science* 222, 1229-1230.
- Lunine, J.I., Atreya, S. 2008. *Nature Geo.* 1, 335.
- Malaska, M. et al. 2010. *Lunar Planetary Science Conference*, 1544.
- Mandt, K. E. et al. 2009. *Planetary and Space Science*, 57, 1917-1930.
- McCord, T.B. et al. 2008. *Icarus* 194, 212-242.
- Meriggiola, R., Iess, L. 2012. *European Planetary Science Congress* 7, 2012-593.
- Mitchell, J.L. et al. 2008. *J. Geophys. Res.* 113(E8), E08015.
- Mitchell, J.L. et al. 2011. *Nat. Geosci.* 4, 589-592.
- Mitri, G. et al. 2007. *Icarus* 186, 385-394.
- Mitri, G., Showman A.P. 2008. *Icarus* 193, 387-396.

Mitri, G. et al. 2013. *Icarus* submitted.

Mitri, G. et al. 2010. *J. Geophys. Res.* 115, 10002.

Moore, J.M., Pappalardo, R.T. 2011. *Icarus* 212, 790-806.

Mousis O. et al. 2009. *Icarus* 204, 749-751.

Mousis O. et al. 2011. *The Astrophysical Journal Letters* 740, L9.

Neish, C.D. et al. 2013. *Icarus* 223, 82-90.

Nelson, R.M. et al. 2009a. *Icarus* 199, 429-441.

Nelson, R.M. et al. 2009b. *Geophys. Res. Lett.* 36, L04202.

Niemann, H. B., et al. 2005. *Nature* 438, 779-784.

Nimmo, F., Bills, B.G. 2010. *Icarus* 208, 896-904.

Owen, T., Niemann, H.B. 2009. *MPES* 367, 607-615.

Paillou, P. et al. 2008. *Geophys. Res. Lett.* 35, L18202.

Porco, C.C. et al. 2005. *Nature* 434, 159-168.

Radebaugh, J. et al. 2007. *Icarus* 192, 77-91.

Radebaugh, J. et al. 2008. *Icarus* 194, 690-703.

Raulin, F. 2008. *Nature* 454, 587-589.

Rodriguez, S. et al. 2009. *Nature* 459, 678-682.

Roe, H. G. 2012. *Annu. Rev. Earth Planet. Sci.* 40, 355-382.

Rubin, D. M., Hesp, P. A. 2009. *Nat. Geosci.* 2, 653-658.

Schaller, E. L. et al. 2009. *Nature* 460, 873-875.

Schinder, P. J. et al. 2012. *Icarus* 221, 1020-1031.

Schneider, T. et al. 2012. *Nature* 481, 58-61.

Seidelmann, P.K. et al. 2006. *Celestial Mechanics and Dynamical Astronomy* 98, 155-180.

Soderblom, L.A. et al. 2007. *Planetary and Space Science* 55, 2025-2036.

Soderblom, L. et al. 2009. *Icarus* 204, 610-618.

Sotin, C. et al. 2005. *Nature* 435, 786-789.

Sotin, C. et al. 2012. *Icarus* 221, 768-786.

Stephan, K. et al. 2010. *Geophys. Res. Lett.* 37, L07104.

Stiles, B.W. et al. 2008. *The Astronomical Journal* 135, 1669-1680.

Stiles, B.W. et al. 2009. *Icarus* 202, 584-598.

Stofan, E. 2007. *Nature* 445, 61-64.

Stofan, E.R. et al. 2010. *Lunar and Planetary Science Conf.* 5270.

Tobie, G. et al. 2006. *Nature* 440, 61-64.

Tobie, G. et al. 2005. *Icarus* 175, 496-502.

Tokano, T. 2009. *Icarus* 204, 619-636.

Tokano, T. 2010. *Aeolian Res.* 2, 113-127.

Tokano, T. 2013. *Icarus* 223, 766-774.

Tokano, T., Neubauer, F. M. 2002. *Icarus* 158, 499-515.

Tokano, T., Neubauer, F. M. 2005. *Geophys. Res. Lett.* 32, L24203.

Tokano, T. et al. 2006. *Nature* 442, 432-435.

Turtle, E. et al. 2009. *Geophys. Res. Lett.* 36, L02204.

Turtle, E.P et al. 2011a. *Geophys. Res. Lett.* 38, L03203.

Turtle, E. P. et al. 2011b. *Science* 331, 1414-1417.

Wall, S.D. et al. 2009. *Geophys. Res. Lett.* 36, L04203.

Wall, S.D. et al. 2010. *Geophys. Res. Lett.* 37, L05202.

Whar, J. et al. 2006. *J. Geophys. Res.* 111, E12005.

Wei, H.Y. et al. 2010. *J. Geophys. Res.* 115, E10007.

West, R.A. et al. 2011. *Geophys. Res. Lett.* 38, L06204.

Wood, C.A. et al. 2010. *Icarus* 334-344.

Zebker, H. A. et al. 2009. *Science* 324, 921-923.

Zimmer, C. et al. 2000. *Icarus* 147, 329.

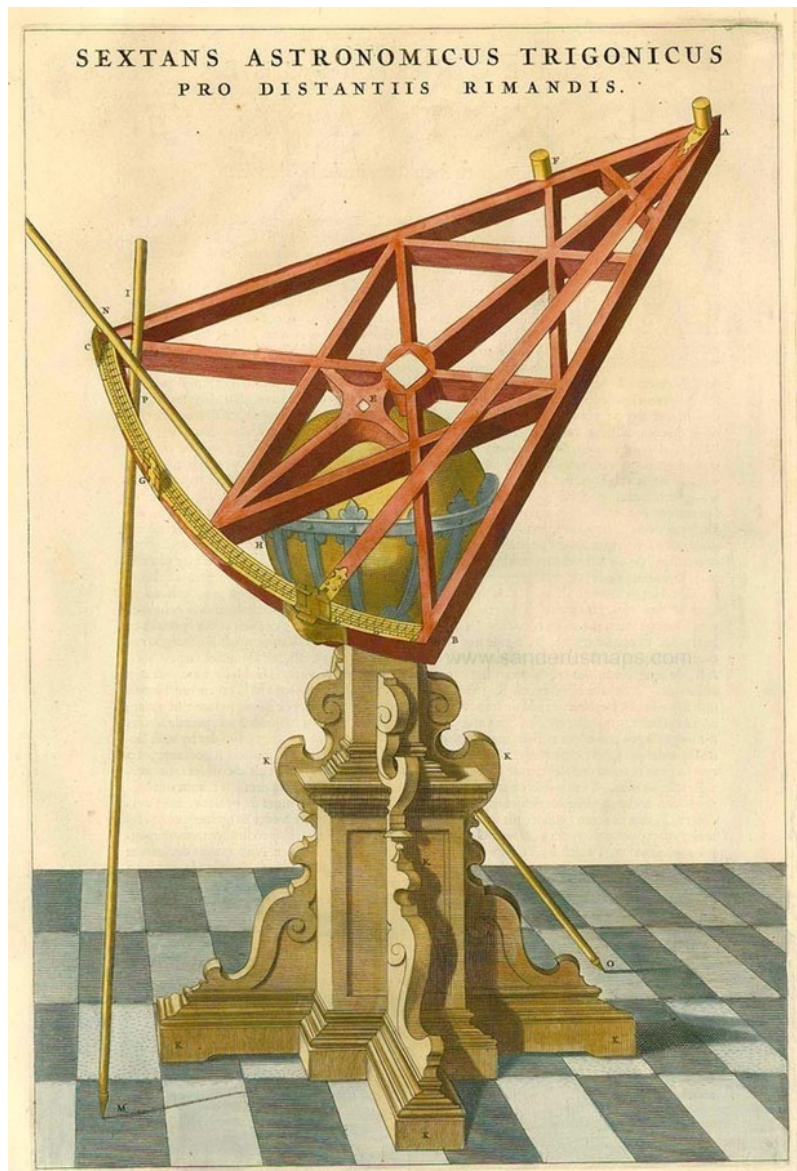
Astrometry for Dynamics

A response to the call by ESA for *White Papers* for the definition of *Large* missions

Contact details:

Erik Høg
Skovsøen 23
2880 Bagsværd
Denmark
Tlf: +45 4449 2008
Email: erik.hoeg@get2net.dk

Erik Høg - emeritus of
Niels Bohr Institute
Juliane Maries Vej 30
2100 Copenhagen Ø
Denmark
Website: www.astro.ku.dk/~erik



This page should list supporters for the proposal, but I will not name any because this proposal is not meant to compete with any other proposal, only to be a possible supplement.

Acknowledgements: I am indebted to Claus Fabricius for drawing my attention to the problem of low resolution in Gaia photometry in a correspondence about Gaia, and furthermore for mentioning the ESA call for proposals. I am also grateful to Floor van Leeuwen for his prompt answer to my question about the low resolution, quoted in the text, and to Frederic Arenou, Ulrich Bastian, Anthony Brown, Nicole Capitaine, Jørgen Christensen-Dalsgaard, Dafydd Wyn Evans, Laurent Eyer, Claus Fabricius, Gerry Gilmore, Carme Jordi, Uffe Gråe Jørgensen, Sergei Klioner, Jens Knude, Floor van Leeuwen, Dimitri Pourbaix, Annie Robin and Chris Sterken for valuable comments to a previous version. I am especially grateful to Lennart Lindegren for writing: "*I strongly encourage you to submit the proposal, which I think is definitely a very viable concept*" and "*The argument for high angular resolution photometry is compelling, for it is certainly one of the weak points of Gaia.*"

Cover figure: Sextans Astronomicus Trigonicus by Tycho Brahe

Astrometry for Dynamics

Proposal by Erik Høg, Niels Bohr Institute, Copenhagen, Denmark

23 May 2013

Preamble

This proposal of a Gaia-like mission may possibly supplement the proposal expected from Anthony Brown on behalf of the GST, and it is not meant to compete. The proposal is so short because I saw the call for proposals only on 13 May while on vacation. With the deadline of 24 May for submission, a draft was distributed after a few days, and two days were left for an email discussion with colleagues who are anyway very busy, four months before the launch of Gaia.

ABSTRACT: Studies of the kinematic and dynamics of our Galaxy, of nearby galaxies and of our own and other planetary systems require very accurate positions obtained over long periods of time. Such studies would greatly benefit from a new Gaia-like mission. It would capitalize on the experience in Europe with Hipparcos and Gaia and on the results from these missions. Proper motions with much smaller formal errors than from Gaia alone can be obtained for one billion stars from the positions observed by Gaia and a new mission. Equally important, these proper motions will be much less affected systematically by motions in the very frequent unresolved binaries. Studies of our Galaxy and nearby galaxies will benefit from the proper motions. The study of orbits for over 100,000 objects in the solar system will profit greatly when it can be based on the positions from two global astrometry space missions by ESA. It is proposed that the photometry with low-dispersion spectra is replaced by filter photometry in 3 or 4 bands. This will provide photometry of all stars, sufficient for the required chromatic corrections of astrometry and it will give photometry of narrow double stars which cannot be obtained from the ground nor can it be obtained with Gaia because the long spectra of the two stars overlap. Ground-based surveys of multi-colour photometry and spectra will be available for astrophysical studies for a large fraction of the stars.

Infrared and Nanoarcseconds missions

Infrared and Nanoarcseconds missions for astrometry are worth considering, but I will leave the detailed discussion to others. An infrared mission concept, JASMINE, has been studied by Japanese colleagues for more than ten years and a small version, Nano-JASMINE is due for launch in November 2013 and has a 5 centimeter primary mirror. But Nano-Jasmine is NOT infrared, it is a CCD-based version of Hipparcos (scanning satellite on an Earth-centered circular orbit). It is the first of a planned series of three satellites of increasing size: Nano-JASMINE, Small-JASMINE, and JASMINE. The two last satellites shall observe in the infrared, allowing for better observation towards the center of the Galaxy.

It is my impression from the discussion I have seen or heard of such missions that they have great potential. I consider the nanoarcseconds region to be an important long-term goal for space astrometry in the effort to serve the needs of astrophysics. But it would be too risky to focus only on this goal for a mission already around 2030 in view of all the technical challenges, known as well as unknown ones. The astrometric community cannot soon focus on the development of a difficult new mission because the Gaia mission will take the attention in the years to come. Instead

of such a single goal for astrometry, ESA should capitalize on the experience from Hipparcos and Gaia and plan for a Gaia-like mission about 2030. This would provide excellent and unique science data as I shall outline below and the mission cost may be lower because of the experience with Gaia. The mission study should include a consideration of some infrared detection capability in an otherwise Gaia-like mission in order to get closer to the center of the Galaxy. All this would require that ESA plans for astrometry on a time scale of 40 years.

Gaia-like mission

A Gaia-like mission is proposed in response to the call by ESA (2013). It should have a scientific performance as expected for Gaia, according to the section included below, but exceeding this where further considerations during the coming years will show the technical possibilities combined with deeper consideration of the scientific wishes. At present, two improvements are pointed out: (1) An observing epoch about 2031 for an L2 mission would provide large epoch differences from the Hipparcos mission about 1991 and the Gaia mission about 2016, (2) the proposed introduction of filter photometry instead of low-dispersion spectra could provide 3 or 4-colour photometry of all stars, especially also of narrow double stars.

An epoch difference of 15 years from Gaia means that the proper motions derived from the positions obtained at each epoch will have formal standard errors 7.5 times smaller than those from Gaia alone. This follows from the performance figures given on the Gaia website. The parallaxes from both missions will also improve because the proper motions will no longer interfere with the parallaxes.

Proper motions derived from positions observed in a short interval of time, e.g. from one space mission are very often affected by systematic errors due to the motions in unresolved binaries, it therefore takes time to obtain very good proper motions, e.g. two or more space missions. The systematic error from the orbital motions depends on the orbital period of the pair.

Unresolved double star can be discovered from astrometric observations in a single mission by the large residuals of the standard solution of linear motion for a single star. Many such discoveries were obtained with Hipparcos, in fact only 80 per cent of the stars could be solved as single stars without problems (see section 6.5.1 in ESA 2000). Many more will be problematic in Gaia observations which will thus lead to new discoveries.

With two missions a very large fraction of stars can be discovered as binaries from the residuals of either mission and from a comparison of the proper motions from each mission and from the motions derived from the positions at the two epochs. Furthermore, the acceleration in the orbit can be determined. The additional use of Hipparcos results will lead to further discoveries, depending on the orbital periods.

Many binaries and planetary systems will be discovered with these methods (see sections 1.5 and 1.6.2, respectively, in ESA 2000). What has changed significantly since Gaia was approved by ESA is the situation of the extrasolar planets. Rather than a few tens, there are now a few thousands already identified, essentially, by Kepler. Furthermore, the final Gaia epoch precision is somehow worse than what was originally asked by the scientific community. So, the expectation in terms of planet discovery and orbit fitting should be lowered. However, the discovery and characterisation of binary and multiple stars remain as valuable as they were ten years ago. A space based proper motion catalogue derived from the two Gaia missions, 15 years apart, for 1 billion stars would supersede any ground base catalogue ever compiled.

The benefit of the much more accurate proper motions for studies of the Galaxy deserves to be elaborated. The same is true for the observations of star positions in nearby galaxies and of solar system objects.

From the scientific point of view, the impact of a significantly higher accuracy of absolute proper motions can be seen in studies of galactic dynamics (probing much deeper into the galaxy as well as into the halo), dynamics of star clusters and OB associations (conditions of star formation, cluster formation), proper motions and orbits of globular clusters (galactic potential), apparent proper motions of QSOs, resulting from the orbital acceleration of the solar system in our galaxy and the aberration effects resulting from the velocity vector of the solar system in space. Direct comparisons with the Gaia results would also provide statistics on orbital binary systems from proper motion disturbances.

Photometry

There are two reasons to perform photometry of stars with the mission itself: (1) enable chromatic corrections of the astrometric observations, and (2) provide astrophysical information for all objects, including astrophysical classification (for instance object type such as star, quasar, etc.) and astrophysical characterisation (for instance interstellar reddenings and effective temperatures for stars, photometric redshifts for quasars, etc.).

For a new mission the first reason is still valid to the same extent as for Gaia, but the justification for the second purpose has changed. By the time of a new mission, multi-colour photometry will be available for all the observed stars with higher accuracy and better spectral resolution than the mission itself can provide. Such photometry will be provided and be available from large surveys as Pan-STARRS and LSST, providing five or six spectral bands from 300 to 1100 nm. The angular resolution of these surveys could not be found, but since they are ground based, the resolution is hardly better than 0.5 arcsec. Also spectra can be obtained in the millions by new spectrographs.

For Gaia the angular resolution along scan of the astrometric observation is about 0.12 arcsec (FWHM of the sampled line-spread function). For photometry however the low-dispersion spectra limit the resolution greatly for double and multiple stars because the spectra of the two stars will overlap, each spectrum having a length of 1-2 arcsec. This was a penalty of going from filter photometry to spectrophotometry proposed by the industrial contractor in 2005.

This is not apparent from the Gaia information quoted below: *"Photometric observations will be collected with the photometric instrument, at the same angular resolution as the astrometric observations and for all objects observed astrometrically"*. To my recent question about this issue Floor van Leeuwen replied: *"For the spectra there are limitations in densely populated areas, although we will try as far as we can to de-blend polluted spectra. So, in principle the data are there, but they may not be easy or even possible to reduce."* – This confirms my understanding of the problem of Gaia photometry for double and multiple stars.

For a new mission, it is proposed to return to filter photometry. The two prisms should be replaced by two filters which may, e.g., cut out 330-460 and 460-600 nm corresponding to B and V. The intensity in the band beyond 600 nm may be obtained from the intensity of the wide-band 330-1050 nm measured in the astrometric field by subtracting the intensities of the observed B and V, and the result may be called R. This is just an example and the exact choice of band need careful study, but the basic idea is that 3-band photometry would be obtained: B, V, R, but three bands are probably sufficient for the chromatic correction of astrometric observations.

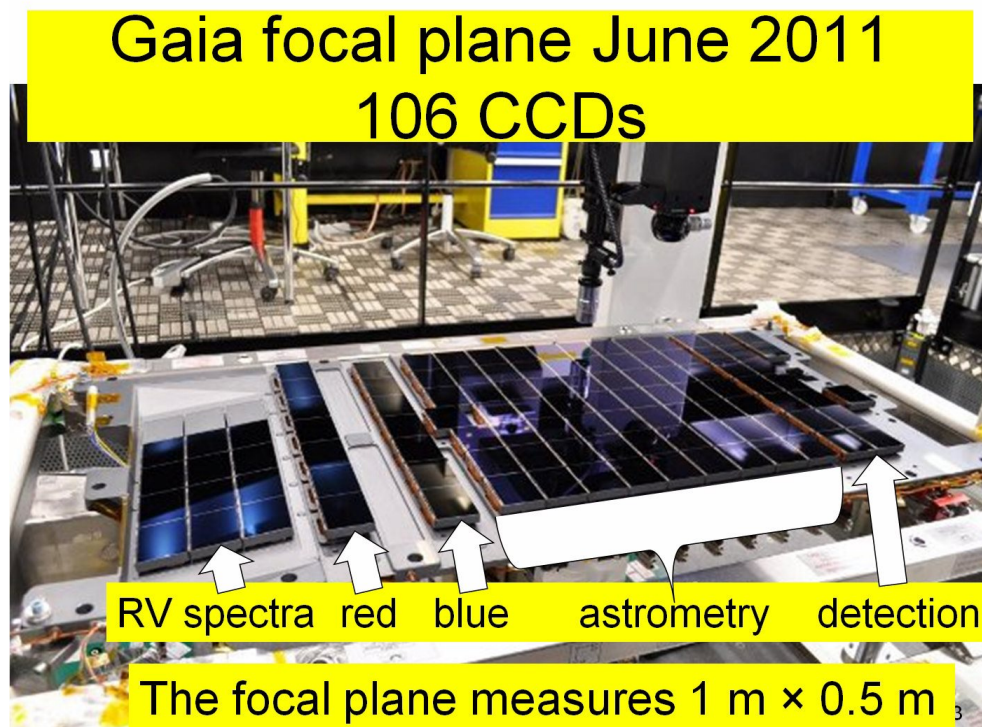


Figure The focal plane of Gaia with 106 large CCDs. The stars enter from right, are detected and then measured during the passage of various CCDs. At first comes astrometry on 9 CCDs, then photometry on two CCDs measuring a short spectrum in blue, BP, and in red, RP. At left the stars are observed at high dispersion around the red Ca-triplet in order to obtain the radial velocity of the star and the intensity in the spectrum.

It should be considered to include 3 CCDs for photometry instead of two. It appears from the figure that the space would be available without any shift of the CCDs for astrometry and RV. If this is done 4-band photometry could be obtained, B, V, R, I, where one of the bands results from subtraction of the three bands from the photometry obtained in the wide astrometric band.

The accuracy of such filter photometry will be much better than the Gaia photometry from the short spectra in two respects: (1) All stars, especially the faint ones, will be less affected by noise from readout, background and parasitic stars, and (2) for components of double and multiple stars the improvement will be most pronounced, and they will obtain good photometry in many cases where Gaia could not give anything. The further advantage of the simpler data reduction is worth mentioning.

It must be emphasized that the 4-colour photometry is meant for use to make the chromatic astrometric corrections and for astrophysics when nothing else is available. Photometry was an historical necessity, when spectra were too expensive to obtain. Dedicated spectrographs are now being built which will obtain millions of stellar spectra. That provides the quantitative science. Superb multi-colour all-sky IR data are already available, and the multi-colour optical surveys are being rapidly completed. It does not make sense to duplicate those with an astrometric mission.

Origins of Gaia

In October 1993 a proposal was submitted to ESA to study for astrometry “*a large Roemer option and an interferometric option*”, GAIA. They should be studied as two concepts for an ESA Cornerstone Mission for astrometry “*without a priori excluding either*”, as the cover letter said. Interferometry was completely dropped in January 1998, and further design was based on the Roemer concept with direct imaging on CCDs using Time Delayed Integration (TDI), a technique not previously used in space. Therefore, Gaia is a large Roemer option. – The history of the events before and after 1993 are briefly reviewed by Høg (2013).

References

ESA 2000, GAIA – Composition, Formation and Evolution of the Galaxy. Concept and Technology Study Report. ESA-SCI(2000)4.

ESA 2013, ESAs call for White Papers for the definition of the L2 and L3 missions in the ESA Science Programme at <http://sci.esa.int/Call-WP-L2L3>.

Gaia 2013, The website at:
http://www.rssd.esa.int/index.php?project=GAIA&page=Science_Performance

Høg E. 2013, Origins of Gaia. <https://dl.dropbox.com/u/49240691/GaiaOrigins.pdf>

Science Performance of Gaia

Copied from the Gaia website in 2013

Gaia will perform micro-arcsecond (μas) global astrometry for all $\sim 1,000$ million stars down to $G \approx 20$ mag — except for the $\sim 6,000$ brightest stars in the sky — by linking objects with both small and large angular separations in a network in which each object is connected to a large number of other objects in every direction. Over the five-year mission lifetime, a star transits the astrometric instrument on average ~ 70 times, leading to ~ 630 CCD transits. Gaia will not exclusively observe stars: *all* objects brighter than $G \approx 20$ mag will be observed, including solar-system objects such as asteroids and Kuiper-belt objects, quasars, supernovae, multiple stars, etc. The Gaia CCD detectors feature a pixel size of $10\ \mu\text{m}$ (59 milli-arcsecond) and the astrometric instrument has been designed to cope with object densities up to 750,000 stars per square degree. In denser areas, only the brightest stars are observed and the completeness limit will be brighter than 20^{th} magnitude.

Photometric observations will be collected with the photometric instrument, at the same angular resolution as the astrometric observations and for all objects observed astrometrically, to:

- enable chromatic corrections of the astrometric observations, and
- provide astrophysical information for all objects, including astrophysical classification (for instance object type such as star, quasar, etc.) and astrophysical characterisation (for instance interstellar reddenings and effective temperatures for stars, photometric redshifts for quasars, etc.).

Spectroscopic observations will be collected with the spectroscopic instrument for all objects down to $G_{\text{RVS}} \approx 16$ mag, to:

- provide radial velocities through Doppler-shift measurements using cross-correlation (~ 150 million stars);
- provide astrophysical information, such as interstellar reddening, atmospheric parameters, and rotational velocities, for stars brighter than $G_{\text{RVS}} \approx 12$ mag (~ 5 million stars); and
- provide element abundances for stars brighter than $G_{\text{RVS}} \approx 11$ mag (~ 2 million stars).

The spectroscopic instrument has been designed to cope with object densities up to 36,000 stars per square degree. In denser areas, only the brightest stars are observed and the completeness limit will be brighter than 16th magnitude.

In the scientific performance assessments for Gaia, all known instrumental effects are included under the appropriate in-flight operating conditions (temperature, CCD operating mode, etc.). All error sources are included as random variables with typical deviations (as opposed to best-case or worst-case deviations). All performance estimates include a 20% contingency margin. This margin is a DPAC science margin, neither meant for nor available to the Gaia industrial prime contractor. The science margin is assumed to cover, among others:

- "scientific uncertainties" in the on-ground data analysis, including uncertainties related to relativistic corrections, aberration corrections, and the spacecraft and solar system ephemeris;
- scientific effects such as the contribution to the astrometric error budget from the mismatch between the actual and the calibrating point spread function, estimation errors in the sky background and total detection noise values that need to be fed to the centroiding algorithm, etc.;
- the fact that the sky does not contain, as assumed for the performance assessments, "perfect stars" but "normal stars", which can be photometrically variable, have spectral peculiarities such as emission lines, have unrecognised companions, be located in crowded fields, etc.;
- other astronomical environmental factors such as, e.g., localised enhanced sky-background surface brightness, unrecognised small-scale sky-background-brightness gradients, unrecognised prompt particle events, etc.

Here follow on the Gaia website (Gaia 2013) the detailed performances for astrometry, photometry and spectroscopy.

+++++

White paper for an ESA L2 or L3-class mission “Europe returns to Venus”

Emmanuel Marcq
emmanuel.marcq@latmos.ipsl.fr

Due on 2013-05-24



Artistic representation of a balloon in the atmosphere of Venus
(from the *European Venus Explorer* mission proposal)

Author list

- Emmanuel Marcq, Maître de Conférences at Université de Versailles Saint-Quentin / LATMOS.
- Anni Määttanen, Researcher at CNRS / LATMOS
- Franck Montmessin, Researcher at CNRS / LATMOS
- Sébastien Lebonnois, Researcher at CNRS / LMD

Supporters

- Jean-Loup Bertaux, CNRS/LATMOS, France
- Éric Chassefière, Université Paris Sud/IDES, France
- Sanjaye Limaye, University of Wisconsin, U.S.A.
- Dmitri Titov, ESA/ESTEC, Netherlands

Contents

A. Scientific and societal interest of Venus's exploration	3
A.1. Real world occurrence of a runaway greenhouse effect	3
A.2. Constraining the inner edge of the habitability zone	4
A.3. Comparative studies of sulphur-based aerosols	4
B. What we know and do not know about Venus	4
B.1. Clouds and chemistry	4
B.2. Atmospheric dynamics	5
B.3. Radiative balance	6
B.4. Volcanism and surface	6
B.5. Water and climate history	7
B.6. Habitability and exobiology	8
C. Resulting specifications	8
C.1. Long-term monitoring of Venus	8
C.2. Isotopic measurements	9
C.3. Analysis of clouds and hazes	12
C.4. Surface and lowest atmosphere measurements	13
D. Possible technical solutions for a L-class mission	15
D.1. UAV or balloon in the clouds	15
D.2. Orbiter	16

A. Scientific and societal interest of Venus's exploration

A.1. Real world occurrence of a runaway greenhouse effect

If we had to summarise as briefly as possible the essential difference between Earth and Venus, it could be stated in this way: *all other things being equal*, if Earth was located at the same distance from the Sun as Venus is, its mean surface temperature would be 66°C, whereas the mean surface temperature of Venus is 460°C. Yet, Venus is the most Earth-like planet that we know in many ways: it is our closest neighbour in the solar system, its size, bulk composition are very similar to Earth's.

The huge differences in surface conditions can be ascribed to the very different present-day atmospheres. Earth's atmosphere causes a greenhouse warming of about 33°C, primarily caused by water vapour first and then by trace amounts of carbon dioxide, whereas Venus's thick (surface pressure 92 times higher than on Earth) atmosphere

causes a 500°C greenhouse warming, mainly due to the carbon dioxide which is its main constituent.

The currently accepted scenario that accounts for the present-day conditions on Venus is named *runaway greenhouse*⁴⁰. Assuming similar starting volatile inventories on both planets, both atmospheres contained several hundred bars of water vapour, and several tens bars of carbon dioxide. Most of water vapour probably did not reside in the atmosphere, but formed instead water oceans similar to present day Earth's. But beyond a certain threshold of solar flux (which is twice higher on Venus than on Earth), it becomes impossible for the atmosphere to be saturated with water vapour – since water vapour is a very potent greenhouse gas, more water vapour yields higher temperatures in the atmosphere, therefore rising the saturation pressure in a positive feedback loop. Eventually, the whole oceanic reservoir evaporates into the atmosphere, causing a tremendous H₂O-CO₂ greenhouse effect. Water is then present up to the upper atmosphere, where it is easily dissociated into hydrogen and oxygen atoms. Hydrogen atoms escape then eas-

ily into outer space since they are the lightest atoms, leaving free oxygen to oxidise other molecules.

Once most of water has escaped, the atmosphere reaches a state very similar to present-day Venus's, with a high CO₂ pressure. On Earth, the solar flux was low enough so that no positive feedback loop could kick in, liquid water oceans remained stable and could dissolve CO₂ into carbonate rocks. Only molecular nitrogen remained in the atmosphere (molecular oxygen resulting from the photosynthetic activity and became a major constituent of our atmosphere at a much later stage).

With the assumption that Venus started water-rich, it is therefore the only world we know that most likely experienced such a runaway greenhouse – although we do not know precisely *when* this runaway started. However, the radiation output of the Sun is slowly increasing with time, so that within a billion year Earth itself will experience the solar flux that Venus did during its early history, putting our own world at risk of eventually experiencing the same fate as Venus⁴³, ultimately rendering Earth uninhabitable.

More worryingly, these runaway greenhouse studies are part of a wider topic known as climate stability. Greenhouse gases are currently increasing in our atmospheric due to human activities. Some⁷⁰ fear that this anthropogenic perturbation could trigger a runaway greenhouse much sooner, since all previous changes of comparable magnitude in Earth's atmosphere occurred at a much slower pace. All the various thresholds (in terms of composition or solar irradiance) are still a matter of active research^{54,31,69}, and a better knowledge of Venus radiative budget and past atmospheric history could provide modellers with better observational constraints.

A.2. Constraining the inner edge of the habitability zone

This phenomenon (the runaway greenhouse insolation threshold) is actually thought to be the limiting factor in constraining the habitable zone around main sequence stars⁴⁷. More than thousand extra-solar planets are

confirmed as of 2013, among which about ten (roughly 1%) are located in the habitable zone, defined as the range of orbits that allow for stable liquid water on the surface. Also, 20 to 30 extrasolar gas giants are located in this zone, and may host habitable satellites. All scientific advances regarding the runaway greenhouse will help in better constraining these numbers, and are crucial in determining the potential for Earth-like life in our galaxy.

A.3. Comparative studies of sulphur-based aerosols

Even though Venus's surface temperature is the hottest in the solar system, it could be even higher if it were not enshrouded by ubiquitous and very reflective cloud layers which reflect about 70% of the incoming solar light. These thick clouds consist mainly in tiny (about 1 μm in radius) droplets, whose chemical composition is mostly sulphuric acid mixed with some water³⁴.

Very interestingly, such acidic aerosols can be observed in Earth's stratosphere after volcanic eruptions large enough to inject sulphur dioxide at such high levels, causing momentary cooling of Earth's atmosphere through an albedo increase⁸¹. It has even been advocated⁷⁴ that intentional engineering of such sulfate aerosols could be used to mitigate the enhanced anthropogenic greenhouse effect on Earth.

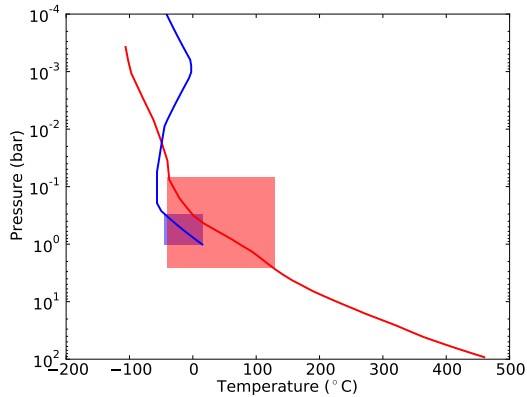
Given the very high stakes of such projects, a detailed study of the naturally occurring analogous clouds of Venus and of their exact influence on the radiative budget of their host planet is of paramount importance even beyond the scientific community.

B. What we know and do not know about Venus

B.1. Clouds and chemistry

Venus has a large and complex cloud system. Clouds and hazes extend from 45 to 80 km, spanning an enormous range of environments from temperatures and pressures of 100°C at the cloud base to -70°C and 1 mbar at the upper hazes. Visible opacity of the

cloud deck is in the range 20 to 40, and total mass loading is thought to vary from 0.1 to 10 mg/m³. Most of the opacity appears to be in a main cloud deck that stretches from 50 to 60 km. Temperature profiles from descent probes and radio occultation show this region to be convectively unstable, suggesting that most of the cloud formation here is convectively driven, as are cumulus clouds on Earth.



Earth (blue) and Venus (red) temperature profiles with respect to pressure. Significant overlapping exists in the cloud region (rectangles) of both planets

Remote sensing observations show that the upper cloud, at least, consists of micrometre-sized droplets of 75 % H₂SO₄:25 % H₂O. This sulphuric acid is photochemically produced at the cloud tops by combining water vapour and sulphur trioxide. Sulphuric acid was also detected by *in-situ* analysis on descent probes, but they also measured the presence of chlorine, phosphorous and perhaps even iron in the lower clouds. These measurements cannot be explained by current cloud models. The mechanisms involved in the formation of this convective cloud are still unclear, despite insights gained from *Venus Express*^{5,60}.

An additional puzzle in the cloud composition is the nature of the unknown UV absorber that is responsible for the distinctive features seen at the cloud tops at UV-blue wavelengths. Dozens of possible substances have been proposed, but it will be impossible to resolve without *in situ* chemistry measurements^{59,85}. *Venus Express* is currently monitoring morphology and global structure of the cloud layer but cannot provide a definitive answer to the question of its chemical origin.

Measurements by the *Venera* and *Pioneer Venus* descent probes sounded the vertical

structure of the cloud. The *Pioneer Venus* large probe also carried the Large Cloud Particle Spectrometer (LCPS) which measured number density, size distribution, shape and refractive index²⁶. One of the surprises was the discovery of a multi-modal size distribution that could indicate that several distinct processes are involved in the formation of the cloud system, and large particles over 3 μ m in size and number density ~ 10 cm³, which may well be solid crystals and thus cannot be composed of sulphuric acid. The existence of these large “Mode 3” particles and their composition remains a hotly debated topic, which is impossible to resolve without *in-situ* measurements. All cloud models to date are based on the particle size distributions observed from a single descent profile – that of the *Pioneer Venus* Large Probe⁴⁶ – but observations from *Venus Express* have now revealed strong variability of the cloud properties across the planet^{91,5}.

The chemistry of the atmosphere is fundamentally different above and below the clouds: above the clouds, it is dominated by photochemical processes in a convectively stable environment⁶³; while below, UV radiation is much reduced and high temperatures play a more significant role in governing reactions⁴⁸. Chemical abundances above the clouds have been well probed by generations of orbiters, in particular by the occultation spectrometry of *Venus Express*^{7,88}. *Venus Express* is also the first orbiter to exploit the near-infrared windows to measure chemical abundances in the lower atmosphere^{87,56,9,8}. However, chemical processes in the intervening cloud layer remain poorly understood and very difficult to measure either from orbit or from rapidly dropping descent probes. Profiles of H₂SO₄ vapour abundance may be obtained by studying radio occultation from *Venus Express*, but the difficult analysis has not yet been attempted.

B.2. Atmospheric dynamics

The most striking feature of the Venusian atmospheric circulation is the zonal super-rotation found in its troposphere and mesosphere (0 – 100 km)²⁹. Although the solid planet only rotates once every 243 Earth days,

the atmosphere at the cloud tops rotates some 50 to 60 times faster, with westwards wind speeds often reaching 400 km/h. The zonal (East-West) winds are at their maximum at the cloud tops, decreasing steadily to zero at the surface and at the mesopause near 100 km. In addition to zonal super-rotation, there appears to be a slower overturning of the atmosphere from equator to pole with meridional velocities of 10 to 20 m/s and giant vortexes at each pole recycling the air downwards⁵⁵. The mechanisms causing the super-rotation are still unknown. Two leading hypotheses suggest that *vertical* momentum transport by thermal tides and *meridional* momentum transport (by mean meridional flow and large eddies), respectively, are the most important mechanisms driving the super-rotation.

Semi-direct measurements of wind fields at the cloud tops (60 to 70 km) have been obtained from several spacecrafts by tracking cloud features on the day side. *Venus Express* has provided more cloud-top wind fields⁶⁴ and has added wind fields at cloud base altitude (about 50 km) using infrared spectral windows on the night side⁷⁶. However, it has not been possible to reconstruct a global tide model because the wind field at each of these heights is incomplete, spanning only half of the globe.

Profiles of convective stability obtained by radio occultation from *Venus Express* show that a convective region extends from 51 to 60 km⁸³. This convective zone, extending over more than a scale height, is radiatively heated from below by the deep atmosphere, and radiatively cooled to space from about 60 km. There have been many attempts to model this convective activity, but only *in situ* measurements could provide the vertical velocities within the cloud layer.

B.3. Radiative balance

As previously stated, the climate of Venus is notable in particular for its extreme greenhouse effect. The greatest uncertainties in radiative balance on Venus, as on Earth, relate to the role of clouds. The standard work on Venus' radiative balance¹⁵ assumes a single, best-guess cloud model. The observations of

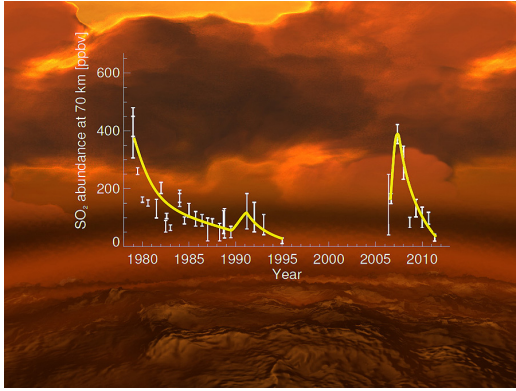
Venus Express have revealed a highly variable cloud layer, whose optical depth varies by over an order of magnitude³⁹. Local measurements of radiative flux would provide context for all other measurements there, to ascertain whether the instrument is above a thin or thick lower cloud, or below a region of high or low UV absorber amount.

Understanding the radiative fluxes will also be crucial in order to understand the dynamics. The “engine” driving the atmospheric circulation on Venus, as on Earth and Mars, is differential heating of the atmosphere by solar radiation. On Earth and Mars, the sunlight is absorbed mainly at the surface, leading to vigorous tropospheric dynamics; in contrast, on Venus it is in the cloud layers that most sunlight is absorbed, so this is the most important altitude range which must be studied in order to understand the global dynamics. Global circulation models have confirmed a strong sensitivity of the atmospheric circulation to the assumed vertical profile of heat deposition^{84,52}.

B.4. Volcanism and surface

The radar images of the surface delivered by the *Pioneer Venus*, *Venera-15* and *-16*, and *Magellan* orbiters surprisingly revealed that Venus possesses a young surface suggesting a geologic active past. Volcanism and tectonics have strongly altered the Venusian surface^{35,17} forming highly deformed old plateaus (*Tesserae*) and extensive lowlands (*Planitia*) – vast young volcanic plains covering about 80 % of the surface. Relatively rare and uniformly distributed impact craters suggest that global resurfacing of Venus happened about 700 Myr ago^{6,61}.

Venus Express has provided exciting new observations suggestive of presently or recently active volcanism, including measurements of surface alteration near volcanic features occurring in the last 2.5 Myr^{65,80}; a sudden tenfold increase in mesospheric SO₂ followed by several years of decline^{57,58}, echoing earlier work from UVIS on-board *Pioneer Venus* orbiter²⁵. These observations have also prompted a reanalysis of *Magellan* brightness maps, suggesting that warm lava flows were observed¹².



Long-term SO_2 evolution above the clouds of Venus⁵⁸

B.5. Water and climate history

The large uncertainties on the results of existing radiative transfer codes of H_2O - CO_2 temperatures do not enable us to know whether the surface temperature of Venus was ever low enough to allow the condensation of water even in the reduced solar illumination conditions prevailing at the beginning of the Solar system, although this is the most probable case⁵³. But even if thermodynamical conditions allowed the formation of an ocean at some early stage, the necessary condition for such an ocean to have formed is that large amounts of water were available in the early atmosphere. It could not have been the case, for example if Venus was initially endowed with significantly less water than Earth, as occurs in some existing accretion simulations⁷⁵, and/or if this water was rapidly lost through the combined effects of (1) the progressive crystallisation of the magma ocean, releasing water to the atmosphere, and (2) hydrodynamic escape removing this water from the atmosphere to space³⁰.

An interesting outcome of this scenario is that during the first few hundred million years, Venus could have developed a dense molecular oxygen atmosphere (typically around 10 bar) formed by photolysis of water, with substantial amounts of water vapour. The physical and/or chemical processes of oxygen loss whereby this oxygen shifted into the interior are not well understood. In other circumstances, could Venus have kept a substantial amount of oxygen in its atmosphere for billions of years? In such a case, an exoplanet similar to the young Venus

could appear as a false positive of a planet hosting photosynthetic life³⁰.

Because of the active geology of Venus, such as the global resurfacing event about 700 Myr ago, traces of any primitive ocean may seem unlikely to be extensive enough to even confirm its existence. From an analysis of *Galileo*/*NIMS* data, it has been suggested that the majority of highlands on Venus have a lower emissivity in the near IR than the majority of the lowlands, which could be attributed to the presence of granitic walks signing an ancient ocean³⁶. The survival of a signature of liquid water at the surface cannot be therefore excluded. In a global perspective, the only way to reconstruct the detailed history of volatile reservoirs on Venus, from accretion to the end of the heavy bombardment, that is during the first billion years, is to constrain numerical models of the interior-magma ocean-atmosphere-interplanetary space system evolution with present day noble gases abundances and isotopic fractionation patterns, and with ratios of stable isotopes through the use of the powerful techniques of isotopic geodynamics. If any evidence for past liquid water is found at the surface by future landers, the mineralogical records could be used to better constrain such evolution models.

Venus Express has been providing key information on the different channels of ion escape thanks to the ASPERA-4 ion spectrometer coupled to the magnetometer observations during solar minimum conditions^{3,42}. Since *Venus Express* should remain operational at least until late 2014, it will also provide data during solar maximum conditions. It will also provide the state of the hydrogen exosphere thanks to the UV spectrometer SPICAV¹³. Unfortunately, ASPERA cannot directly measure the neutral escape, nor provide any isotopic information on the ion escape⁴. However, ion escape is probably among the major escape channels because of the large gravity of Venus hampering the escape of neutral particles³. Ion measurements outside Venus's ionosphere provide also indirect information on the neutral exosphere and on its escaping component. ASPERA and the *Venus Express* magnetometer are also

providing the rate of solar wind penetration into Venus’s atmosphere⁹³, key information for reconstructing the past evolution of ^4He in Venus’s atmosphere⁴⁹ which depend on several factors, including the solar wind contribution, on Venus’s past and present out-gassing rate and on Venus’s loss rate.

The contrast between low and high solar activity (as well as during the energetic solar events) will provide important clues on the evolution of the interactions between Venus and the solar wind during the past few billion years. It is most probable that the present channels are representative of atmospheric escape since the end of the late heavy bombardment³⁰. Measurements of ^{129}Xe , ^{40}Ar and ^4He abundances would provide key information on the cumulated past degassing and atmospheric loss rates, thereby complementing the information gathered by *Venus Express* and allowing a reconstruction of the last 3.5 billion years of atmospheric and out-gassing evolution.

B.6. Habitability and exobiology

To understand the emergence of life in our solar system, we clearly need to understand (1) which regions of our solar system offered viable environments for the emergence of life, and (2) what were the inventories in these locations of the building blocks of life, ranging from oxygen and water through to more complex carbon-based compounds found for example in carbonaceous meteorites.

Improvements in our understanding of the history of Venus, including the key issues for habitability of climate, liquid water, volatiles and volcanicity, considered jointly with Earth and Mars, will bring important lessons about the evolution of habitability on terrestrial planets in our solar system and beyond. As to the second point, the clouds of Venus are one of the very few places in the solar system in which liquid water can be presently found, albeit in a very acidic form, which benign temperatures of 20°C and intriguing mixed phase chemistry including gaseous, liquid and possibly solid phases. As the climate of Venus has heated and cooled, this region of liquid clouds particles will have risen and fallen in altitude to occur approximately at a constant

temperature range³³, so there is every reason to believe that the chemical cycles of the cloud layer have remained relatively unchanged. Indeed, it has been suggested that life on Venus, if it has been established in supposed early oceans, could have migrated to the clouds of Venus, where in present days habitable niches with pressure and temperatures comparable to terrestrial conditions exist³³. These conditions are not far from habitable as might at first be thought: extremophiles have been found on Earth living in condition of $\text{pH}=0$ ^{78,22}, and other bacteria have been found living in cloud droplets²⁸. The clouds contain diverse sulphur compounds in various oxydo-reduction states, which are known to be used by life to gather energy¹⁴. Lightning, whether associated with cloud processes or volcanicity, could serve as an energy source for the synthesis of pre-biotic compounds⁶², as can UV radiation: it has for example been suggested⁷⁹ a potential biotic contribution to the UV absorber. Although it is extremely unlikely that one would find life in the combination of high acidity and low nutrient availability found in the clouds of Venus¹⁴, it is still of biochemical interest to study whether complex carbon compounds can be found here, whether of indigenous origin or carried by meteorites. This is especially of interest given the parallels between this environment and that of Earth in its first few tens of million of years, when it exhibited a mainly CO_2 atmosphere⁹².

C. Resulting specifications

C.1. Long-term monitoring of Venus

We now have usable scientific observations of Venus’ atmospheric chemistry and dynamics with sufficient spatial resolution for about 40 years, which together constitute one of the longest climatological data set in the solar system after Earth’s. In this emerging realm of comparative climatology, some evolution, perhaps even cycles, can be found. Thus, SO_2 measurements at cloud top level from the 1980’s²⁷ are very reminiscent of the same measurements in the late 2000’s and early 2010’s⁵⁸. Decadal variations are also predicted by some general circulation models⁶⁶. Shorter-term variability is also observed in

mesospheric wind fields⁴⁵ and well as for SO₂. This strong variability is even more unexpected considering that, unlike for Earth or Mars, Venus experiences no seasonal forcing since its obliquity is too low. These variabilities, especially of the cyclic kind, are more probably related to proper frequencies of the complex Venusian atmosphere not directly related to external forcings, such as the Quasi-Biennial Oscillation (QBO) or the El Niño/La Niña alternation on Earth. Such long-term cycles are thus also an active topic in Earth atmospheric science.

Long-term (on a decadal timescale) monitoring of atmospheric circulation and chemistry are therefore of the highest scientific interest. After the failure of the orbital insertion of *Akatsuki* and the expected end of the *Venus Express* mission by late 2014, the short-term hope of Venusian exploration entirely resides in Roskosmos' *Venera-D* mission, scheduled for a launch in 2018 at the soonest. After that, it is probably too late for the design of another ambitious mission to Venus during the 2020s if we consider the bleak international context, especially regarding NASA reduced participation and leadership in planetary science missions after the severe planned budget cuts. Thus, ESA has the opportunity, after the success of *Venus Express*, to return there twenty years later in order to monitor Venus in its decadal evolution and re-assess its global leadership in Venusian science. Moreover, considering the strong decadal changes that are expected on Earth due to human influence, monitoring Venus during a similar time frame may constitute a sort of "control-case" study so that external variability (e.g. solar variations) can be taken properly into account. Ground-based monitoring, although necessary and helpful to compensate for the shortcomings of spatial missions (which are too short-lived on a decadal timescale and whose light and compact instruments are often outperformed by the best available instruments mounted on ground-based telescopes), lack the necessary spatial resolution and coverage to properly monitor the changes on the whole extent of the Venusian atmosphere.

C.2. Isotopic measurements

Noble gases in planetary atmospheres are key to decipher the history of planets from early stages to the current epoch. Similarities and differences in the noble gases compositions and budgets of Earth, Venus and Mars reflect similarities and differences in the evolution of the three large terrestrial planets. *A robust database on the noble gas composition in planetary atmospheres is thus indispensable for a comprehensive understanding of planets, including Earth.*

It is their chemical inertness and volatility that make noble gases ideal tracers in many respects. First, we can assume that the noble gases in an atmosphere are globally well mixed, so that any local measurement reflects global abundances and long-term evolution of a planet. Second, the rareness of noble gases in telluric atmospheres allows recognition of contributions from sources which would go unnoticed for other elements. Radiogenic noble gas isotopes from the decay of long- and short-lived radioactive parents enable the reconstruction of the early and later degassing history of a planet as well as potential early and later atmospheric losses. Noble gases are also unique tracers for possible source materials of the planets themselves, helping to constrain their accretionary history. Finally, isotopic fractionations of noble gases relative to a source composition again constrain degassing and potential losses. More specifically:

C.2.a. Radiogenic noble gases isotopes

They constrain the degassing history of a planet. The radiogenic and fissionogenic fractions of ¹²⁹Xe and ¹³⁶Xe, produced from now extinct ¹²⁹I and ²⁴⁴Pu respectively (with half-lives of 15.7 Myr and 80 Myr) in the atmospheres of Earth and Mars reveal that the two planets were both efficiently degassed and did not retain an atmosphere until about 100 Myr of solar system history⁷². For the Earth, this is sometimes attributed to the moon-forming giant impact only, but the data from Mars suggest that a vigorous early degassing may be a common fate to all terrestrial planets.

⁴⁰Ar (from the decay of long-lived ⁴⁰K)

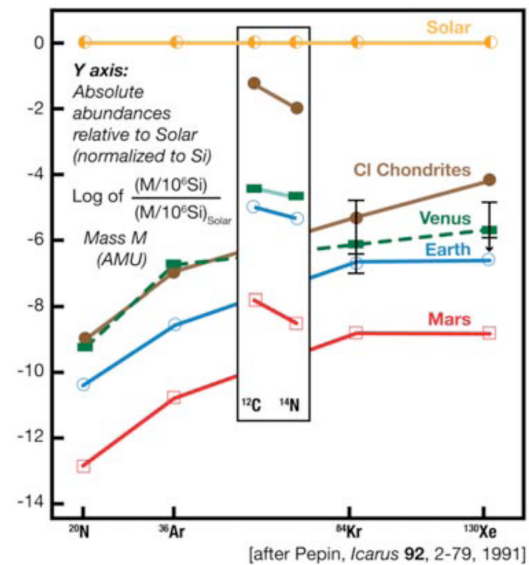
likely has been continuously accumulated over more than 4 Gyr and therefore constrains the volcanic/tectonic evolution of Venus throughout its history. Its mixing ratio in the atmosphere of Venus is 70 ± 25 ppmv⁸⁹. The atmosphere contains about two to four times less ^{40}Ar per gram planet than does Earth's. Although the bulk K content of Venus is also uncertain, the general view is that Venus has degassed less of its ^{40}Ar than Earth⁴⁴. The absolute concentration of ^{40}Ar is already known to within 35%⁸⁹. To constrain the degassing history better, this measurement will have to be combined with the *in situ* K measurements to be performed by future landers.

^4He is another radiogenic noble gas isotope from long-lived parents (U and Th). It is not retained in Venus' atmosphere over its entire history, although its atmospheric lifetime of several 10^8 yr is considerably longer than on Earth. Yet, ^4He still can be used to constrain Venus' long-term degassing behaviour and to compare it with those of the two other large terrestrial planets⁵⁰. ^4He is also useful to constrain the atmospheric escape rate. Currently, ^4He concentration data are only available for Venus's upper atmosphere, with large uncertainties of a factor about two. The model-dependent extrapolation of these data even leads to an uncertainty of a factor ~ 20 for the ^4He concentration in the mixed atmosphere²⁰. $^3\text{He}/^4\text{He}$ ratio, when measured, can be used to better constrain the non-radiogenic fraction of the total ^4He in the atmosphere (which is assumed to be negligible by some).

C.2.b. Non-radiogenic noble gases isotopes

They constrain gas acquisition and loss processes. They are thus invaluable for understanding Venus's atmospheric and interior evolution and its accretionary history, even though uncertainties of the available data are uncomfortably large. The most important observations⁷³ are (1) a strong depletion of the lighter noble gases relative to heavier gases and solar abundances. At first glance, the depletion for Venus is quite similar to that for Earth and Mars and also that in primitive meteorites, although the abundance patterns differ in crucial details (2) Venus's atmosphere

contains roughly one and two orders of magnitude more gas per gram planet than Earth, and Earth in turn contains roughly two orders of magnitude more noble gases than Mars. These observations (plus the few available isotopic ratios for Venus) have led to a variety of scenarios on accretion and loss processes. The development of such scenarios benefits strongly from comparisons of the similarities and differences among the three planets. Such Venus data, as scarce as they are, therefore also help to understand Earth and Mars.



Concentrations of the four heavier noble gases in the atmospheres of the three major terrestrial planets normalised to the rock-forming element Si and to solar composition⁶⁷. Also shown are the data of the chemically most primitive meteorites, the CI chondrites, which might serve as a possible analogue of planetary building blocks. The uncertainties on the abundances of Kr and Xe are nearly one order of magnitude.

Both “gas-poor” and “gas-rich” scenarios have been studied. “Gas-poor” models – stipulating that a planet's original gas inventory never waqs very much larger than the present one – were once popular, because of the roughly similar depletions of light relative to heavy noble gases in planets and meteorites. This seemed to imply that the non-radiogenic gases of the planets had been supplied by their planetesimals, represented by meteorites. Venus's atmosphere would then retain essentially the entire inventory ever delivered, whereas Earth and Mars would have lost most of their original inventories or degassed largely incompletely. Another “gas-

poor” hypothesis postulated the decreasing gas concentration from Venus to Mars to reflect decreasing noble gas partial pressures in the solar nebula with increasing heliocentric distance⁷¹. Isotopic data – especially recent data available for Earth – appear to reconcile the idea that noble gases in planetary interiors can be derived as a trapped component in accreting material², but cannot be simply related to the gases now in the atmosphere³⁸. The strong early losses of We from Earth and Mars are also hardly compatible with the idea that the decreasing present-day noble gas abundances from Venus to Mars should mainly reflect initial inventories. If a strong early Xe loss could be proven for Venus also, this would be clear evidence that the similarity of noble gases concentrations in Venus today with those in meteorites is merely fortuitous.

For the reasons mentioned above, among others, “gas-rich” accretion scenarios have become popular, where the present-day planetary noble gas inventories are remnants of some initial much larger inventory. Gas-rich models have an advantage in that they accommodate a wider range of gas accretion and loss mechanisms to preordain the observed ranges of isotopic and elemental compositions. The price for this is the need to introduce further free unconstrained parameters. Potential acquisition mechanisms for primordial noble gas inventories include gravitational capture from the accretion disk (by growing planets, planetary embryos or dissolution of gravitationally captured primordial atmospheres in a magma ocean), accretion of material irradiated by an early solar wind or accretion of volatile-rich planetesimals. Among early loss mechanisms, hydrodynamic escape has gained much popularity. Hydrodynamic escape from the exosphere (e.g. due to early solar EUV radiation or impacts) drags noble gases along, leading to elemental and isotopic fractionation. Some⁶⁸ were able to simultaneously reproduce the isotopic compositions of terrestrial atmospheric noble gases in Earth and Venus with a two stage scenario. First, Earth lost most of its gases by hydrodynamic escape driven by the giant moon-forming impact. In the second stage, a weaker, solar EUV-driven

hydrodynamic escape episode led to a further loss on both Earth and Venus. Models like this are non-unique but have predictive power. To test such models further and reduce their number of free parameters, accurate isotopic data of all noble gases in Venus atmosphere, including Kr and Xe, are required. As mentioned above, some³⁰ have explored the history of water on Venus with a hydrodynamic escape model constrained by the present-day isotopic composition of Ne and Ar in Venus’s atmosphere. Such investigations would immensely benefit from accurate isotopic data of also the heavy noble gases.

C.2.c. Non-noble gases isotopes

The composition of the stable isotopes (that is, isotopes not produced or destroyed by nuclear reactions including radioactivity) of the light elements, namely H, C, O, N, etc. provide exceptional insights into the origin and processing of planetary atmospheres. These isotopic ratios can be modified significantly by chemical exchanges and by kinetic processes, and the resulting effects can be drastic. For instance, the deuterium/hydrogen (D/H) ratio varies by one order of magnitude among major cosmochemical reservoirs (e.g. the solar protonebula, meteorites, planetary atmospheres) and several orders of magnitude at a micrometer scale in meteoritic matter.

The D/H ratio of the Venusian atmosphere has been measured repeatedly by several spacecraft instruments, and there is agreement that its value is (160 ± 20) times the terrestrial ratio of the oceans and most meteorites¹⁸, and about 1000 times the original ratio of the solar nebula. It also overcomes the cometary ratio which is about 3 times the terrestrial ratio. Escape of atmospheric atmospheres could in principle explain the depletion of the light isotopes (e.g. H) relative to the heavy ones (e.g. D) and non-thermal escape processes have been effectively advocated in the case of Venus. However, there is no consensus as to whether Venus started with large amounts of water (oceans) that was dissociated and H energised by early solar UV, resulting in efficient escape to space^{21,43,30}, or if water would have been more or less in a steady state, the large

fractionation being the result of a large escape from a small reservoir³². In the latter case, a steady state would be maintained by continuous supply of volatile elements by impacts and mantle degassing, and the water content and D/H ratio would have evolved with time. Deciding between these scenarios has important consequences for understanding the origin and evolution of the Venus atmosphere. Unfortunately, the use of a single stable isotope tracer alone does not permit to do this, and it is necessary to consider other tracers such as noble gases (see above), elemental abundances, and other stable isotopic ratios.

The extent of stable isotopic fractionation depends on the relative mass of the isotopes and is therefore variable for different elements. Furthermore, different cosmochemical end-members have contrasted extents of fractionations for different elements. The isotopic composition of carbon, $^{12}\text{C}/^{13}\text{C}$, does not vary significantly at a reservoir scale in the solar system, with most values including estimates of the protosolar nebula value being around the terrestrial ratio of 89. Venus does not make an exception with a value of 88.3 ± 1.6 measured by the *Venera 11* and *12* instruments⁴¹.

The isotopic composition of oxygen constitutes a very important cosmochemical tracer of the origin of matter in the solar system and subsequent mixings between forming dust, grains and planetesimals. Mass-independent fractionation (MIF) effects have affected O which resulted in heterogeneities that can be identified well despite “normal” fractionation that affected the isotopes of oxygen during planetary processing. Unfortunately, these effects are small, of the order of 1%, and require samples to be brought to the laboratory and analysed with high precision. Below we argue that the analysis of the nitrogen isotopes could partly overcome this problem and provide an alternative tracer of origin and processing of the Venusian atmosphere. Besides MIF, mass-dependent fractionation of oxygen is likely to have occurred in Venus progenitors and during processing in Venus, such as atmosphere-water (if any)-rock exchange and chemical/isotopic exchanges be-

tween atmospheric molecules (e.g. CO_2 -CO-OCS-SO₂). The current uncertainty on the $^{16}\text{O}/^{18}\text{O}$ ratio is about 20%. Improvements would greatly advance our understanding of origin and chemistry of oxidized components.

In stark contrast, the isotopic composition of nitrogen presents important variations which are only matched by those of D/H ratio. The Genesis (NASA) mission obtained a direct solar wind measurement of the $^{14}\text{N}/^{15}\text{N}$ ratio of the Sun, presumably representing the value of the protosolar nebula. Surprisingly, this value is lower by 40% than the terrestrial value, implying that the Earth and meteorites are largely enriched in ^{15}N relative to the starting initial gas composition. Furthermore, comets are enriched by 80% in ^{15}N relative to Earth, and therefore by 120% relative to the nebula from which all these objects are thought to be derived. Meteorites also present measurable variations. The origin of these ^{15}N enrichments in the solar system is unknown and could be due to either low temperature ion-molecule exchange in the cold regions of the disk, or to isotopic fractionation during the dissociation of N_2 and trapping of products in forming solids, to self-shielding in the gas, or a combination of these processes. In planetary atmospheres, nitrogen can also be isotopically fractionated during escape processes, and the Martian atmospheric value is enriched by 60% relative to the terrestrial one. This enrichment has been attributed to the non-thermal cumulative escape of nitrogen. The only N isotope measurement of the Venusian atmosphere indicates a terrestrial-like composition³⁷, but the uncertainty is too large (20%) to definitively explore these possibilities.

C.3. Analysis of clouds and hazes

We have already mentioned above the main questions regarding the particulate matter in Venus’s atmosphere. It also appears that the evolution and/or cyclic pattern observed for chemical composition and atmospheric dynamics are also reflected in the upper haze. For example, UV observations from *Pioneer Venus*²⁵ have found that the upper haze column opacity was following a steady decline during the 1980s, mirroring closely the de-

cline observed for SO₂. Further back in time, observations from Earth¹⁹ have evidenced decadal changes in the UV/visible brightness and appearance of Venus. These changes most likely imply a variability of other parameters than the upper haze opacity: changes in composition (refractive index) or size distribution (whether the proportion of the different identified modes, or changes in the mean size and variance of the modes themselves). Such a variability is already known with respect to latitude, with the polar haze seemingly distinct from the low latitude haze^{90,77}.

Multi-wavelength polarimetry is a tool of choice to constrain the properties of the scattering. Unfortunately, *Venus Express* instruments only had a limited capability in this domain, with only the IR channel of the SPICAV spectrometer exhibiting some sensitivity for polarization – these data have just begun being processed as of 2013. In particular, no polarization data from *Venus Express* is available in the visible and UV range, where the signature of the upper haze would be most distinct. The scientific community therefore has to rely mostly on *Pioneer Venus* data, which will be almost 50 year old at the expected time of operation for the L2 or L3 mission. The progress already done and expected in the two next decades warrants that a modern, spectral imager/polarimeter would greatly improve our knowledge of Venusian hazes and clouds – ground-based polarimetry would not be able to see the polar regions of Venus, neither be able to operate in the UV range due to Earth’s atmospheric opacity.

C.4. Surface and lowest atmosphere measurements

The surface of Venus experiences maybe the harshest possible conditions at the surface of any celestial body in the solar system. Our most sophisticated, temperature-proof electronics would fail there if not actively refrigerated. Fortunately, remote sensing of this region is possible thanks to the existence of the so-called “night side infrared windows”⁸²: the thick CO₂ atmosphere exhibits some transparency at certain wavelengths that enable the study of the thermal emission of the surface up to an altitude of 15 km, as well as

the lower atmosphere near 15 – 30 km and 30 – 40 km. These windows were discovered only from 1984 onwards¹, so that they could not be studied by spacecraft sent before *Venus Express*. Study of the surface temperature and emissivity, as well as of the atmospheric composition in the 15 – 30 and 30 – 40 km range were then possible and provided very insightful results mentioned here above.

However, further progress would depend on more precise measurements, especially in the shorter wavelengths windows near 1 micrometer that can probe to the surface and the first atmospheric scale height. The Planetary Fourier Spectrometer (PFS) instrument onboard *Venus Express* could have brought some useful data, but unfortunately did never work once the spacecraft was in orbit. Very high-resolution spectroscopy could measure trace species and isotopes (most notably H₂O and HDO) in these windows. Since the atmosphere below an altitude of about 10 km is expected to be convectively stable¹⁶, this means that any relatively gentle (i.e., not perturbing this stable layer by producing a convective plume) outgassing from the interior could remain concentrated for a certain amount of time before being picked up by the convection once reaching an altitude of 10 km, thus keeping it concentrated enough for detection. The detection of locally enhanced water vapour abundance and/or reduced D/H ratio close to the surface would be the unequivocal signature of present-day outgassing of fresh, not yet fractionated water vapor, and thus volcanic activity. This is especially important in the context of the young surface of Venus already mentioned, or to assess if volcanism may be one of the contributing factor to the variability of SO₂ and upper haze some 60 km higher^{58,24}.

The survey of the surface emissivity does not require much spectral resolution and thus could already be done with VIRTIS measurements on board *Venus Express*. However, if some geological activity is present, there shall be noticeable difference between the emissivity maps^{65,80} from 2006-2014 (*Venus Express*-era) compared to a future emissivity map from the 2030s onwards, which would also bring a definitive answer to the question of active volcanism on Venus.

Scientific objective	Required measurements	Instruments
<i>History of the atmosphere</i>		
Initial volatile content	Atmospheric elemental and isotopic composition of noble gases and stable isotopes	<i>MS, GCMS, TDL</i>
Water and volatile history	Isotopic measurements of noble gases and stable isotopes	<i>MS, GCMS, TDL</i>
Climate history. When will Earth also experience a runaway greenhouse?	Detailed characterisation of present greenhouse effect. Understanding of past evolution.	All (except radar)
Past and present-day habitability	Isotopic measurements of noble gases and stable isotopes. Search for organic compounds within droplets	<i>MS, GCMS</i>
<i>Present day climate</i>		
Impact of clouds on the greenhouse effect	Detailed characterisation of cloud properties and processes. Radiative fluxes through the clouds. Temperatures and winds in the cloud layer. Nature of the UV absorber.	<i>Nephelometer, Meteo package, TDL, GCMS, Polarimetric spectral imager</i>
Chemistry in the cloud region and lower/upper atmosphere	Measurements of key trace species. Chemical composition of cloud particles.	<i>XRF, GCMS, TDL, High-res spectrometer, Longwave spectrometer</i>
Super-rotation and heat redistribution on a slow rotating planet	Study of upwards and downwards radiative fluxes. Measurements of temperature and winds.	<i>Meteo package, Radiometer, Polarimetric spectral imager</i>
<i>Geological activity</i>		
Degree of volcanic activity	Thermal remote-sensing of the surface. Chemical and isotopic composition of the first atmospheric scale height.	<i>High-res spectrometer</i>
Surface processes: erosion, lifting, alterations	Surface altimetry and radar reflectivity. Surface infrared emissivity	<i>Radar, Shortwave IR spectral imager</i>

Table 1: Traceability matrix of science objectives. Orbiter-borne instruments are shown in red, and *in-situ* instruments in italic blue.

D. Possible technical solutions for a L-class mission

D.1. UAV or balloon in the clouds

Since remote sensing is unable to measure the abundances and isotopic ratios of noble gases, *in-situ* measurements from within the homosphere (the well-mixed atmosphere below 140 km) are necessary. There are two technical solutions that could be implemented in order to fulfill our scientific objectives: (1) a balloon, whose lift force comes from buoyancy and (2) an Unmanned Aerial vehicle (UAV), whose lift force comes from relative wind speed on the wings.

In order to maximize the other science output, they should fly in the cloud region, where remote sensing is comparatively less easy than above the clouds or deeper in the atmosphere – below the clouds, the temperature rises to values above 130°C where most of the available electronics would fail. In contrast, the pressure and temperature range experienced in the clouds makes it one of the most Earth-like environment in the whole solar system. Furthermore, the greater density of Venusian atmosphere compared to Earth’s (about 1.5 denser in same (P, T) conditions) makes both buoyancy and lifting more efficient. The sulphuric acid in the clouds would not constitute a real threat if suitable exterior coating is applied (e.g. polytetrafluoroethylene, also known as Teflon®).

D.1.a. Comparative pros and cons

The main advantage of the **balloon** technical solution is that it has been already tested in past Venus exploration: the Franco-Soviet *Vega* mission¹⁰. Two helium balloon aerobots were designed to float at 54 km from the surface and had enough battery power for 60 hour of operation with a modest scientific payload (pressure and temperature sensors, light sensor, anemometer, nephelometer). They nevertheless reported surprisingly higher vertical velocities and temperatures contrasts than expected and measured by other descent probes¹¹. Technical expertise is still available at the CNES (Centre National d’Études Spatiales) which was

the prime contractor of these balloon experiments, although more than 30 years old. Recent missions proposals such as the *European Venus Explorer (EVE)* have reactivated this expertise. With technological progress relative to battery mass capacity and solar cell technology since the 1980s, several hundred kilowatts of solar power would be readily available for a total collecting surface on the order of 1 m², thus extending the lifetime of such a mission to several months at least (the limiting factor probably being buoyant gas leaking through the envelope). Drawbacks for this solution are (1) space agencies are generally reluctant to sending large amounts of gas or liquid into space, mostly due to insufficient technical readiness and (2) for a given volume of the envelop, the altitude equilibrium range is very narrow, thus impeding vertical exploration of the atmosphere; this can however be mitigated using a mixed liquid/gas content in the balloon, yielding some control on the cruise altitude.

On the other hand, the **aircraft** solution offers much more flexibility, being able to move freely in the vertical direction between about 50 to 70 km, and offers also some possibility for horizontal maneuvering. Solar designs have been advocated⁵¹, with the electric propeller being supplied through solar panels located on both top and bottom of the aircraft since sunlight is more or less isotropically scattered in the cloud region. Also, no major fluid component would need to be put in orbit. Simple, back-of-the-envelope calculations show that in order to be able to generate enough lift force for a 200 kg total mass, the aircraft would require a true airspeed of about 40 m/s near an altitude of 55 km, which is enough to overcome mean vertical and meridional (north-south) wind speeds, but not enough to counteract the zonal super-rotation. The required mechanical power needed to sustain this airspeed, at a constant altitude, is on the order of several hundred watts, which can easily be supplied by the solar cells on the day side since the solar constant on Venus is twice that on Earth, at least above the clouds. The major technical problems to overcome are night flight: assuming about 48 h spent on the night side, fly-

ing following the super-rotation to the morning terminator, the energy needed to sustain the flight is on the order of 10^8 J, which would require, using Li-Ion batteries, about 250 kg of batteries with present-day technology, outweighing all other components. The aircraft design should therefore incorporate some passive gliding ability, so that the aircraft could be operate at night with a much lower power – a drop of 10 km on Venus only yields about 10^7 J of potential gravitational energy to the aircraft. It should also be capable of fully autonomous flight, since operating from Earth would be highly impractical (light speed travel from Earth to Venus takes, at best, a little more than 2 min). Another unknown parameter is the level of turbulence in the atmosphere of Venus. It is expected to be weaker than on Earth, since no release of latent heat takes place in Venustian clouds (such a release is a major energy source in Earth convective clouds, especially for thunderstorm-generating cumulonimbus). But some cumuliforms clouds have been seen by *Venus Express*⁸⁶, so that we know that local convection exists even at cloud top level. The aircraft should be able to detect and avoid these convective formations well in advance, perhaps with the help of the relay orbiter.

D.1.b. Possible payload

The scientific payload does not depend on the type of platform (balloon or aircraft) finally chosen. It should include at least: (1) a mass spectrometer (MS) and gas chromatograph – mass spectrometer (GC-MS) in order to analyse the composition of the gaseous atmosphere (with an emphasis on the noble gases composition and isotopic ratios) as well as that of the particulate matter; (2) a camera with several monochromatic filters at wavelengths well chosen, in order to assess the variability of key components of the Venustian atmosphere (e.g., the unknown UV absorber) as well as for public outreach, possibly in conjunction with a microphone; (3) a basic meteorology payload to measure winds, pressure and temperature fields; (4) a tunable diode spectrometer (TDL) for precise measurements of key minor species in the gaseous

phase (especially H_2O , HDO and SO_2); (5) an X-ray fluorescence spectrometer to measure elemental abundances in the collected aerosols; (6) a nephelometer with polarization capability in order to characterise cloud and hazes particles; (7) a radiometer in order to study the radiative balance in a selected key broad spectral bands from the UV to the thermal IR. A synthetic view of this payload is summarized in the traceability matrix (Table 1).

A new strategy that could be tested along with the aforementioned measurements is the “dumb rover” paradigm: a demonstration rover built with extremely heat-resistant mechanical parts, as well as minimal electronics, primarily for some basic sensors and communication purposes, would be able to move at the surface. This rover would be piloted from the balloon/aircraft or the orbiter, where complex electronics would be hosted in much safer operational conditions.

D.2. Orbiter

An orbiter should also be part of the mission, considering the available budget for a L-class mission: the whole budget of the *Venus Express* mission was substantially less than the billion EUR budget allocated to a L-class mission). Such an orbiter would have two main uses: (1) act a communication data/relay with the *in-situ* instruments and (2) perform advanced remote sensing in order to meet the scientific targets not obtained from the *in-situ* flying device.

D.2.a. Payload

The primary target for the orbiter would be the study of the surface and the lowest-most 10 km of atmosphere. In order to perform such a study, we would require: (1) a very high-resolution (with $R > 10^5$) spectrometer, or better, imaging spectrometers, in the $0.9 - 2.5 \mu\text{m}$ range. These spectrometers would be able to measure accurately the composition (including isotopic composition) of trace species in all known night-side infrared spectral windows (2) an aperture synthesis radar, comparable to the one used by the NASA space mission *Magellan*. This in-

strument could track small changes in the topography and emissivity by comparison with *Magellan* data (3) a UV to near IR spectrometer/imaging spectrometer with polarimetric capabilities, in order to monitor global changes in the upper cloud region (clouds and hazes, circulation, minor species). (4) a thermal IR spectrometer from $5\text{ }\mu\text{m}$ to $20\text{ }\mu\text{m}$ in order to study the cool upper part of the atmosphere on both day and night sides, in order to supplement ground based observations²³. A synthetic overview is available in Table 1

D.2.b. Orbit choice

Two different strategies are possible for the orbit choice. The first one consists in accompanying the *in-situ* device in its super-rotation in order to maximize the *in-situ* data rate. This strategy only makes sense if the super-rotation is fairly constant, so it requires a constant altitude for the *in-situ* device (most likely a balloon in such a case). If the balloon operates at cloud top level, the super-rotation period is nearing 4 Earth days, yielding an equatorial orbit at an altitude of about 90000 km, which is a little far to achieve a good spatial resolution on the ground level, but provide an outstanding full hemispheric view of Venus (comparable with weather satellites on Earth) as well as a good tracking the super-rotation. Furthermore, an equatorial orbit would greatly impede our ability to monitor polar regions, which are also difficult to monitor from Earth. The second choice is then to adopt a polar orbit at a much lower altitude (several hundred of kilometers). This enables a much better spatial resolution for the observations, as well as a good visibility of both polar regions. Conversely, this requires good data storage on-board the flying device, since the link with the orbiter would only be intermittent in such a case. Depending on the relative emphasis between science objectives, both choices could be valid, including even a polar elliptical orbit close to the one of *Venus Express*.

References

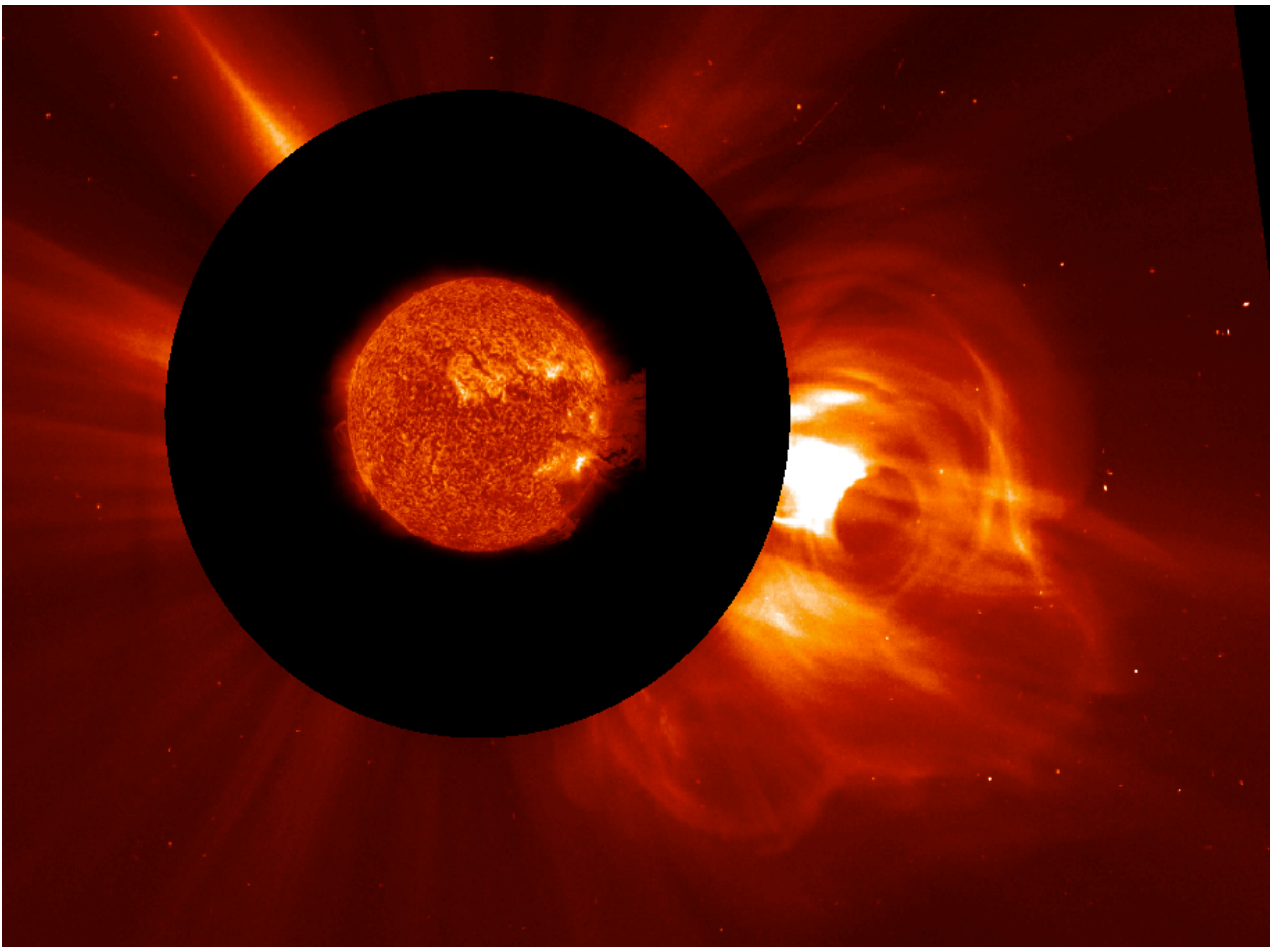
- [1] D. A. ALLEN AND J. W. CRAWFORD. Cloud structure on the dark side of Venus. *Nature*, **307**:222–224 (1984).
- [2] C. J. BALLENTINE; ET AL. Neon isotopes constrain convection and volatile origin in the Earth’s mantle. *Nature*, **433**:33–38 (2005).
- [3] S. BARABASH; ET AL. The Analyser of Space Plasmas and Energetic Atoms (ASPERA-4) for the Venus Express mission. *Planetary and Space Science*, **55**:1772–1792 (2007).
- [4] S. BARABASH; ET AL. The loss of ions from Venus through the plasma wake. *Nature*, **450**:650–653 (2007).
- [5] J. K. BARSTOW; ET AL. Models of the global cloud structure on Venus derived from Venus Express observations. *Icarus*, **217**:542–560 (2012).
- [6] A. T. BASILEVSKY; ET AL. The Resurfacing History of Venus. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 1047 (1997).
- [7] J.-L. BERTAUX; ET AL. SPICAV on Venus Express: Three spectrometers to study the global structure and composition of the Venus atmosphere. *Planetary and Space Science*, **55**:1673–1700 (2007).
- [8] B. BÉZARD; ET AL. Water vapor abundance near the surface of Venus from Venus Express/VIRTIS observations. *Journal of Geophysical Research (Planets)*, **114**:E00B39 (2009).
- [9] B. BÉZARD; ET AL. The 1.10- and 1.18- μ m nightside windows of Venus observed by SPICAV-IR aboard Venus Express. *Icarus*, **216**:173–183 (2011).
- [10] J. BLAMONT AND R. Z. SAGDEEV. The VEGA mission. *Naturwissenschaften*, **71**:295–302 (1984).
- [11] J. E. BLAMONT; ET AL. Implications of the VEGA balloon results for Venus atmospheric dynamics. *Science*, **231**:1422–1425 (1986).
- [12] N. V. BONDARENKO; ET AL. Present-Day Volcanism on Venus: Evidence from Microwave Radiometry. *Geophysical Research Letters*, **37**:L23202 (2010).
- [13] J.-Y. CHAUFRAY; ET AL. Hydrogen density in the dayside venusian exosphere derived from Lyman- α observations by SPICAV on Venus Express. *Icarus*, **217**:767–778 (2012).
- [14] C. S. COCKELL. Life on Venus. *Planetary and Space Science*, **47**:1487–1501 (1999).
- [15] D. CRISP. Radiative forcing of the Venus mesosphere. II - Thermal fluxes, cooling rates, and radiative equilibrium temperatures. *Icarus*, **77**:391–413 (1989).
- [16] D. CRISP AND D. TITOV. The Thermal Balance of the Venus Atmosphere. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 353 (1997).
- [17] L. S. CRUMPLER; ET AL. Volcanoes and Centers of Volcanism on Venus. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 697 (1997).
- [18] C. DE BERGH; ET AL. The composition of the atmosphere of Venus below 100 km altitude: An overview. *Planetary and Space Science*, **54**:1389–1397 (2006).
- [19] A. DOLLFUS. *Synthesis on the Ultra-violet Survey of Clouds in Venus’ Atmosphere*, p. 133. Brand, J. C. and McElroy, M. B. (1968).
- [20] T. M. DONAHUE AND J. B. POLLACK. *Origin and evolution of the atmosphere of Venus*, pp. 1003–1036. Hunten, D. M. and Colin, L. and Donahue, T. M. and Moroz, V. I. (1983).
- [21] T. M. DONAHUE; ET AL. Venus was wet - A measurement of the ratio of deuterium to hydrogen. *Science*, **216**:630–633 (1982).
- [22] K. J. EDWARDS; ET AL. An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science*, **287**(5459):1796–1799 (2000).
- [23] T. ENCRENAZ; ET AL. HDO and SO₂ thermal mapping on Venus: evidence for strong SO₂ variability. *Astronomy and Astrophysics*, **543**:A153 (2012).
- [24] L. W. ESPOSITO. Sulfur dioxide - Episodic injection shows evidence for active Venus volcanism. *Science*, **223**:1072–1074 (1984).
- [25] L. W. ESPOSITO. Long term changes in Venus sulfur dioxide. *Advances in Space Research*, **5**:85–90 (1985).
- [26] L. W. ESPOSITO; ET AL. *The clouds are hazes of Venus*, pp. 484–564. Hunten, D. M. and Colin, L. and Donahue, T. M. and Moroz, V. I. (1983).
- [27] L. W. ESPOSITO; ET AL. Sulfur dioxide at the Venus cloud tops, 1978–1986. *Journal of Geophysical Research*, **93**:5267–5276 (1988).
- [28] S. FUZZI; ET AL. Fog dropletsan atmospheric source of secondary biological aerosol particles. *Atmospheric Environment*, **31**(2):287 – 290 (1997). ISSN 1352-2310.
- [29] P. J. GIERASCH; ET AL. The General Circulation of the Venus Atmosphere: an Assessment. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 459 (1997).
- [30] C. GILLMANN; ET AL. A consistent picture of early hydrodynamic escape of Venus atmosphere explaining present Ne and Ar isotopic ratios and low oxygen atmospheric content. *Earth and Planetary Science Letters*, **286**:503–513 (2009).
- [31] C. GOLDBLATT; ET AL. The Runaway Greenhouse: New Model Results and Implications for Planets and Anthropogenic Global Change. *LPI Contributions*, **1675**:8095 (2012).
- [32] D. H. GRINSPOON. Implications of the high D/H ratio for the sources of water in Venus’ atmosphere. *Nature*, **363**:428–431 (1993).
- [33] D. H. GRINSPOON AND M. A. BULLOCK. The history of water on Venus. *Geochimica et Cosmochimica Acta Supplement*, **69**:750 (2005).
- [34] J. E. HANSEN AND J. W. HOVENIER. Nature of the Venus clouds as derived from their polarization. In A. WOSZCZYK AND C. IWANISZEWSKA, eds., *Exploration of the Planetary System*, vol. 65 of *IAU Symposium*, pp. 197–200 (1974).
- [35] V. L. HANSEN AND J. A. WILLIS. Structural Analysis of a Sampling of Tesseræ: Implications for Venus Geodynamics. *Icarus*, **123**:296–312 (1996).
- [36] G. L. HASHIMOTO; ET AL. Felsic highland crust on Venus suggested by Galileo Near-Infrared Mapping Spectrometer data. *Journal of Geophysical Research (Planets)*, **113**:E00B24 (2008).
- [37] H.-J. HOFFMANN; ET AL. The Pioneer Venus Bus neutral gas mass spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, **18**:122–126 (1980).
- [38] G. HOLLAND; ET AL. Meteorite Kr in Earth’s Mantle Suggests a Late Accretionary Source for the Atmosphere. *Science*, **326**:1522– (2009).
- [39] N. I. IGNATIEV; ET AL. Altimetry of the Venus cloud tops from the Venus Express observations. *Journal of Geophysical Research (Planets)*, **114**:E00B43 (2009).
- [40] A. P. INGERSOLL. The Runaway Greenhouse: A History of Water on Venus. *Journal of Atmospheric Sciences*, **26**:1191–1198 (1969).
- [41] V. G. ISTOMIN; ET AL. Composition of Venus lower atmosphere from mass-spectrometer data. *Cosmic Research*, **17**:581–584 (1980).
- [42] R. JÄRVINEN; ET AL. Widely different characteristics of oxygen and hydrogen ion escape from Venus. *Geophysical Research Letters*, **37**:L16201 (2010).
- [43] J. F. KASTING. Runaway and moist greenhouse atmospheres and the evolution of earth and Venus. *Icarus*, **74**:472–494 (1988).
- [44] W. M. KAULA. Constraints on Venus Evolution from Radiogenic Argon. *Icarus*, **139**:32–39 (1999).
- [45] I. V. KHATUNTSEV; ET AL. Temporal variations of zonal wind speed at Venus cloud tops from Venus Monitoring Camera UV images. In *European Planetary Science Congress 2012*, p. 18 (2012).
- [46] R. G. KNOLLENBERG AND D. M. HUNTEN. The microphysics of the clouds of Venus - Results of the Pioneer Venus particle size spectrometer experiment. *Journal of Geophysical Research*, **85**:8039–8058 (1980).
- [47] R. K. KOPPARAPU; ET AL. Habitable Zones around Main-sequence Stars: New Estimates. *The Astrophysical Journal*, **765**:131 (2013).

- [48] V. A. KRASNOPOLSKY. Chemical kinetic model for the lower atmosphere of Venus. *Icarus*, **191**:25–37 (2007).
- [49] V. A. KRASNOPOLSKY AND G. R. GLADSTONE. Helium on Mars and Venus: EUVE observations and modeling. *Icarus*, **176**:395–407 (2005).
- [50] V. A. KRASNOPOLSKY AND J. B. POLLACK. H₂O-H₂SO₄ system in Venus' clouds and OCS, CO, and H₂SO₄ profiles in Venus' troposphere. *Icarus*, **109**:58–78 (1994).
- [51] G. A. LANDIS; ET AL. Venus atmospheric exploration by solar aircraft. *Acta Astronautica*, **56**:750–755 (2005).
- [52] S. LEBONNOIS; ET AL. *Models of Venus Atmosphere*, p. 129. Bengtsson, L. and Bonnet, R.-M. and Grinspoon, D. and Koumoutsaris, S. and Lebonnois, S. and Titov, D. (2013).
- [53] T. LEBRUN; ET AL. Thermal evolution of an early magma ocean in interaction with the atmosphere. *Journal of Geophysical Research: Planets* (2013). ISSN 2169-9100.
- [54] J. LECONTE; ET AL. 3D climate modeling of close-in land planets: Circulation patterns, climate moist bistability and habitability. *ArXiv e-prints* (2013).
- [55] S. S. LIMAYE. Venus atmospheric circulation: Known and unknown. *Journal of Geophysical Research (Planets)*, **112**:E04S09 (2007).
- [56] E. MARCQ; ET AL. A latitudinal survey of CO, OCS, H₂O, and SO₂ in the lower atmosphere of Venus: Spectroscopic studies using VIRTIS-H. *Journal of Geophysical Research (Planets)*, **113**:E00B07 (2008).
- [57] E. MARCQ; ET AL. An investigation of the SO₂ content of the venusian mesosphere using SPICAV-UV in nadir mode. *Icarus*, **211**:58–69 (2011).
- [58] E. MARCQ; ET AL. Variations of sulphur dioxide at the cloud top of Venus's dynamic atmosphere. *Nature Geoscience*, **6**:25–28 (2013).
- [59] W. J. MARKIEWICZ; ET AL. Morphology and dynamics of the upper cloud layer of Venus. *Nature*, **450**:633–636 (2007).
- [60] K. MCGOULDRIK AND O. B. TOON. Observable effects of convection and gravity waves on the Venus condensational cloud. *Planetary and Space Science*, **56**:1112–1131 (2008).
- [61] W. B. MCKINNON; ET AL. Cratering on Venus: Models and Observations. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 969 (1997).
- [62] S. L. MILLER. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science*, **117**:528–529 (1953).
- [63] F. P. MILLS AND M. ALLEN. A review of selected issues concerning the chemistry in Venus' middle atmosphere. *Planetary and Space Science*, **55**:1729–1740 (2007).
- [64] R. MOISSEL; ET AL. Venus cloud top winds from tracking UV features in Venus Monitoring Camera images. *Journal of Geophysical Research (Planets)*, **114**:E00B31 (2009).
- [65] N. MUELLER; ET AL. Venus Surface Thermal Emission Observed by VIRTIS on Venus Express. *LPI Contributions*, **1470**:45–46 (2009).
- [66] H. F. PARISH; ET AL. Decadal variations in a Venus general circulation model. *Icarus*, **212**:42–65 (2011).
- [67] R. O. PEPIN. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus*, **92**:2–79 (1991).
- [68] R. O. PEPIN. Evolution of Earth's Noble Gases: Consequences of Assuming Hydrodynamic Loss Driven by Giant Impact. *Icarus*, **126**:148–156 (1997).
- [69] R. T. PIERREHUMBERT. *Principles of Planetary Climate* (2010).
- [70] R. T. PIERREHUMBERT. The Runaway Greenhouse: Could it Happen Here? *LPI Contributions*, **1675**:8035 (2012).
- [71] J. B. POLLACK AND D. C. BLACK. Noble gases in planetary atmospheres - Implications for the origin and evolution of atmospheres. *Icarus*, **51**:169–198 (1982).
- [72] D. PORCELLI AND K. TUREKIAN. 4.11 - the history of planetary degassing as recorded by noble gases. In E. IN CHIEF: HEINRICH D. HOLLAND AND K. K. TUREKIAN, eds., *Treatise on Geochemistry*, pp. 281 – 318. Pergamon, Oxford (2003). ISBN 978-0-08-043751-4.
- [73] D. P. R. O. PEPIN. Origin of noble gases in the terrestrial planets. *Rev. Mineral. Geochem.* (2002).
- [74] P. J. RASCH; ET AL. Exploring the geoengineering of climate using stratospheric sulfate aerosols: The role of particle size. *Geophysical Research Letters*, **35**:L02809 (2008).
- [75] S. N. RAYMOND; ET AL. High-resolution simulations of the final assembly of Earth-like planets I. Terrestrial accretion and dynamics. *Icarus*, **183**:265–282 (2006).
- [76] A. SÁNCHEZ-LAVEGA; ET AL. Variable winds on Venus mapped in three dimensions. *Geophysical Research Letters*, **35**:L13204 (2008).
- [77] M. SATO; ET AL. Photopolarimetry Analysis of the Venus Atmosphere in Polar Regions. *Icarus*, **124**:569–585 (1996).
- [78] C. SCHLEPER; ET AL. *Picrophilus oshimae* and *picrophilus torridus* fam. nov., gen. nov., sp. nov., two species of hyperacidophilic, thermophilic, heterotrophic, aerobic archaea. *International Journal of Systematic Bacteriology*, **46**(3):814–816 (1996).
- [79] D. SCHULZE-MAKUCH; ET AL. A Sulfur-Based Survival Strategy for Putative Phototrophic Life in the Venusian Atmosphere. *Astrobiology*, **4**:11–18 (2004).
- [80] S. E. SMREKAR; ET AL. Recent Hotspot Volcanism on Venus from VIRTIS Emissivity Data. *Science*, **328**:605– (2010).
- [81] G. L. STENCHIKOV; ET AL. Radiative forcing from the 1991 Mount Pinatubo volcanic eruption. *Journal of Geophysical Research*, **103**:13837 (1998).
- [82] F. W. TAYLOR; ET AL. Near-Infrared Sounding of the Lower Atmosphere of Venus. In S. W. BOUGHER; ET AL., eds., *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 325 (1997).
- [83] S. TELLMANN; ET AL. Structure of the Venus neutral atmosphere as observed by the Radio Science experiment VeRa on Venus Express. *Journal of Geophysical Research (Planets)*, **114**:E00B36 (2009).
- [84] D. V. TITOV; ET AL. Radiation in the atmosphere of Venus. *Washington DC American Geophysical Union Geophysical Monograph Series*, **176**:121–138 (2007).
- [85] D. V. TITOV; ET AL. Atmospheric structure and dynamics as the cause of ultraviolet markings in the clouds of Venus. *Nature*, **456**:620–623 (2008).
- [86] D. V. TITOV; ET AL. Morphology of the cloud tops as observed by the Venus Express Monitoring Camera. *Icarus*, **217**:682–701 (2012).
- [87] C. C. C. TSANG; ET AL. Tropospheric carbon monoxide concentrations and variability on Venus from Venus Express/VIRTIS-M observations. *Journal of Geophysical Research (Planets)*, **113**:E00B08 (2008).
- [88] A. C. VANDAELE; ET AL. Composition of the Venus mesosphere measured by Solar Occultation at Infrared on board Venus Express. *Journal of Geophysical Research (Planets)*, **113**:E00B23 (2008).
- [89] U. VON ZAHN AND V. I. MOROZ. Composition of the Venus atmosphere below 100 KM altitude. *Advances in Space Research*, **5**:173–195 (1985).
- [90] V. WILQUET; ET AL. Optical extinction due to aerosols in the upper haze of Venus: Four years of SOIR/VEX observations from 2006 to 2010. *Icarus*, **217**:875–881 (2012).
- [91] C. F. WILSON; ET AL. Evidence for anomalous cloud particles at the poles of Venus. *Journal of Geophysical Research (Planets)*, **113**:E00B13 (2008).
- [92] K. ZAHNLE; ET AL. *Emergence of a Habitable Planet*, p. 35. Fishbaugh, K. E. and Lognonné, P. and Raulin, F. and Des Marais, D. J. and Korablev, O. (2007).
- [93] T. L. ZHANG; ET AL. Little or no solar wind enters Venus' atmosphere at solar minimum. *Nature*, **450**:654–656 (2007).

Fundamental Processes in Solar Eruptive Events

Spokesperson: S.A. Matthews
UCL Mullard Space Science Laboratory,
Holmbury St. Mary, Dorking RH5 6NT, U.K.
sarah.matthews@ucl.ac.uk
+44 (0)1483 204208

May 23, 2013



1 Authors and supporters

1. D.R. Williams, UCL Mullard Space Science Laboratory, UK
2. K.L. Klein, LESIA, Observatoire de Paris, CNRS, Universites Paris 6 et 7, France
3. E. Kontar, School of Physics & Astronomy, University of Glasgow, G12 8QQ, UK
4. D.M. Smith, Physics Department and Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, USA
5. S. Krucker, Institute of 4D Technologies, School of Engineering, University of Applied Sciences and Arts North-western Switzerland & Space Sciences Laboratory, University of California, USA
6. A. Lagg, Max-Planck-Institut fur Sonnensystemforschung, Germany
7. G. Hurford, Space Sciences Laboratory, University of California, USA
8. N. Vilmer, LESIA, Observatoire de Paris, CNRS, Universites Paris 6 et 7, France
9. A. MacKinnon, School of Physics & Astronomy, University of Glasgow, UK
10. V. Zharkova, Department of Mathematics, School of Computing, Informatics and Media, University of Bradford, UK
11. L. Fletcher, School of Physics & Astronomy, University of Glasgow, UK
12. H. Hudson, School of Physics & Astronomy, University of Glasgow, UK & Space Sciences Laboratory, University of California, USA
13. I.G. Hannah, School of Physics & Astronomy, University of Glasgow, UK
14. P. Browning, Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, UK
15. D.S. Bloomfield, Astrophysics Research Group, School of Physics, Trinity College Dublin, Ireland
16. D. Innes, Max-Planck-Institut fur Sonnensystemforschung, Germany
17. G. Trottet, LESIA, Observatoire de Paris, CNRS, Universites Paris 6 et 7, France
18. C. Foullon, College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK
19. V. Nakariakov, Centre for Fusion, Space and Astrophysics, Department of Physics, University of Warwick, UK
20. L.M. Green, UCL Mullard Space Science Laboratory, UK
21. M. Mathioudakis, Astrophysics Research Centre, School of Mathematics and Physics, Queen's University, UK
22. A. Gandorfer, Max-Planck-Institut fur Sonnensystemforschung, Germany
23. V. Martinez-Pillet, Instituto de Astrofisica de Andalucia (CSIC), Spain
24. O. Limousin, CEA/DSM/Irfu/Service d'Astrophysique, France
25. E. Verwichte, Centre for Fusion, Space and Astrophysics, Department of Physics, University of Warwick, UK
26. S. Dalla, Centre for Astrophysics, University of Central Lancashire, UK
27. G. Mann, Leibniz-Institut fur Astrophysik Potsdam (AIP), Germany
28. H. Aurass, Leibniz-Institut fur Astrophysik Potsdam (AIP), Germany
29. I. de Moortel, School of Mathematics and Statistics, University of St Andrews, UK
30. B. Kliem, Institut fur Physik und Astronomie, Universitat Potsdam, Germany
31. T. Neukirch, School of Mathematics and Statistics, University of St Andrews, UK
32. B.R. Dennis, NASA Goddard Space Flight Center, Code 671, USA
33. L.K. Harra, UCL Mullard Space Science Laboratory, UK
34. M. Battaglia, Institute of 4D Technologies, School of Engineering, University of Applied Sciences and Arts North-western Switzerland
35. A. N. Zhukov, Solar-Terrestrial Center of Excellence - SIDC, Royal Observatory of Belgium, Belgium
36. S. Zharkov, Department of Physics and Mathematics, University of Hull, UK
37. M. Temmer, Kanzelhöhe Observatory-IGAM, Institute of Physics, University of Graz, Austria

2 Introduction

The Sun is our closest star, and with space now firmly established as part of our society's environment, its unique proximity has inescapable consequences for us. While its radiation provides the energy source of our whole ecosystem, our understanding of how the variations in that radiation control, e.g. our climate, still contains huge gaps. As well as the long term variations in the solar output, the Sun exhibits a cycle of activity the constituents of which are explosive events which release energy. This explosive energy release occurs on a myriad of scales, from nanoflares to huge eruptive flares, which are accompanied by the bulk eruption of plasma and magnetic field known as coronal mass ejections (CMEs) and whose impacts can be seen globally across the Sun and throughout the heliosphere. The most extreme of these events constitute the largest examples of explosive energy release within our solar system, during which upwards of 10^{26} J of energy is released. Electrons are routinely accelerated to tens of keV, occasionally to tens of MeV, while ions achieve energies that may attain several GeV, on timescales of seconds to hundreds of seconds. Acceleration episodes may last only a few seconds, but sometime they may last hours. The accelerated electron component of solar flares contains a substantial fraction of the total energy released, while solar energetic particle (SEP) events contain $\sim 10\%$ of the total energy of the associated CME, for those events that are eruptive. The coupling of so many scales during the course of these eruptions represents a significant challenge to our ability to understand how and where the energy is released, its transport through the solar atmosphere, and beyond.

The ramifications of these explosive events within the heliosphere and the near-Earth environment are considerable. The increased electromagnetic radiation produces significant changes in the Earth's upper atmosphere which impact communications and satellite orbits, with 'immediate' consequences for GPS L-band signals. The fluxes of accelerated particles that are produced during solar energetic particle events (SEPs) lead to significantly increased ionisation in the polar atmosphere, affecting radio transmissions and the chemistry of the upper atmosphere. Particularly for large events, ozone levels can be subsequently affected for months and even years (e.g. Jackman et al. 2000). SEPs also present a high risk to both manned and un-manned space platforms, while the interactions between the magnetic field of the CME and our own magnetosphere can lead to well documented impacts on Earth.

Our ability to probe many layers of the solar atmosphere simultaneously offers unique opportunities to study the relationship between fundamental physical processes such as magnetic reconnection, wave generation and particle acceleration. As well as a compelling need to understand how these processes operate and couple in order to develop reliable tools for the prediction of solar flares & CMES, these are the same processes also operating in other astrophysical environments. On the Sun we can observe their time evolution, and map the evolving magnetic field configuration where the energy release, particle acceleration and subsequent energy transport, takes place. Only by fully understanding the parameter space can we begin to appreciate how these processes are operating in other more extreme astrophysical environments, and critically assess how they might effect the formation of life on planets outside of our own solar system.

Substantial and significant progress has been made in our understanding of eruptive processes on the Sun in recent decades through missions such as SOHO, TRACE, RHESSI, Hinode, STEREO and now SDO, and Europe has played a crucial role in many of these missions both in science (observation and theory and modelling) and engineering. Bold new measurements are now required to move our understanding of the fundamental processes that lead to solar flares & CMEs to the next level. One key way to gain new insights is through the study of previously unexplored wavelength ranges, and the application of new and improved measurement techniques to rarely observed wavelength ranges. An area ripe for such study is the sub-mm spectrum from solar flares (frequencies above 100 GHz) which, until recently, was expected to be the continually decreasing extension of the gyrosynchrotron spectrum from mildly relativistic electrons observed for many years at cm-mm wavelengths. However, new observations brought a major surprise: in the largest flares the spectrum starts to increase with decreasing wavelength (see e.g. Kaufmann et al. 2004). The development of new detectors and techniques for probing this part of the spectrum are now well underway, and offer the potential for a wealth of discoveries about how stars release the energy stored in their atmospheres.

Recent observations from FERMI in the γ -ray range indicate that proton acceleration in solar eruptions is more common than previously thought, and longer lived (Ackermann et al., 2013; Ajello et al., 2013), and RHESSI has shown us unexpected offsets between the locations of γ -ray and HXR emission in some solar flares (Hurford et al., 2006), suggest-

ing differences in the electron and ion acceleration and transport processes that require further investigation. In addition, although we have reasonable diagnostics of > 1 MeV ions, we have few probes of the sub-MeV component. This could imply a very substantial under-estimate of the total energy budget for ions, with implications for viable acceleration processes. New developments in γ -ray imaging and spectroscopy offer the opportunity to probe this comparatively poorly explored regime of the spectrum and to make major advances.

At the heart of understanding all forms of solar activity, and explosive energy release in particular, is the magnetic field. The energy must be built up, stored and then released in the coronal magnetic field, and the subsequent evolution of the event, including the transport of accelerated particles depends critically on the magnetic field topology. New techniques and theoretical advances are now making the possibility of routine measurements of the chromospheric field an achievable reality. This will provide major advances in our understanding of how explosive energy release is triggered, and how CMEs and energetic particles are accelerated.

3 Fundamental questions

Through decades of investigations our picture of solar flares & CMEs has evolved substantially. We now know that they are intimately related, that their visible effects on the solar surface can be global, and that their impacts propagate not just out into the heliosphere, but also down into the solar interior. Although we still have no consistently clear picture of how these eruptions are triggered, we know that there are many scales involved, and thus that we must understand the details of that coupling. Where we once believed particle acceleration to be an exceptionally energetic phenomenon occurring in only a fraction of flares, we now know that energetic particles are seen in nearly every manifestation of magnetic energy conversion, from large flares down to minor explosive events in active regions and sometimes even the quiescent solar atmosphere. The evaluation of the energy budget of flares suggests that non-thermal electrons, and probably also protons, carry a significant fraction of the energy released, and that energy release occurs where the magnetic field changes topology. However, where the energy is stored and how it is released in that magnetic field remains a question, and how the energy is partitioned once released also remains uncertain because of gaps in our coverage of the spectrum, and large uncertainties in some of our measurements. Sound knowledge of this partitioning provides critical constraints for modelling of the entire eruptive process. The overarching questions to be answered are the following:

- How does the magnetic field govern the onset and evolution of solar flares & CMEs?
- What conditions determine the distribution of energy and momentum during the energy release?

The route to answering these questions, as in all areas of science, involves combining all the relevant pieces of the puzzle. By opening new windows in the spectrum and exploiting new advances in measurement techniques over a broad range of wavelengths all the relevant pieces would be in place.

3.1 How does the magnetic field control the onset and evolution of solar flares & CMEs?

The magnetic field is fundamental to our understanding of how solar flares & CMES occur. It stores the energy and defines the environment and direction in which it is released explosively, but still the initiation of explosive energy release on the Sun eludes us, both its prediction and its immediate consequences. The processes involved span many spatial and temporal scales. Ideal MHD processes must play the dominant role in the acceleration and expansion of the coronal mass ejection, while in the accompanying flare there are ideal and resistive MHD, and kinetic processes such as wave transport, reconnection, heating and particle acceleration all at work. How do these processes couple? Is the coupling the same in every flare?

- *What is the pre-eruption magnetic field configuration and evolution?*

The storage and release of energy that powers eruptive flares is intimately tied to the configuration of the magnetic field. Twisted and sheared structures which are current-carrying, and not in their potential state, provide the energy

reservoir. Schrijver (2009) emphasizes the importance of the emergence of highly non-potential magnetic field into an already stressed coronal field for the initiation of the most energetic events. However, while the flux emergence can be observed as it progresses from the photosphere up into the corona, direct measurements of the magnetic field can only routinely be obtained from magnetic field measurements in the photosphere where the coronal field is rooted. It remains the case that we cannot consistently directly determine the magnetic field strength and structure in the corona, where the initial energy release is believed to occur.

Several approaches have been developed to make the best use of the photospheric magnetic field observations available, and to avoid full MHD-modeling of the active region corona. One technique assumes that the corona is force-free and all currents flow along field lines. The vector magnetic field measured in the chromosphere/photosphere gives the lower boundary conditions for the extrapolations. These extrapolated magnetic structures are then compared to EUV and X-ray emission structures that outline the configuration of the magnetic field in the corona. Two important associated assumptions are that the magnetic forces dominate over thermal and gravity forces and that the coronal field adjusts instantaneously to the boundary conditions. These assumptions do not hold for the photosphere, where most of our measurements are made, but are somewhat better in the chromosphere. Estimates of coronal currents are significantly more reliable when extrapolating vector chromospheric fields rather than photospheric fields (Metcalf et al, 2008), and indeed these are essential to understand the magnetic environment in which the energy release occurs.

Many theoretical models for eruptive flares have been developed but they are all united in one aspect - they all invoke a twisted bundle of magnetic field, known as a flux rope, to be present either before the eruption or formed as a result of magnetic reconnection between sheared arcade field lines in the very early stages of the eruption. Therefore, distinguishing between these pre-eruption topologies is key to narrowing in on the physical mechanisms that are responsible. This requires simultaneous observations from the photosphere to the corona to be combined with the extrapolations that give snapshots of the magnetic configuration. Although there are some indications that a flux rope may be present before some eruptive flares from coronal observations (Green et al, 2011), there remain many unanswered questions regarding the exact magnetic topology, flux content, current distribution and the evolution in the run up to the eruption.

Studies of the formation of the flux rope require observations and plasma diagnostics throughout the atmosphere to probe the signatures of reconnection that builds helical field lines from a sheared arcade in the hours or days before the eruptive flare, as in the model of van Ballegooijen & Martens (1989). Magnetic reconnection may also play a key role in destabilising a flux rope through the removal of overlying field. It is not currently possible to probe this occurrence directly with current instrumentation which cannot detect the faint HXR emission from the accelerated particles.

The transformation of a sheared arcade into a flux rope may be the key to understanding why some flares become eruptive whilst others do not. To narrow down competing theories the role of magnetic reconnection needs to be studied along with other mechanisms that build twist and into the coronal field such as the rotation of the photospheric footpoints and the propagation of torsional Alfvén waves from the sub-photospheric portion of the coronal field.

- *What is the partitioning and relative importance of ideal and non-ideal processes at the onset of the eruption?*

The impulsive increase of SXR emission during an eruptive flare is temporally coincident with the period of rapid acceleration of the ejecta of up to velocities of 100 to 1000 km/s (Zhang et al, 2001). In the case of the eruption of a flux rope, the reconnection that occurs in the current sheet beneath is self-amplifying (“runaway”) magnetic reconnection that both powers the flare and transforms the tethers of the overlying magnetic field into poloidal field of the flux rope aiding the eruption. One key unanswered question regards the driver of the eruption and whether an ideal MHD instability, such as the kink or torus instability (Török & Kliem, 2005; Kliem & Török, 2006) is responsible or whether it is the reconnection that acts to cut the magnetic tethers that is the dominant process. There is a close timing and coupling of both ideal and non-ideal processes in eruptive flares.

If reconnection is the driver, the speed of the reconnection outflow should be faster than the rise of the overlying flux rope, while the opposite would be true if an instability is the dominant process. Distinguishing between these two scenarios is challenging due to the close temporal occurrence of a rise of the flux rope and the onset of reconnection and also due to the low emission measure of the reconnection outflow. Combining more sensitive EUV imaging and spectroscopy with higher dynamic range HXR imaging and spectroscopy would allow measurements of the outflow velocities to be robustly made.

The magnetic field can be measured at the photospheric level but in spite of intensive efforts, no unambiguous pre-eruption signatures using photospheric field observations have been found (Barnes & Leka 2008). This indicates that the key physics and signatures of the eruption are likely to be found at higher heights, in the chromosphere and above. The observations detailed above provide an exciting opportunity to finally develop a method of forecasting eruptive flares.

- *How are particles accelerated and transported during the eruptive flare?*

The magnetic field conditions will strongly affect the evolution of the eruption, the distribution of energy and momenta, and the evolution of particle acceleration. In the solar atmosphere we have access to a broad range of diagnostics of particle acceleration. In particular, both the flux and the spectral index of the HXR spectrum provide important diagnostics of the accelerated electrons, and hence of the acceleration process. They provide a means to explore from point to point, and from flare to flare, the acceleration efficiency in footpoint and coronal X-ray sources, and their relationship to the changes in the local magnetic field (Zharkova et al, 2011).

There is clearly a link between the energization of electrons and the changing topology of the field. For example, anti-correlation between magnetic flux transfer rate and minimum overall spectral index (Liu & Wang 2009), points to harder spectra at the outer edges of (the rarely observed) HXR ribbons (e.g Masuda et al., 2001, Liu et al., 2008) and the possible periodic triggering of flares by externally generated MHD waves and anomalous resistivity (Foullon et al. 2005; Nakariakov et al. 2006). However, in practice, while some features of the HXRs are well described by current models, the picture overall is unclear. HXRs, and impulsive optical emission are primarily restricted to compact footpoints, not the long ribbons detected in $H\alpha$ and (E)UV. Even taking into account the restricted dynamic range of current HXR imaging techniques this means that there are highly localised, preferential sites within the overall magnetic field structure where electron acceleration takes place, which map onto these very compact footpoints.

The more elongated $H\alpha$ and (E)UV ribbons that are associated with eruptive flares often exhibit a complex structure, motion across the chromosphere and can last for several hours after the main eruptive phase which indicates a continued release or transport of energy from the corona. A 2D 'standard' model has been developed as a cartoon for eruptive flares which describes the ribbons as being the footprints of newly reconnected field lines in the corona. Recent work has made significant progress in developing a 3D standard model for eruptive flares (Aulanier et al, 2012) and this has given important insight into the evolving electric currents as the eruption proceeds. It has also provided predictions that now need to be tested observationally. High cadence chromospheric vector magnetic field data are crucial to follow the evolution of the electric currents. This model provides a new frame work in which to investigate the locations of the breakdown of ideal MHD and the sites of reconnection in 3D, the rate of reconnection and release of magnetic energy. We must explore the injection of magnetic flux accurately, in order to understand how this controls the evolution of the coronal field. This, together with higher HXR spectral, spatial and temporal resolution, are required to distinguish between different scenarios of particle acceleration and transport.

3.2 What conditions determine the distribution of energy and momentum during the eruption?

The standard thick-target model (e.g., Brown, 1971) invokes a beam of non-thermal electrons as the agent for transporting the energy. However, flares are initiated in a strongly magnetised, low-beta corona, so a large fraction of the magnetic

free energy must be transported away from the immediate reconnection site by kinetic or MHD perturbations (e.g. shrinking loops, Petschek shocks). Coronal structure and topology play an important role in directing these perturbations and converting their energy to that of accelerated particles by generating parallel and perpendicular shock geometries, providing environments where turbulence can be initiated and trapped (LaRosa et al. 1994), trapping energetic particles and generating particular configurations where long-lived acceleration sites can exist and recur. For example, quasi-periodic components in HXR, γ -ray and microwave light-curves suggest either the presence of MHD oscillations or auto-oscillatory regimes of magnetic reconnection (e.g. Nakariakov & Melnikov 2009). Understanding these mechanisms will provide new insights in the basic physics of the energy release and particle acceleration process.

Although we believe the total energy emitted during solar flares & CMEs is well established to an order of magnitude, there remain significant uncertainties in the energies associated with the various different components of the eruption (Emslie et al., 2005), and in the momentum as a whole (Hudson et al., 2012). In particular, our estimates to date cannot assess properly the energy component in the sub-mm/FIR range, the energy in $< 1\text{ MeV}$ ions, nor quantities such as turbulent mass motions. Recent results by Woods et al. (2011) and Su et al. (2012) also highlight that the recently discovered second phase of the flare may contain as much energy as the first phase. A robust determination of the energy and momentum partitioning is critical if we are to constrain the energy release processes involved in the eruption as a whole, and more complete coverage of the whole range of enhanced emission is clearly required. Below we discuss specific questions that will lead to this determination.

- *What is the distribution of non-thermal electrons below 10-20 keV?*

The X-ray spectrum of energetic electrons at energies below 10-20 keV is dominated by the thermal emission from the corona. However, this low energy part of the electron spectrum is the one most affected by transport effects, such as wave-particle interactions and return currents (Zharkova & Gordovskyy, 2006). The electron acceleration rate in a large flare above 20 keV is $10^{34}\text{-}10^{36}\text{ e}^{-1}$, which poses severe difficulties for understanding the electrodynamics of the beams. Their self-induced electric fields should be important, but the dynamic range of current HXR instruments does not allow us to identify the energy loss mechanism(s). An understanding of how the deka-keV electrons evolve can uncover the dominant processes for electron transport in solar flares and hence provide the observations for new models beyond those relying on collisional transport.

- *What is the angular distribution of energetic electrons, and can emission coming from precipitating and returning electrons be distinguished?*

While the energy spectrum of energetic electrons is relatively well known, our knowledge of the angular distribution of energetic electrons is very limited. In the standard flare scenario the accelerated electrons are beamed downwards. Since relativistic electrons emit in the direction of their propagation, this should be reflected in strong directivity and noticeable polarization of the emitted HXR (e.g. Zharkova et al. 2011). Attempts to measure this polarisation signal have been made over the years, but the difficulties are such that most are not considered to be reliable. HXR photons emitted downwards are also effectively scattered by the electrons in the photosphere (both free and bound) and hence a substantial fraction of HXRs will be back-scattered into the observer's direction. RHESSI observations have demonstrated that this Compton backscattering (X-ray albedo) can be detected in a spatially integrated spectrum via a broad bump in the range between 20 and 50 keV (Kontar et al. 2006). However, the observations are inconsistent with strong downward beaming of electrons (e.g. Kontar & Brown 2006). This dilemma questions the standard flare model, and can only be addressed with new albedo, directivity and X-ray polarization measurements. Images of the albedo component are crucial to constrain new physical modelling of the main processes leading to the formation of the observed angular distribution.

- *What is the relationship between thermal and non-thermal plasmas in a flare?*

It is usually assumed that in a flare the majority of the plasma heating in the corona and at the chromospheric foot-points is due to the same electron populations that generate the non-thermal HXR footpoint emission. On the other hand some observational evidence and theoretical reasons suggest that non-thermal distributions can be accelerated

out of pre-heated thermal distributions. Current imaging observations provide a confused picture, showing sources with both thermal and non-thermal characteristics around the presumed energy release sites. Observations with high spatial and temporal resolution and a high dynamic range will be necessary to tie down the spatial and temporal relationships between these components, and to provide constraints for modelling the entire chain of acceleration, particle transport, and energy deposition.

- *How is the flare energy transported from the corona to the chromosphere?*

The plasma of the chromosphere is heated by the energy transported from the primary energy release site in the corona. Thermal far infrared (FIR) radiation from the heated chromosphere is a free-free continuum that is easier to interpret than line emissions, enabling measurements of the energy deposited in deep atmospheric layers. Comparison of the HXR and γ -ray emission from non-thermal particles with the thermal FIR at sub-second resolution will allow us to distinguish between energy transport by conduction fronts and energetic particles.

The relatively low dynamic range of current X-ray observations makes imaging of the HXR emission from the low density corona challenging. Improvements now make dynamic ranges of 50-100:1 possible in SXR, HXR and γ -ray imaging, and will, for the first time, allow us to study how and where coronal energetic electrons and ions are propagating, to produce the first images of the direct and scattered components of the albedo and to distinguish between thermal and non-thermal electron populations. FIR imaging of flares will provide completely new insights into our understanding of the transport of flare energy to the chromosphere.

What is the role of Alfvén waves in energy transport? Birn et al. (2009) showed that during reconnection in coronal conditions Poynting flux is the main downward energy component outside the diffusion region, and Fletcher & Hudson (2008) postulate that this Poynting flux is a plausible additional transport mechanism to the thick-target electron beam. More recent 1-D simulations by Russell & Fletcher (2013) also conclude that Alfvénic waves with periods \leq seconds are capable of heating the chromosphere. To determine whether Alfvén waves are present and what their periods and associated energy flux are we require better spectroscopic diagnostics, coupled with more sophisticated HXR timing information.

- *How is acoustic emission generated by flares & CMES?*

Acoustic emission associated with solar flares was first observed by Kosovichev and Zharkova (1998). Observations shown that the so-called 'sunquake' originates in the impulsive phase of the flare (not necessarily accompanied by an eruption), and that it is generally well-correlated with the locations of enhanced white-light emission, transient changes in the magnetic field, and the hard X-ray footpoints (Kosovichev and Zharkova, 1998; Donea and Lindsey, 2005). Most observations to date have been limited to image cadences of \approx one minute, corresponding to a frequency \leq 8.33 mHz, although observations at 10 mHz have recently been seen by Zharkov et al. (2013).

How the energy and momentum are transferred from the corona into the solar interior is still not well-understood. Three plausible mechanisms have been proposed, including hydrodynamic shock-wave heating originating in the chromosphere (e.g. Kosovichev and Zharkova, 1998), Lorentz forces from the magnetic field reconfiguration (e.g. Hudson, Fisher, and Welsch, 2008), and photospheric backwarming (e.g. Martinez-Oliveros, Moradi, and Donea, 2008). More recently new evidence from Zharkov et al. (2011) suggests a potential link between the sites of the enhanced acoustic emission and the feet of the erupting flux rope. In principle, more accurate determination of energy and momenta of the various different components of the eruptive event, plus the requirement for momentum conservation could distinguish between these mechanisms. The ability to combine seismic observations with high dynamic range HXR and γ -ray imaging and spectroscopy, plus SXR, UV/EUV and WL imaging and spectroscopy from a single space platform would allow us to determine the origins of the acoustic emission.

- *What determines the maximum particle energy?*

The most direct quantitative diagnostics of energetic particles in the solar atmosphere are provided by HXR/ γ -ray observations. They give information on electron and ion energy spectra, numbers and energy contents. While

energetic electrons are characterised by their bremsstrahlung X-ray continuum emission we have to look through a different window at ions: γ -ray lines from 0.4 - 8 MeV tell us about ions above a few MeV in energy, and the ions > 0.2 GeV/nuc are characterised by the continuum above 100 MeV (e.g. Share and Murphy, 2006; Vilmer, MacKinnon, Hurford, 2011). Intriguingly, in a few large flares the energy content of ions above a few MeV is comparable to the total energy (e.g. Murphy et al., 1997) found for electrons from the X-ray continuum. Because of the historically low sensitivity in the γ -ray range, our present knowledge of the acceleration histories of different particles during flares is extremely poor.

Even for those flares that have produced detectable γ -rays a confusing picture has developed regarding the relationship between HXR and γ -ray emissions. There are three well-documented types of observations that appear contradictory: observations by SMM and later RHESSI (see Shih et al., 2009) show that the (event-integrated) fluences of electron bremsstrahlung above 300 keV and of the 2.223 MeV line emission, which is indirectly produced by protons above 30 MeV, correlate well over the entire observable range, which extends over three orders of magnitude. This shows a close physical relationship between the acceleration processes producing relativistic electrons and ions, but the individual studies of large events show differences in the time evolution of electron bremsstrahlung and nuclear line emission. Differences do not only show up in the time evolution: one of the most intriguing results from RHESSI comes from the first imaging observations of a γ -ray line (GRL) event. The 2.2 MeV neutron capture line location was found to be displaced by $20''$ from the centroid of the HXR sources in the 50 - 100 keV range imaged in the same conditions (Hurford et al., 2003). The spatial and temporal differences of emissions from electrons and ions are an unanswered challenge to our present understanding of particle acceleration and particle transport in flares.

Imaging in the GRL domain with RHESSI has been achieved now for 5 events (Vilmer, MacKinnon & Hurford, 2011). Statistically significant displacements between HXR and GRL sources were observed in three of them. In four of the five events where nuclear line emission was imaged by RHESSI, a single unresolved source was observed in the GRL domain. Given the shortcomings of the imaging technique, this is not evidence that the inherent GRL sources are predominantly single sources, a result that would be in stark contrast to typical hard X-ray double sources. Clearly, our understanding of the most surprising feature revealed by RHESSI is limited by the dynamic range and sensitivity of contemporary imaging observations.

The different electron and ion interaction sites observed in a few events can be interpreted as revealing either different electron and ion acceleration sites or showing different transport mechanisms for electrons and ions accelerated in the same site. So far we have only event-integrated images, whereas the full Sun time histories of large flares suggest a strong variation during a given event. The association of images with a higher dynamic range and higher sensitivity in HXR above a few 100 keV and in the GRL domain with high-resolution measurements of the chromospheric magnetic field will provide vital information to reconstruct the path between the regions of acceleration and emission, and to better understand the locations and the conditions for acceleration of electrons and ions in solar flares.

Energy content in non-thermal electrons and ions can be estimated from the analysis of HXR and GRL observations of flares (Emslie et al., 2005), although with quite large uncertainties, in particular for ions because of the unknown lower limit of the accelerated ion spectrum. HXR/ γ -ray observations obtained with high spectral resolution enable detailed analysis of the bremsstrahlung continuum and the resolution of individual γ -ray lines. In the HXR domain bremsstrahlung spectra at high resolution can be directly inverted to get the effective mean electron flux spectrum in the source (see e.g. Piana et al. 2003). The electron flux spectrum is the essential quantity to really constrain acceleration models. Such spectra have so far only been obtained for very few (<10) events (Kontar et al., 2011), due to sensitivity constraints.

High spectral resolution in the GRL domain is essential to better constrain the line fluences and to analyse line shapes. Together with detailed calculations of GRL shapes observations at high spectral resolution provide strong constraints on the ratio of accelerated helium with respect to accelerated protons (α/p ; e.g. Kiener et al., 2006).

However, many parameters determine line fluences and line shapes: the angular distribution of interacting ions, spectral index of the energetic ions and α/p ratio, so that only the combination of line shapes and line fluences can provide strong constraints on those parameters directly linked to particle acceleration and transport models. The number of solar flares for which GRL spectra at high resolution have been obtained is still very small (around five combining RHESSI and INTEGRAL/SPI observations), and even in these cases the separation of the line emission from the bremsstrahlung and broad line background remains a difficult issue. The relative importance with time of the narrow (proton and alpha-particle) and broad (heavier accelerated species) line shapes should also be key evidence for the progress of the energisation process but this sort of study also has been beyond all existing data.

γ -ray lines are not detected in all flares, even those where bremsstrahlung is observed above 10 MeV. We do not know if accelerated ion distributions do not extend into the MeV energy range, or if ion acceleration plays no important role in these events. However, the ions below a few MeV/nucleon are those whose energetics we need to compare with the thermal plasma and the accelerated electrons. A sensitive, low-background spectrometer could measure lower limits on the already well-observed lines. MacKinnon (1988) pointed out the possible diagnostic role of radiative capture lines, which are weak but sensitive to ions in the 0.1 - 1 MeV range.

- *How do flare-accelerated electrons escape into the heliosphere?*

Escape of energetic electrons along open magnetic field lines is normally observed via radio emission, generated by escaping electrons at tens of keV. However, the complexity of the underlying wave-wave interactions currently makes the use of plasma radio emission as a diagnostic tool for the number or energetics of escaping electrons ambiguous. The HXR emission from escaping electrons can be a direct measure of the energetic electrons in the corona, and with the advances in direct HXR imaging now available, combined with spectroscopic measurements these beams could be directly imaged and information on their energy content and spectra derived. Comparison with in situ measurements would provide a powerful tool for understanding the transport processes at work between the Sun and 1AU, and consequently the acceleration mechanisms.

3.3 How and where are energetic particles accelerated in solar flares?

Particle acceleration is a ubiquitous process found everywhere throughout the universe in environments as diverse as stellar coronae, active galactic nuclei, the coronae of accretion disks around black holes, the magnetospheres of neutron stars and planetary atmospheres (including our own) interacting with the wind of their star. It operates both in very dynamic and explosive situations and in more steady phenomena where steep gradients, turbulence and instabilities exist. The product of this process, or processes, are energetic particles that fill the universe and as a result can tell us how that universe originated and what it is made of. They shape the way in which our own and other solar systems work, how planets are formed and what the conditions for the emergence and continuation of life might be. Despite the critical role of particle acceleration in shaping the universe as we know it, the details of the physical processes themselves and how specific environmental conditions affect their onset and evolution remain poorly understood.

- *Is particle acceleration a universal process during energy release in the solar atmosphere or are special conditions required to make it occur?*

Particle acceleration can clearly no longer be understood as a process that operates only in particularly strong events. We do not know if it is the primary phenomenon of the conversion of magnetic energy, with the accelerated particles accounting eventually also for the presence of the hot X-ray emitting plasma or the heating of the chromosphere, or if there are cases where the energy release is primarily in the form of bulk heating. We may expect that in small events the different manifestations will be easier to disentangle. X-ray spectroscopic imaging with increased sensitivity is needed to identify the presence or absence of non-thermal electron populations in small energy conversion events. This will show if, and under which conditions, purely thermal events of energy conversion occur in the solar atmosphere.

- *What are the similarities and differences of magnetic field configurations involved in small and large events?*

The identification of flare-like phenomena in the chromospheric network suggests that small-scale energy release may occur in different configurations than major flares. The comparison of X-ray imaging of small events with the magnetic field structure that eventually governs the process of energy release is a crucial element of the study of similarities and differences of particle acceleration scenarios.

- *How, when and where does the Sun accelerate electrons to 100 MeV and ions to GeV energies?*

GeV protons have been shown for several decades to be present in large solar flares & CMES by ground-based neutron monitors. γ -ray emission from similarly high-energy protons and neutrons above 100 MeV were first detected aboard SMM/GRS (Chupp & Ryan 2009; Vilmer, MacKinnon and Hurford 2011). The existence of such high-energy particles is a considerable challenge for our contemporary understanding. At γ -ray energies the bremsstrahlung continuum from electrons is dominant below 1 MeV and again significant in the 10-50 MeV range. When ions over a few hundred MeV/nuc are produced, their interactions with the ambient medium produce pions whose decay products lead to a broad-band continuum at photon energies above 10 MeV (with a broad peak around 70 MeV from neutral pion decay photons; e.g. Vilmer, MacKinnon and Hurford, 2011) and also neutrons which, if energetic enough, may escape from the Sun and be directly detected (e.g. Chupp and Ryan, 2009). All observations of pion decay photons have come so far from very large flares. It is unclear if this is because only large flares accelerate ions to GeV energies, or because of instrumental threshold effects. Threshold is clearly not the only reason: the quantitative analysis of a few events exhibiting significant pion production generally led to the conclusion that the ion energy distribution does not have a simple power-law form from the γ -ray line emitting energy domain (1-10 MeV) to the pion producing energy domain above 300 MeV, but may involve breaks and cutoffs (e.g. Dunphy et al. 1999; Vilmer et al. 2003). These analyses mostly rely on observations by different instruments. Major contributions from Fermi LAT to high-energy solar γ -ray emission studies are starting to emerge, but observations from 150 keV to 150 MeV with a single instrument, as well as observations of the environments where particles are accelerated to such high energies, and of the regions where the emissions are produced are necessary to make further breakthroughs.

A different, and hitherto unexploited, tool to probe relativistic electrons and positrons in flares is their gyrosynchrotron emission at far infrared wavelengths (THz frequencies). The emission may be generated by the most energetic electrons (tens to hundreds of MeV) produced in solar flares, or by relativistic positrons resulting from nuclear reactions of high-energy protons and α particles (>300 MeV/nuc).

The maximum particle energy attained during solar flares should be reflected by a maximum photon energy, but this has never been clearly observed. The relative magnitudes of the charged and neutral pion decay spectral components also carry information on the ion distribution, as well as the magnetic field where the radiation is produced. All existing observations have yielded spectra with rather broad photon energy channels, limiting the extent to which these diagnostics could be used. A systematic survey is a pre-requisite to understanding the importance of relativistic particles in the flare process.

- *What are the time scales of particle acceleration at the highest energies?*

A comprehensive study of the timing of electron bremsstrahlung, pion decay radiation, and the gyrosynchrotron radiation of relativistic electrons is required to assess the acceleration times at relativistic energies. It came as a huge surprise when the Compton Gamma-Ray Observatory and Gamma-1 observed pion decay continuum during several hours after the start of a flare (e.g. Kanbach et al. 1993). Why such high-energy particles are seen in the solar atmosphere in the apparent absence of other manifestations of a flare (Ryan, 2000) presents us with a major challenge. The often quoted alternative of acceleration at a CME-driven shock faces problems, because particles have to reach the low solar atmosphere to undergo the nuclear reactions producing the pion decay γ -rays. The 20 January 2005 flare demonstrated the possibility that very high-energy ions are accelerated in the flare process,

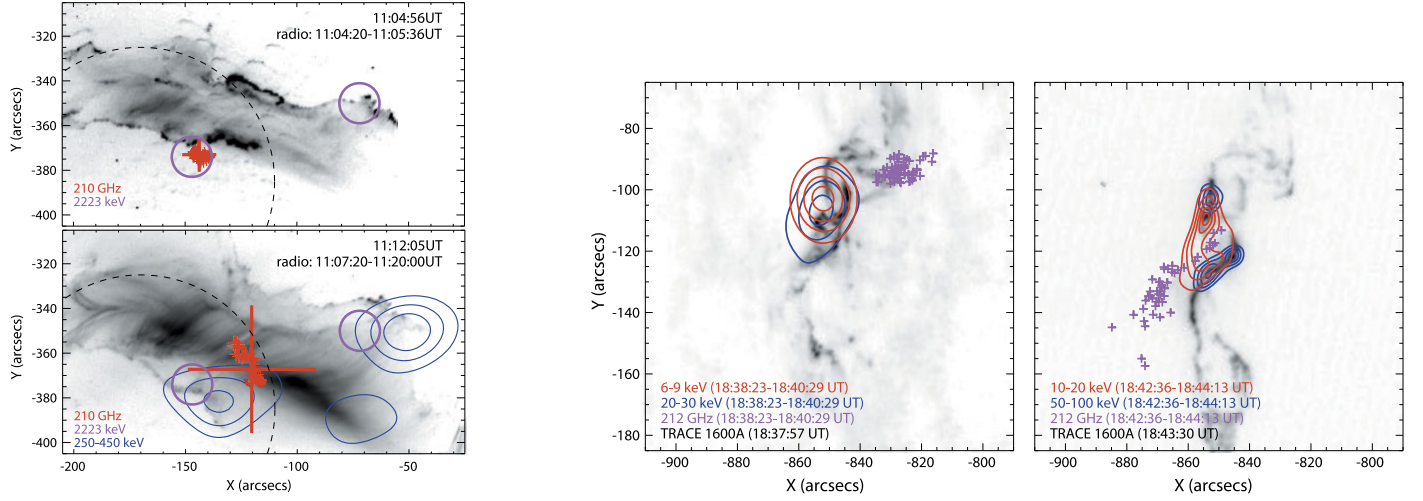


Figure 1: Left: Superposed images of TRACE (UV; grey-scale), RHESSI (blue: 250-450 keV, pink: 2.223 MeV) contours from Hurford et al. 2006) and KOSMA (210 GHz, red crosses) during a large solar flare (adapted from Trottet et al., 2008). Right: TRACE 1600 Å images of 6 Dec 2006 flare during the rise phase of the event (left), and impulsive peak (right). Red contours show thermal soft X-rays, blue contours show non-thermal hard X-rays, purple crosses show the 212 GHz source from SST data at different times (adapted from Kaufmann et al., 2009)

rather than in the interplanetary medium following a flare (e.g. Grechnev et al. 2008). Observations of the high-energy emissions will be essential in unraveling the relative roles of flare and interplanetary medium processes in accounting for high-energy ions, a question that is also crucial to understanding the long-duration events.

- *How is the FIR/sub-mm emission of non-thermal particles generated in flares?*

The sub-mm spectrum from solar flares (frequencies above 100 GHz) was, until recently, expected to be the continually decreasing extension of the gyrosynchrotron spectrum from mildly relativistic electrons. However, in the largest flares the spectrum starts to increase with decreasing wavelength (see e.g. Kaufmann et al. 2004). Interestingly, these events are also found to be associated with significant pion-decay radiation (Trottet et al. 2008). The origin of this new component is hotly debated, as summarised by Krucker et al. (2013). One possibility is that this emission is produced by synchrotron emission from pion-decay positrons (e.g. Lingenfelter and Ramaty, 1967). This interpretation is supported by the unique observation combining spatial information of the emission at 210 GHz with emission from energetic ions in the γ -ray line domain. Figure 1(left) shows impulsive emission at 210 GHz (red cross, top panel) from the site of interacting ions (pink contours), not from the site of interacting electrons (blue contours, bottom panel) from the flare on 28 October 2003. On the right in figure 1, time-resolved imaging of the 212 GHz emission in relation to soft and hard X-rays is shown for another flare on 6 December 2006. Kaufmann et al. (2009) report a displacement of $30''$ from the rise to the peak of this flare, with the source motion primarily parallel to the flare ribbons. No clear correlation is seen between the 212 GHz and other emissions in this event. But while the close correlation in time and space of the impulsive sub-mm emission and the strong production of neutral pions in the 28 October 2003 flare suggests synchrotron emission from charged pion-decay positrons, the sub-mm flux predicted from the number of positrons derived from charged pion decay seems inadequate to account for observed fluxes. The alternative interpretation (e.g., Silva et al. 2007; Trottet et al. 2008) that the sub-mm emission is gyrosynchrotron emission from non-thermal electrons would then suggest that electrons in different energy ranges are accelerated in different regions. These as well as the other interpretations have a number of weaknesses (Fleishman & Kontar, 2010) and new observations into the THz range promise to reveal tantalising new insights..

- *How and where are SEPs accelerated, and what is their relationship to trapped accelerated particles?*

SEP events represent one of the greatest threats to our increasing reliance on space-related assets. While their maximum energy does not reach that of cosmic rays, they also represent some of the most energetic particles in the solar system, and the opportunity to probe in detail the underlying acceleration processes in order to inform our understanding of cosmic ray acceleration. They include the so-called impulsive events which are short-lived and display enhancements of ^3He to ^4He , with a strong connection to EUV jets observed at the Sun (e.g. Pick et al., 2006); electron dominated events, which have a strong association to type III radio bursts; gradual events which can last for several days and have been observed to display abundance variations, and hybrid events which show characteristics of more than one of these types. Recent observations with STEREO have also shown that the longitudinal extent of these events can be as large as 360° , making their potential impact throughout the heliosphere event greater (e.g. Rouillard et al., 2011). While the largest and longest lived SEP events are believed to be accelerated by CME associated shocks, there is still considerable debate about the connection between the escaping particles and those that remain trapped in the solar atmosphere; over the origin of impulsive SEP events which do not have an accompanying CME, and over the relative importance of perpendicular versus parallel shocks. Major unanswered questions include whether flare accelerated particles provide a seed population for shock acceleration; whether the population of impulsive SEP particles is the same as the population that remains trapped at the Sun that radiates in HXR and γ -rays, and why not every CME driven shock produces an SEP event.

In the case of what determines whether or not a CME will produce an SEP, many previous studies have shown that there is a large (up to three orders of magnitude) variation in SEP intensity for a given CME speed. This suggests either an acceleration mechanism that isn't related to the CME at all, or a dependence on the existence of a seed population of lower energy particles from a previous event, or series of events. Kahler (2001) found that intensity of particles remaining from previous events was the only discriminating factor in sorting intensities. More recently, a study by Mewaldt et al. (2012) found that upper limits can be placed on the Fe 10-40 MeV fluence on the basis of measurements of the intensity in 0.01-2.0 MeV/nucleon energy range measured at 1AU in the previous 24 hours, highly suggestive of the importance of a seed population. Currently, however, we are only able infer the presence of such a seed population in the corona where the acceleration takes place; actual measurements are only available at 1 AU. A recent work by Laming et al. (2013) proposes a method for detecting the existence of a seed population through measurements of the enhanced wings of the Lyman α line profile in the region between 1.5 - 3.0 R_s . Different origins of the supra thermal population would produce different spectral and temporal signatures, allowing them to be distinguished. For example, a population of protons produced by decaying solar neutrons would be produced within a few hours of the associated flare, whereas a population produced by continuous large-scale reconnection in the corona or magnetic pumping would not have such a time dependence. Combining γ -ray imaging and spectroscopy with observations of the Lyman α line profile could be a powerful tool in understanding whether such seed populations exist, where they originate and how they vary, ultimately providing us with the means to predict which CMEs will subsequently produce SEPs.

The association of SEP angular extent with global EUV waves and CME expansion by Rouillard et al. (2011) and Dresing et al. (2012) has been suggested to support shock acceleration as the primary agent for the origin the events studied by these authors. However, recent work by Masson et al. (2013) raises interesting questions about the role of interchange reconnection during the CME expansion. More sensitive HXR and γ -ray imaging and spectroscopy, coupled with EUV imaging and spectroscopy, magnetic field extrapolations, and coronagraph information on the CME expansion will allow us to determine the roles of both shock and reconnection driven acceleration in these important events.

4 Observable parameters and general measurement strategy

In order to constrain the mechanisms responsible for eruptive events in the solar atmosphere it is necessary to make direct measurements of directly measure the physical conditions present in a) the energy release and acceleration sites, and b) the environment in which the accelerated particles propagate and deposit their energy. A dedicated set of instruments on a

single platform can achieve this through the combination of imaging and spectroscopy in wavelengths that sample the vast range of energies over which energy release signatures are observed. Making these measurements requires a combination of spectroscopy and imaging that can rapidly probe the energies and fine spatial scales over which particle acceleration signatures are observed. Such measurements can now be accomplished by a single payload of dedicated instruments.

The chromospheric and photospheric magnetic field vector needs to be sampled quickly at spatial scales of a few hundred kilometres to provide details of the pre-eruption topology, its evolution during the eruption, and the energetic particle precipitation and energy dissipation processes. The vector field measurements will provide vital information on the evolution of the magnetic field through the atmosphere and its role in the initiation of the energy release (e.g. e.g. the identification of structures such as coronal nulls and quasi-separatrix layers) and subsequent evolution. Similar sub-arcsecond scales are needed to image in the photosphere (WL), chromosphere (UV and HXR) and corona (SXR) to supply important information on energy partition, and the spatial scales at which the accelerated particles deposit their energy.

The properties of the spectra of accelerated particles spectra should be probed through the combination of high resolution HXR and γ -ray spectra and with observations in the FIR/sub-mm. Through the application of both magnetography and imaging in the SXR, HXR, γ -ray and FIR range, transport effects can be separated from changes in the initial acceleration process. This would provide an unprecedented opportunity to observe and constrain the particle acceleration processes operating in the solar atmosphere. Determining the location as well as the time-varying spectrum and directivity of accelerated ions and electrons requires us to make spatially resolved observations in HXR and γ -rays. Although there have been good reasons to believe that the acceleration region lies in the tenuous corona, while the strongest signatures of the accelerated particles are observed in the lower and denser chromosphere, more recent work is beginning to unsettle this picture (e.g. Fletcher & Hudson, 2008). Current indirect indirect HXR imaging instruments have notoriously limited dynamic range that, which makes it impossible for them to simultaneously image X-rays produced by electrons in or close to the acceleration site at the same time as in the sites where the particles precipitate, and thus. Because of this, it is almost impossible to deconvolve the properties of the acceleration mechanism from the effects of transport effects. To break through this barrier, an improvement in sensitivity and dynamic range of order 50 - 100 times is required. This improvement can now be achieved through the use of direct HXR imaging techniques employing grazing-incidence optics.

Increased sensitivity in γ -ray imaging spectroscopy is needed to provide a step change in our ability to image γ -rays, allowing us to produce a time series of images during X and high M-class flare events. This will provide vital clues to the relationship between ion and electron acceleration mechanisms, and strong constraints for current models. Solar flares have never been imaged before in the FIR range.

The observable parameters necessary to achieve these science goals must include:

Simultaneous rapid (dekasecond) and fine (sub-arcsecond) structure measurements of the vector magnetic field in the chromosphere and photosphere: these measurements are required to determine the magnetic environment where energy is released and particles are accelerated via modelling of the coronal field. Chromospheric magnetic field measurements are an essential new achievement, which is required for more realistic coronal magnetic field extrapolations, because the chromosphere cannot be considered as a force-free boundary.

Rapid photon spectra in the energy range 4 keV – 150 MeV: These measurements are the key to establishing the energy distributions of accelerated electrons and protons. Going beyond 100 MeV will let us identify the spectral signatures of relativistic protons; extending and to low energies probes where the non-thermal electron spectra join the thermal background.

Simultaneous images in the range 1 keV – 30 MeV capable of viewing faint and bright X-ray and γ -ray sources within the same field of view: These observations provide source localisations for both electrons and ions, and contain at low energies the likely transition between non-thermal and thermal electron populations, as a crucial ingredient to investigating the role of non-thermal particles in the energetics of the solar atmosphere. The imaging of nuclear line emissions, together with the extended spectroscopy addressed in the previous item, ensure the substantial improvement of the radiative diagnostics of protons.

The brightness distribution of the Sun in far IR wavelength ranges: This is a first in solar flare observations, intended to open the FIR spectrum as a novel diagnostic of the electrons, and possibly positrons, of the highest energies produced in solar flares. Furthermore, thermal bremsstrahlung in this range is an alternative tool to probe energy transport

during flares, because it ties the response of the chromosphere to energy deposition coming from the corona.

Measurements of the X-ray polarization in the energy range 5 – 200 keV and γ -ray imaging polarimetry between 300 – 1000 keV: This long-neglected diagnostic will provide a unique perspective on the angular distribution of high-energy accelerated electrons.

High spatial and temporal resolution images of the photosphere in the 300 – 530 nm range: These will reveal the transfer of energy and momentum to the lowest layers that can be reached by a flare. In the context of the other observation this diagnostic will provide constraints on the extension and location of the energy release sites, and will show how flare-accelerated particles relate to the still poorly understood white-light signatures of flares and to the dynamical photospheric phenomena revealed by sunquakes.

High spatial and temporal resolution images of the corona in the EUV range ($T = 10^4 - 10^7$ K with a full-Sun FoV): As already established with STEREO and SDO these reveal the global nature of flares & CMES, connections between remote magnetic environments, and the extent of particle acceleration sites. They are essential context information for all other measurements.

High spatial and temporal resolution UV/EUV imaging spectroscopy covering temperatures between 10^4 and 10^7 K: This will provide essential diagnostic information on temperatures, densities, velocities and abundances in the energy release and particle acceleration sites, including identifying shocks and turbulence, and links between interacting and escaping particle populations.

WL measurements of the erupting CME as it propagates out into the heliosphere: This will reveal the expansion of the CME in relation to signatures observed in the lower corona, and the existence and extent of shocks driven by it.

5 Overview of a strawman payload

Many decades of observing the Sun have shown that co-ordinating observations of transient events such as flares & CMES with different observing platforms, while achievable, is notoriously difficult. Even with much improved pointing stability and information for space-based observations, co-alignment of datasets also remains subject to errors. Although of relatively minor importance for large-scale phenomena, this can be a serious impediment when studying small-scale processes. A single platform with co-ordinated instrumentation is essential for producing the best and most comprehensive datasets for studying the underlying processes in flares & CMES.

Matthews et al. (2012) outlined the SPARK (Solar Particle Acceleration Radiation and Kinetics) concept in response to ESA's call for M3 missions in 2010. Additionally, in response to the call for white papers in the 2010 Decadal Review, Lin et al. (2011) outlined a more comprehensive, but complementary mission concept called SEE2020. Both of these concepts put forward strawman payloads that would be capable of making the measurements required to address the science questions outlined in this white paper. Given that the scope of the science in this white paper exceeds that outlined in the SPARK proposal, it would be necessary to augment that payload with a full-Sun EUV imager, EUV/UV imaging spectrometer and a coronagraph, which rather naturally leads to a very similar payload to that in SEE2020. We would envisage that should a mission to address the fundamental process in solar flares & CMES be selected, that there would be only one, and that it would involve collaboration between ESA and NASA at some level. However, Europe has a strong heritage in all the instrumentation concepts outlined here, from involvements in SOHO, STEREO, Hinode, Solar Orbiter and other astrophysical observatories.

A strawman payload that could achieve the required measurements could include the following:

- **FIR/Sub-mm imaging:** DESIR (Detection of Eruptive Solar Infrared Radiation) was proposed as a two-bandwidth radiometer which provides full-disk observations of the Sun in the far infrared domain. The instrument would image the solar disk on two detectors in the spectral bands 25 – 45 μm and 80 – 130 μm , after spectral selection and splitting. DESIR would be an un-cooled instrument using thermal detectors (micro-bolometer matrices). The design is one that was developed at LESIA/Observatoire de Paris in the framework of the SMESE micro-satellite project. SMESE was declared ready for a phase B study by CNES, but not continued for programmatic and financial reasons.

Other possibilities for this region of the spectrum include developments based on Microwave Kinetic Inductance Detectors (MKIDs, Mazin, 2009), which are high-sensitivity, photon-counting, energy-resolving detectors. While they need cryogen to operate: the Cooper pairs that they rely on have bonding energies of milli-electron volts, so room temperature disrupts this. However, their quantum efficiency remains useful ($>20\%$) over a large range of wavelengths, from approximately 200 nm to 3000. Their current energy resolution - at the level of several percent - is too weak for use in spectrometers, but would readily be useful in measuring the shape of continuum radiation from, e.g., gyro-synchrotron emission in flares or other, more exotic processes. Measuring this with even coarse energy resolution would be sufficient to measure the broad shape of the spectrum, and access a wealth of untapped information on the process of converting magnetic energy, but it is likely that this would improve on the timescales relevant to L3 and L4.

MKIDs detectors are also being extended to higher-energy domains (Mazin et al., 2010), since they are not limited by band-gap considerations. MKIDs could be deployed in other areas of the solar spectrum, like the EUV, X-ray and gamma ray regimes, where their ability to count photons would be highly desirable in searching for faint sources, including reconnection outflows, which are thought to have low emission measure, and have escaped detection so far.

- **ChroME (Chromospheric Magnetism Explorer):** The ChroME concept was developed by the team at MPS- Lindau, and is intended to simultaneously measure the chromospheric and the photospheric magnetic field vector. A filtergraph type instrument would allow the polarization state of the light to be quantified down to a level of 10^{-4} of the continuum intensity in the chromospheric Mg II k 279.6 nm line and a photospheric line in the visible, the baseline choice being the Fe I 525 nm line. The polarization signal is created by the Zeeman effect, atomic level polarization, scattering polarization and is modified by the Hanle effect. The theoretical modeling of these effects allows the retrieval of vector magnetic field maps in the chromosphere and the photosphere simultaneously.

ChroME would be capable of achieving a spatial resolution of $0.28''$, spectral resolution of 50 mÅ and a signal-to-noise ratio (SNR) of 300 per $0.15''$ pixel in the chromospheric channel. The nominal cadence is 30 s for 15 wavelength positions, and the field of view is approximately 300×300 arcsec. The photospheric channel is envisaged to operate approximately in the 300 to 530 nm range, with the Fe I 525 nm line the most likely choice. A spatial resolution of $0.56''$ would be achieved in this channel. This line provides reliable photospheric vector magnetograms (e.g. Solanki et al. 2010) as well as imaging of white light flares.

It would be highly desirable for this design to be augmented to provide photospheric velocity measurements in order to carry out local helioseismology.

- **Super-FOXSI (Super Focusing Optics X-ray Solar Imager):** Super-FOXSI builds on the successful heritage of the FOXSI sounding rocket program (Krucker et al., 2011) and uses grazing incidence focusing optics to image directly in HXR over the energy range 4 - 60 keV with a spectral resolution of 1 keV and a spatial resolution of ~ 8 arcsec. Current HXR imaging measurements are obtained through non-focusing rotation modulation collimator (RMC) imaging techniques, such as those used on RHESSI (Lin et al. 2002). However, these, and other types of non-focusing imaging methods, have three main limitations that can be overcome by the use of focusing optics: effective area; large non-solar background, and limited dynamic range. The dynamic range for RMC imaging is limited by emission from the entire field of view. For focusing optics, the dynamic range (and resolution) is set by the shape of the point spread function (PSF), thus the sharper the PSF, the better the dynamic range. Current HXR focusing optics already provide a dynamic range > 50 (indirect imaging methods such as RHESSI < 30), which with appropriate calibration procedures can be increased to > 100 for events with good statistics. For well-separated sources, this range can be extended to > 1000 .

Super-FOXSI would be an updated instrument that uses similar modules to FOXSI. The effective area would be increased by a longer focal length, by filling the modules with the largest possible number of shells (which almost doubles the effective area at 10keV and adds response to higher energies), by using multilayer coatings and by

increasing the diameter of the instrument. FOXSI's Si strip detectors would be replaced with 60 μm CdTe strip detectors. CdTe detectors have already been investigated by ISAS for the HXI system on Astro-H and tests of 400 μm pitch CdTe detectors have resulted in an energy resolution of 1.5 keV at 60 keV, an angular resolution < 10 arcsec and a low-energy cut-off of 2 keV. The 60 μm pitch detector has been fabricated and its spectral resolution is currently being evaluated. Hence, small pitch CdTe strip detectors will be available on the required timescales. A long focal length of the order of 10 – 15 m is highly preferable in order to improve the high energy response of the telescope. Expandable booms of such length have been successfully used for the NuSTAR mission (NASA SMEX) that was launched in 2012.

- LISSAN (Large Imaging Spectrometer for Solar Accelerated Nuclei): LISSAN combines two sensors in order to provide indirect imaging and spectroscopic information in the HXR and γ -ray regimes. The γ -ray sensor performs γ -ray spectroscopy (150 keV to ~ 150 MeV) with high-resolution scintillators (LaBr₃). Imaging over this range is produced by placing pairs of grids above the detectors that produce Moire patterns on the detector surfaces; these patterns can be resolved by coarsely segmenting the detectors. This is an indirect (Fourier) imaging technique closely related to that used on the STIX instrument for Solar Orbiter (Benz et al., 2012). For LISSAN, in order to modulate photons across the entire energy range of interest, the grids are made from tungsten slats 3 cm deep. LISSAN's HXR sensor uses the same imaging technique in a much smaller package with cadmium zinc telluride (CZT) or Cadmium telluride (CT) detectors and lightweight grids, based on the STIX design (Grimm et al., 2012). The HXR sensor, with energy range ~ 5 keV to 200 keV, bridges the energy gap between LISSAN and FOXSI (60 - 150 keV), a particularly important range for understanding the nonthermal component in large flares, which often feature a downward break in this band that is not yet understood. The LISSAN HXR sensor will be able to image down to 1'' FWHM, producing the sharpest solar images ever produced in the non-thermal high-energy band for comparison with optical and (E)UV data.

The achieved energy resolution would be better than 1.5% FWHM at 6.1 MeV, sufficient to measure the Doppler profiles of the C and O lines at 4.4 and 6.1 MeV. LISSAN can also function as an imaging polarimeter from 300 to 1000 keV. Scaling from RHESSI's polarization analyses by taking the relative energy ranges, detector areas and geometry into consideration (but neglecting systematic errors in a non-rotating instrument), LISSAN should have a polarization sensitivity about 4 times better than RHESSI for a given event.

With this instrumental configuration we expect to be able to image ion interaction sites in the 2.2 MeV line from neutron capture down to flares of GOES class approximately M3. For X-class flares, LISSAN would produce up to a few dozen frames, showing the evolution of ion interaction sites with time during the flare. The C and O nuclear lines, whose Doppler profiles indicate the angular distribution and He/H ratio in accelerated ions at the interaction site, will have sufficient statistics for meaningful analysis of flares of magnitude M5 and higher. The finest grids in the gamma-ray sensor will allow imaging down to 8'' FWHM, a resolution more than adequate for separating foot-point pairs and able to resolve foot-point sizes in some cases.

- Full Sun EUV imager: A full sun FoV EUV imager of similar type to AIA, providing sub-arcsecond spatial resolution at approximately 10s cadence and broadly covering the temperature range from 10^4 - 10^7 K would be necessary. Europe has substantial expertise from SOHO, STEREO, Hinode and Solar Orbiter in this area, which would allow such an instrument to be developed.
- UV/EUV imaging spectrometer: A large FoV UV/EUV imaging spectrometer covering the range 10^4 - 10^7 K with arc second resolution would be required to provide measurements of line profiles, temperatures, densities, velocities and abundances. Again Europe has strong heritage in this area from SOHO, Hinode and Solar Orbiter, and has been involved in the development of similar instrument concepts for Solar C.
- WL Coronagraph: A white-light coronagraph with a FoV from 1.02 to approximately 10 solar radii would also be required. Again Europe has extensive relevant heritage from SOHO, STEREO and the initial development of PROBA-3.

6 Summary and Conclusions

Explosive energy release on the Sun has contributed, and continues to contribute to shaping our solar system. Along with the long period variations in the solar radiation output and the solar wind that control climate, they provide short but intense bursts of enhanced radiation and energetic particles which impact the magnetospheres and upper atmospheres of all the planets in the solar system, and have helped to shape the way these planets evolve. More than 150 years have passed since the first observation of the largest eruptive event ever recorded (e.g. Gonzalez et al., 2011), and although many advances have been made, many questions remain unanswered. In a world in which our reliance on space-based assets is continually growing, and the quest for signatures of life in other solar systems gathers momentum, we cannot afford not to understand what drives these events and to be able to predict them. New observation windows and more sensitive observing techniques are now possible, with further advances on the 5-10 year horizon, and thus we propose a theme which targets these areas in order to provide the most complete picture of the high energy processes operating in solar flares and CMEs ever achieved.

7 References

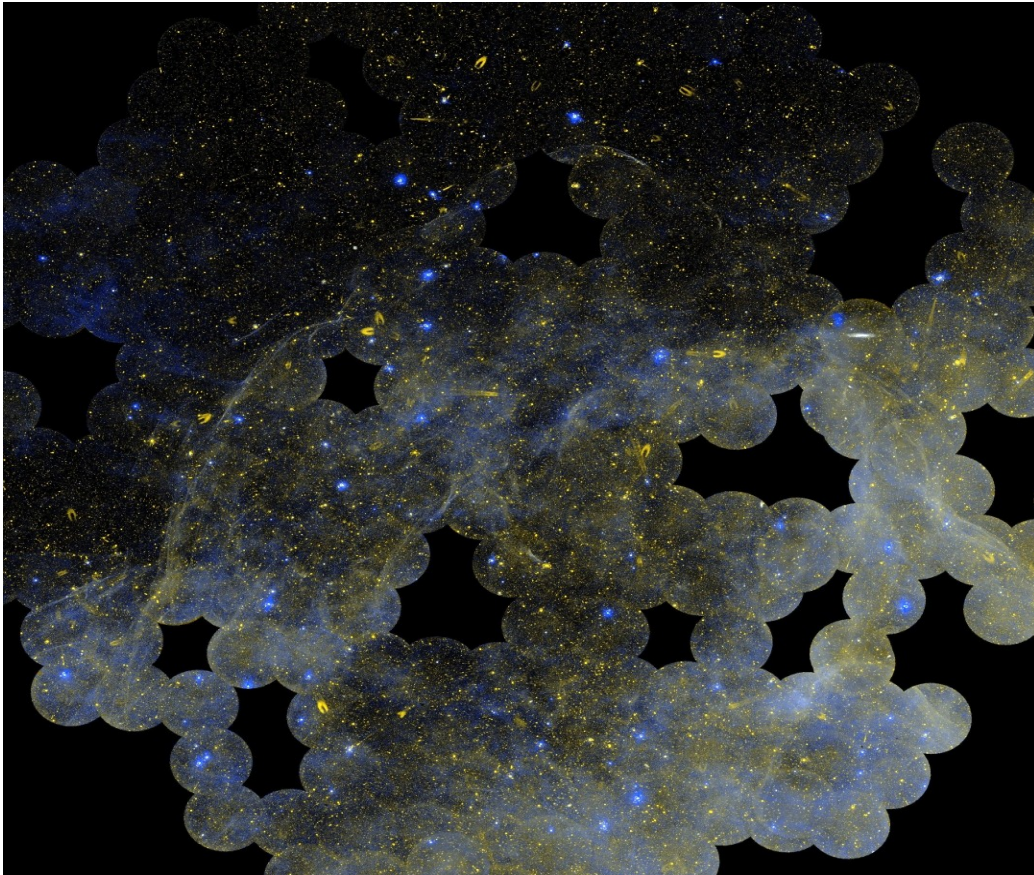
1. Ackerman, M.; Ajello, M.; Albert, A., et al., 2013, ApJ, submitted
2. Ajello, M.; Albert, A.; Allafort, A., et al., 2013, ApJ, submitted
3. Aulanier, G.; Janvier, M.; Schmieder, B., 2012, A&A, 543, A110
4. Barnes, G., & Leka, K.D., 2008, ApJ, 688, L107
5. Benz, A.; Krucker, S.; Hurford, G. J., et al., 2012, SPIE, 8443, article id. 84433L
6. Birn, J., Fletcher, L., Hesse, M. & Neukirch, T., 2009, ApJ, 695, 1151
7. Brown, J.C., 1971, Sol. Phys., 18, 489
8. Chupp, E.L., & Ryan, J., 2009, Res. A&A, 9, 11
9. Donea, A.-C.; Lindsey, C., 2005, ApJ, 630, 1168
10. Dresing, N., Gómez-Herrero, R., Klassen, A., et al., 2012, Sol. Phys., 281, 281
11. Dunphy, P.P., Chupp, E.L., Bertsch, D.L., et al., 1999, Sol. Phys., 187, 45
12. Emslie, A.G., Dennis, B.R., Holman, G.D., & Hudson, H.S., 2005, Geophys. Res. Lett., 110, A11103
13. Fleishman, G.D. & Kontar, E.P., 2010, ApJ, 709, L127
14. Fletcher, L. & Hudson, H.S., 2008, ApJ, 675, 1645
15. Foullon, C., Verwichte, E., Nakariakov, V.M., & Fletcher, L., 2005, A&A, 440, L59
16. Gonzalez, W.D., Echer, E.; Tsurutani, B.T., et al., 2011, Space Sci. Rev., 158, 69
17. Grechnev, V.V., Kurt, V.G., Chertok, I.M., et al., 2008, Sol. Phys., 252, 149
18. Green, L.M.; Kliem, B.; Wallace, A. J., 2011, A&A, 526, A2
19. Hudson, H. S.; Fisher, G. H.; Welsch, B. T., 2008, ASPC, 383, 221
20. Hurford, G.J., Schwartz, R.A., Krucker, S., et al., 2003, ApJ, 595, L77
21. Hurford, G. J.; Krucker, S.; Lin, R. P.; Schwartz, R. A.; Share, G. H.; Smith, D. M., 2006, ApJ, 644, L93
22. Jackman, C.; Fleming, E.L.; Vitt, F.M., 2000, JGR, 105, 11659
23. Kahler, S.W., 2001, JGR, 106, 20947
24. Kanbach, G., Bertsch, D.L., Fichtel, C.E., et al., 1993, A&AS, 97, 349
25. Kaufmann, P., Raulin, J., de Castro, C.G.G., et al., 2004, ApJ, 603, L121
26. Kaufmann, P.; Trotter, G.; Gimenez de Castro, C.G.; et al., 2009, Solar Phys., 255, 131
27. Kiener, J., Gros, M., Tatischeff, V., & Weidenspointner, G., 2006, A&A, 445, 725
28. Kliem, B. & Török, T., 2006, Physical Review Letters, 96, 255002
Kocharov, L.G., Debrunner, H., Kovaltsov, G., et al. 1998, A&A, 340, 257
29. Kontar, E. P., MacKinnon, A.L., Schwartz, R.A., & Brown, J.C., 2006, A&A, 446, 1157
30. Kontar, E., Brown, J.C. Emslie, A.G., et al., 2011, Space Sci. Rev., 159, 301
31. Kontar, E.P., & Brown, J.C., 2006, ASR, 38, 945
32. Kosovichev, A. G. & Zharkova, V. V., 1998, Nature, 393, 317
33. Kostiuk, N.D. & Pilkener, S.B., 1975, Soviet Astronomy, 18, 590
34. Krucker, S., Christe, S., Glesener, L., et al., 2011, SPIE, 8147, 4
35. Krucker, S.; Gimenez de Castro, C. G.; Hudson, H. S., et al., 2013, A&ARv, 21, 58
36. Laming, J.M., Moses, J. D., Ko, Y.-K., et al., 2013, ApJ, in press
37. Larosa, T.N., Moore, R.L., & Shore, S.N., 1994, ApJ, 425, 865
38. Lin, R.P. et al. 2002, Solar Phys., 210, 3
39. Lin, R.P., Krucker, S., Caspi, A., et al., 2011, SPD, 42, 2204

40. Lingenfelter, R.E., & Ramaty, R., 1967, *Plan. Spa. Sci.*, 15, 1303
41. Liu, C., & Wang, H., 2009, *ApJ*, 696, L27
42. Liu, C., Lee, J., Jing, J., et al., 2008, *ApJ*, 672, 69
43. Mackinnon, A.L., 1988, *A&A*, 194, 279
44. Martinez-Oliveros, J. C.; Donea, A.-C.; Cally, P. S.; Moradi, H., 2008, *MNRAS*, 389, 1905
45. Masson, S.; Antiochos, S.K.; DeVore, C.R., 2013, eprint arXiv:1301.0654
46. Masuda, S., Kosugi, T., & Hudson, H.S., 2001, *Sol. Phys.*, 204, 55
47. Matthews, S.A., Williams, D.R., Klein, K.-L., et al., 2012, *ExA*, 33, 237
48. Mazin, B.A.; Sank, D.; McHugh, S., et al., 2010, *APL*, 96, 102504
49. Metcalf, T.R.; De Rosa, M.L.; Schrijver, C.J., et al., 2008, *Sol. Phys.*, 247, 269
50. Mewaldt, R.A.; Looper, M. D.; Cohen, C. M. S., et al., 2012, *SSRv*, 171, 97
51. Murphy, R., Share, G.H., Grove, J.E., et al., 1997, *ApJ*, 490, 883
52. Nakariakov, V.M., Foullon, C., Verwichte, E., & Young, N.P., 2006, *A&A*, 452, 343
53. Nakariakov, V.M., Foullon, C., Myagkova, I.N., & Inglis, A.R., 2010, *ApJ*, 708, L47
54. Piana, M.; Massone, A.M.; Kontar, E.P.; et al, 2003, *ApJ*, 595, L127
55. Pick, M., Mason, G. M., Wang, Y.-M., et al., 2006, *ApJ*, 648, 1247
56. Rouillard, A.P.; Sheeley, N. R.; Tylka, A., et al., 2012, *ApJ*, 752, 44
57. Russell, A.J.B. & Fletcher, L., 2013, *ApJ*, 765, 81
58. Ryan, J.M, 2000, *Space Sci. Rev.*, 93, 581
59. Schrijver, C.J., 2009, *ASR*, 43, 739
60. Share, G.H., & Murphy, R.J., 2006, in *Solar Eruptions and Energetic Particles*, eds. N. Gopalswamy, R. Mewaldt, J. Torsti, *Geophys. Monogr.*, 165, 177
61. Shih, A.Y., Lin, R.P., & Smith, D.M., 2009, *ApJ*, 698, L152
62. Silva, A.V.R., Share, G.H., Murphy, R.J., et al., 2007, *Sol. Phys.*, 245, 311
63. Solanki, S. et al 2010, *ApJ*, 723, L127
64. Su, Y., Dennis, B.R., Holman, G.D., et al., 2012, *ApJL*, 746, L5
65. Török, T & Kliem, B., 2005, *ApJ*, 630, L97
66. Trottet, G., 2006, in *Third French- Chinese Meeting on Solar Physics*, eds. Fang, C., Schmieder, B., Ding, M.D., Nanjing Univ .Press , 82.
67. Trottet, G., Krucker, S., Luthi, T., & Magun, A., 2008, *ApJ*, 678 , 509
68. van Ballegoijen, A.A. & Martens, P.C.H., 1989, *ApJ*, 343, 971
69. Vilmer, N., MacKinnon, A.L., Trottet, G., & Barat, C., 2003, *A&A*, 412, 865
70. Vilmer, N., MacKinnon, A.L., & Hurford, G.J., 2011, *Space Sci. Rev.*, 159, 167
71. Woods, T.N., Hock, R., Eparvier, F., et al., 2011, *ApJ*, 739, 59
72. Zhang, J.; Dere, K. P.; Howard, R. A., et al., 2001, *ApJ*, 559, 452
73. Zharkov, S.; Green, L. M.; Matthews, S. A.; Zharkova, V. V., 2011, *ApJL*, 741, L35
74. Zharkov, S.; Green, L. M.; Matthews, S. A.; Zharkova, V. V., 2013, *Sol. Phys.*, 284, 315
75. Zharkova, V.V. & Gordovskyy, M., 2006, *ApJ*, 651, 553
76. Zharkova, V.V., Kuznetsov, A. A., & Siversky, T.V., 2010, *A&A*, 512
77. Zharkova, V.V. et al., 2011, *Space Sci. Rev.*, 159, 347



EUROPEAN ULTRAVIOLET-VISIBLE OBSERVATORY

"Building galaxies, stars, planets and the ingredients for life among the stars"



The UV sky from GALEX All Sky Survey

Spokesperson: Ana Inés Gómez de Castro
Contact details: AEGORA-Facultad de CC Matemáticas
Universidad Complutense de Madrid
Plaza de Ciencias 3, 28040 Madrid, Spain

email: aig@ucm.es
Phone: +34 91 3944058
Mobile: +34 659783338

LIST OF AUTHORS AND CONTRIBUTORS

Thierry Appourchaux – IAS France
Martin Barstow – University of Leicester, United Kingdom
Mathieu Barthelemy - IPAG, France
Frédéric Baudin – IAS, France
Stefano Benetti – OAPD- INAF, Italy
Pere Blay - Universidad de Valencia, Spain
Noah Brosch - Tel Aviv University, Israel
Enma Bunce - University of Leicester, United Kingdom
Domitilla de Martino – OAC-INAF, Italy
Ana Ines Gomez de Castro – Universidad Complutense de Madrid, Spain
Jean-Michel Deharveng – Observatoire Astronomique Marseille-Provence, France
Roger Ferlet - Institute d'Astrophysique de Paris, France
Miriam García – IAC, Spain
Boris Gaensicke -University of Warwick, United Kingdom
Cecile Gry - Observatoire Astronomique Marseille-Provence, France
Lynne Hillenbrand – Caltech, USA
Eric Josselin - University of Montpellier, France
Carolina Kehrig – Instituto de Astrofísica de Andalucía, Spain
Laurent Lamy - LESIA, France
Jon Lapington – University of Leicester, United Kingdom
Alain Lecavelier des Etangs – Institute d'Astrophysique de Paris, France
Frank LePetit – Observatoire Paris-Meudon, France
Javier Lopez Santiago - Universidad Complutense de Madrid, Spain
Bruno Milliard- Observatoire Astronomique Marseille-Provence, France
Richard Monier - Université de Nice, France
Giampiero Nalletto – University of Padova, Italy
Yael Nazé - Liège University, Belgium
Coralie Neiner - LESIA, France
Jonathan Nichols – University of Leicester, United Kingdom
Marina Orio – OAPD-INAF, Italy
Isabella Pagano – OACT-INAF, Italy
Céline Peroux – Observatoire Astronomique Marseille-Provence, France
Gregor Rauw – University of Liège, Belgium
Steven Shore – University of Pisa, Italy
Marco Spaans - Kaptein Astronomical Institute, The Netherlands
Gagik Tovmassian - Instit. Astr. Sede Ensenada-UNAM, Mexico
José Vilchez – Instituto de Astrofísica de Andalucía, Spain
Kevin France - University of Colorado, USA
Asif ud-Doula - Penn State University, USA

LIST OF SUPPORTERS (see www.nuva.eu/whitepaper/supporters.php)

Executive summary:

The growth of luminous structures and the building blocks of life in the Universe began as primordial gas was processed in stars and mixed at galactic scales. The mechanisms responsible for this development are not well-understood and have changed over the intervening 13 billion years. To follow the evolution of matter over cosmic time, it is necessary to study the strongest (resonance) transitions of the most abundant species in the Universe. Most of them are in the ultraviolet (UV; 950 Å – 3000 Å) spectral range that is unobservable from the ground. A versatile space observatory with UV sensitivity a factor of 50-100 greater than existing facilities will revolutionize our understanding of the Universe.

Habitable planets grow in protostellar discs under ultraviolet irradiation, a by-product of the star-disk interaction that drives the physical and chemical evolution of discs and young planetary systems. The electronic transitions of the most abundant molecules are pumped by this UV field, providing unique diagnostics of the planet-forming environment that cannot be accessed from the ground. Earth's atmosphere is in constant interaction with the interplanetary medium and the solar UV radiation field. A 50-100 times improvement in sensitivity would enable the observation of the key atmospheric ingredients of Earth-like exoplanets (carbon, oxygen, ozone), provide crucial input for models of biologically active worlds outside the solar system, and provide the phenomenological baseline to understand the Earth atmosphere in context.

In this white paper, we outline the key science that such a facility would make possible and outline the instrumentation to be implemented.

1. INTRODUCTION

Any future Cosmic Vision requires probing the conditions for the emergence of life in the Universe. In this white paper we present a path to own this future.

For organic molecules to exist nucleosynthesis needs to have proceeded. Studies of the metal abundance variation up to redshift $z=5$ demonstrate that the metallicity increases steadily with the age of the Universe. However, the metal enrichment of the Universe was clearly neither uniform nor homogeneous; metal-poor clouds have been detected and chemically processed material has been found in the voids of the Cosmic Web. The star formation rate is observed to decrease from $z=1$ to the present. Important clues on the metal enrichment spreading on the Universe depend on inter-galactic transport processes; these are poorly studied because of the lack of high sensitivity spectral-imaging capabilities for detecting the warm/hot plasma emission from galactic halos. Current information derives from ultraviolet (UV) absorption-line spectroscopy of the presence of strong background sources. Most of the intergalactic emission is expected to come from circumgalactic filaments and chimneys that radiate strongly in the UV range. To study these structures a high sensitivity spectral-imaging capability is required with spatial resolution at least ten times better than those provided by the GALEX mission.

Metallicity is relevant for life generation as we know it not only at the DNA level but also at much earlier phases. Silicates and carbonates are the key building blocks of dust grains and planetesimals in protostellar/protoplanetary discs. The far UV radiation is a major contributor to disc evolution. It drives the photo-evaporation of the gas disc setting the final architecture of the giant planets in the system and beginning the epoch of rocky planet formation. Unfortunately, little is known about the FUV radiation from solar-system precursors. The measurements carried out from X-ray to softer UV bands point out that the FUV flux varied significantly during the pre-main sequence (PMS) evolution. Protostellar discs are shielded from the energetic stellar radiation during the early phases ($<1\text{Myr}$), but as they evolve into young planetary discs, the FUV and extreme UV (EUV) radiation from the very active young Suns, irradiates them heavily. Strong stellar winds are expected to interact with the left over particles and produce diffuse Helium and Hydrogen emission that pervades the entire young systems during the planets early evolution and planetary atmosphere formation. Around the Sun, within a modest radius of 500 pc, there are thousands of young solar-like stars of all masses and in all the phases of the PMS evolution. The observation of these sources with high sensitivity mid resolution spectroscopy would provide a unique perspective on atmospheres, magnetospheres and coronal evolution, as well as on their impact on planetary formation and evolution.

Solar system planetary research is fundamental for understanding atmospheres as global systems, including the Earth. For example, the link between upper and lower atmospheres is poorly known for the Earth case despite the fact it can have implications on global warming. Studying the upper atmosphere of not only Solar System planets but also of exoplanets can help the understanding of the mechanism operating at Earth¹ [83].

The stellar or solar FUV-EUV fluxes are the main energy input at high atmospheric altitudes. Many atmospheric atoms, ions and molecules have strong electronic transitions in the UV-visible domain. This wavelength range gives access to fundamental constituents of the atmospheres. In particular, bio-markers like Ozone (O₃) and molecular oxygen (O₂) have very strong absorption transitions in the ultraviolet protecting complex molecules especially DNA from dissociation or ionization. Absorption of O₃ through the Hartley bands occurs between 2000-3000 Å and O₂ has strong absorptions in the range 1500-2000 Å. Atomic oxygen presents a resonance multiplet at 1304 Å and the famous auroral green and red lines in the visible. For other planetary cases or paleo-Earth, CO has strong absorption bands below 1800 Å and forbidden emission bands from 2000 Å to 3000 Å. Lyman-α (Lyα) is the strongest emission line from Earth's atmosphere and is an invaluable tracer for studying the giant planets and hot Jupiters. Hydrocarbons, e.g. methane, are strong FUV absorbers and are sensitive links to the radiative balance in a planetary atmosphere. Observations in the UV and visible wavelength are therefore powerful diagnostics of the structural, thermal, and dynamical properties of planets, be they Solar System or extrasolar.

The UV is an essential spectral interval for all fields in astrophysical research; imaging and spectral coverage at UV wavelengths provides access to diagnostic indicators for diffuse plasmas in space, from planetary atmospheres to elusive gas in the intergalactic medium (IGM). Linking visible and UV spectral features covers the widest possible range of species and vast range of temperatures that cover most astrophysical processes. Moreover, UV observations are essential for studying processes outside of strict thermal equilibrium that produce conditions favorable to complex chemistry, the production pathway for large molecules that absorb and shield planetary surfaces from the harsh space conditions. But UV radiation itself is also a powerful astrochemical and photoionizing agent. Moreover, UV-visible instrumentation provides the best possible spatial resolution for normal incidence optics, since resolution is inversely proportional to the radiation wavelength.

The European community will continue to profit from the refurbished Hubble Space Telescope (HST), a 23 years old space telescope, until the mission end. Following that, the only major telescope world-wide that will provide access to the UV range is the Russian-led WSO-UV (170 cm) space telescope, which includes only minor European participation from Spain. Among the astronomical research lines given the maximum priority in Europe, three require access to the UV range: (1) Planets and Life, (2) The Solar System and (3) The Universe. This even extends to the investigation of the fundamental physical laws, e.g the UV-visible range is without rival for measurement of the variation of the fine structure constant during most of the age of the universe. A versatile UV-visible facility that improves the sensitivity of the dying ultraviolet (UV) facilities by a factor of 50-100 will produce a revolution in our understanding of the Universe. In this White Paper, we outline the key science to be enabled by expanding the astronomical community's access to UV-visible observations in the coming decades.

2. SCIENCE PROGRAM FOR A LARGE ULTRAVIOLET OBSERVATORY

2.1 TRANSPORT PROCESSES IN THE INTERGALACTIC MEDIUM OVER 80% OF THE UNIVERSE LIFETIME

Our Universe is filled by a cosmic web of dark matter (DM) in the form of filaments and sheets. The recent results from the Planck mission [1] show that the DM is one of the two major components of the Universe, forming about 26.8% of its content. The baryonic matter is less than 5%, with the visible baryons being only one-quarter of this amount, ~1% of the total content of our Universe. One important question deriving from this is, therefore, where are most of the baryons.

¹Crip, 2013 – www.lpi.usra.edu/vexag/meetings/ComparativeClimatology/presentations/

The DM large-scale mesh was created by primordial density fluctuations and is detected in the present day galaxy distribution, the large-scale structure (LSS). The pattern of density fluctuations was imprinted and can be observed in the cosmic microwave background. At least some baryonic matter follows the DM web components and, in places where the gravitational potential is strong, this baryonic matter collapses to form visible stars and galaxies. Matter not yet incorporated in galaxies is "intergalactic matter" (IGM), with more than 50% of the baryons predicted to be contained at low redshift in a warm-hot phase (WHIM), shock-heated at about 10^5 K to 10^6 K by the energy released by structure and galaxy formation (e.g. [18,19,24,25]). Evidence for the WHIM from spectroscopic surveys of the low- z IGM has been a major success of FUSE, HST/STIS and HST/COS (e.g. [22,97,93,23]) but the baryon census remains uncertain with 30% still missing ([88,94]).

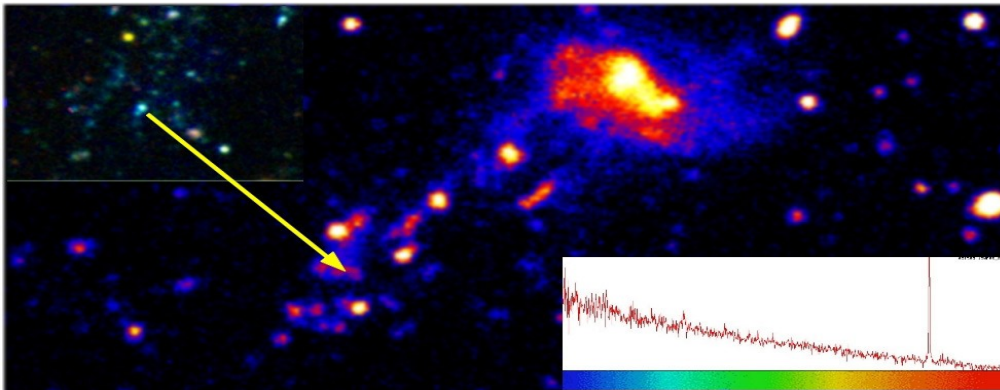
UV absorption line surveys conducted with much larger effective collecting area than HST hold the promise to probe higher temperature ranges with tracers such as NeVIII and low-metallicity volumes with broad Lyman- α absorbers (e.g. [82,68]). This would fully validate the theories of structure formation that predict the development of a warm-hot phase of the IGM at low redshift.

However, while the general characteristics of the process seem clear, not all the steps are fully understood. One troublesome question has to do with the influence of the environment on galaxies; this has been dubbed the "nature vs. nurture" question in the context of galaxy formation and evolution. It is clear, however, that the IGM is responsible for the formation of the visible galaxies and for their fueling during that time. The IGM itself is modified by galaxy ejecta in the form of galactic winds and by stripping of galaxies of their gas when they pass through a cluster of galaxies.

Starburst galaxies (SBs) are known to drive powerful superwinds. The overpressured $10^{(7-8)}$ K metal enriched plasma that is created by stellar winds and core-collapse SNe sweeps up cooler material and blows it out (or even away). Speeds are of the order of 200-1000 km/s, or even 3000 km/s. SBs can contribute significantly to the mass budget, energetics and metallicity of the IGM. Apart from the metal enrichment [89], superwinds may also create large, ~ 100 kpc, holes in the IGM [2].

In order to assess the mechanical and chemical feedback from SBs on the IGM, it is necessary to assess the composition and kinematics of the IGM gas, particularly through elements such as O, Ne, Mg, Si, S, Fe; for $T \leq 10^7$ K. In fact, the dominant IGM component is near $\sim 10^5$ K (WHIM) and its mass has been growing with cosmic time to about half the total amount of all the baryons in the local universe.

UV spectroscopic surveys can detect this part of the IGM through Ly α , resonance lines of H, C, O as well as high ionization states of heavier elements, over the last 5-10 Gyr. Particularly CIV, SiIV and OVI are useful probes in this.



UV and visible images of IC 3418, a galaxy being stripped in the Virgo Cluster. The arrow points to a supergiant star formed in the stripped material (fig 1 from [76]).

With the launch of JWST expected to take place within this decade, and with the full ALMA capability to be available even sooner, the study of the very early Universe will receive a very serious experimental boost. It should be noted however that, while the LSS started to emerge at redshifts higher than 3, its evolution continues to the present and what we see now is mainly the product of evolution at $z \ll 3$. Within the broader cosmological context some unanswered questions are when the first luminous

objects formed, what was their nature, and how did they interact with the rest of the Universe. This is also where the questions about re-ionization and metal enrichment arise.

The IGM is enriched with metals by stellar and galactic winds, by ejecta from supernovae and superbubbles, and by winds from active galactic nuclei. The interaction of strong winds with the IGM may partly quench the IGM accretion. Therefore, we witness a feedback process that modifies the IGM dynamics, physical properties and metallicity. However, it is not clear how are the metals distributed in the IGM following their ejection from a galaxy.

Understanding of this process requires studying the sites where metals are produced, which are in massive stars that explode as SNe and in their surrounding nebulosities. Important information will come from detailed observations of low-metallicity local starbursts [e.g. Izw18, one of the most metal-poor galaxies known to date (e.g. [100]) that will enable the investigation of the early evolution of star-forming galaxies, and the origin of high-ionization nebular lines like H δ . Wolf-Rayet (WR) stars are thought to be primarily responsible for the H δ emission. Nevertheless, based on non-rotating evolution models for single stars, a very small number of WR stars are expected to be present in metal-poor systems (see [60] and references therein). In this context, some questions that might be asked are whether the high-excitation lines are really due to WR stars. If so, how can WR stars form in such a low-Z environment? Are the optical spectra adequate to search for WR galaxies? Do some optically-classified WR galaxies host Of-type stars instead of WR stars, and might be misclassified?

The transport of metals into the IGM must be understood and out-flows and in-flows of metals from and into different types of galaxies must be discovered. Metallicity gradients in galaxies, between the disc and the halo, must also be understood. In this context, we mention the understanding of the extended gas distributions of galaxies that can be achieved by studying absorption lines in the lines of sight to distant objects passing through galaxy haloes.

Interactions in dense groups of galaxies, where the dark matter halos have not yet merged into a single entity and the total mass is much less than that of even a poor GC, are best represented by the Hickson compact groups (HCG). These are collections of fewer than ten galaxies in very close proximity to each other. While the original Hickson collection was based on projected surface density, [52] revised the group association via accurate redshift determinations showing that most groups are real. Therefore, the HCGs represent, environments where galaxies are within a few radii and the light crossing times are very short.

At the other extreme one finds galaxies even near the centers of cosmic voids. Such galaxies have had no interactions for (almost) one Hubble time and thus are the best examples of evolution in isolation. A comparison of the two kinds of objects should, therefore, illuminate the case of "nature vs. nurture" in the evolution of galaxies. In particular, one should expect that isolated galaxies in voids should show almost or no star formation, and that their morphological appearance should be very regular.

In general, galaxies in voids tend to be smaller and to exhibit weaker star formation than similar objects in the general field. There is evidence for galaxy formation in voids taking place along DM filaments, since the galaxies appear aligned (e.g., [108,12]). The latter paper presents three galaxies connected within a quasi-linear HI structure, where one of the galaxies VGS_31b exhibits a tail and a ring.

At least one isolated polar ring galaxy (RG) has been identified in a cosmic void [91]. Since the consensus regarding RGs is that these result from a major interaction whereby one galaxy acquires significant amounts of HI from either a nearby companion or directly from the intergalactic space, the conclusion should be that this baryonic mass transfer takes place even in the regions with the lowest galaxy density. Thus neutral IGM can exist in cosmic voids and, when the conditions are suitable, this IGM can convert into luminous stars and galaxies.

The case of RGs is particularly interesting since in many cases these objects appear to have an elliptical or lenticular galaxy surrounded by, or containing, a gaseous and star-forming disc or ring. Since RGs form by accretion of extra-galactic matter, which does not have to follow the kinematic properties of the central object, one finds this type of galaxies to be a good test case for matter orbiting at different

orientations within a dark matter halo. Thus the study of RGs offers a way to understand the shape of the DM concentrations.

The GALEX mission has been the only UV wide-field imager yielding significant discoveries in the domain of galaxy evolution. Its images uncovered the extended UV (XUV) discs seen around many spiral galaxies; similar features were detected also around other morphological types of galaxies. The XUV discs are essentially star-forming features. Their location around some objects that had been classified as "red and dead" ellipticals points toward the acquisition of fresh gas either through galaxy-galaxy interactions or directly from the intergalactic space. XUV discs and the RGs may be related phenomena. In order to better characterize them a much more capable instrument is needed than GALEX.

Other intriguing objects newly discovered are "green pea" galaxies (e.g. [17]). Such objects may be similar to but of a higher luminosity than the star-forming very compact "knots" identified near star-forming dwarf galaxies [15] or the "H α dots" [59]

With the existing telescopes and the giant ones planned for the next decade (TMT, ELT, etc.) it will still not be possible to directly measure the light from individual main-sequence stars in nearby galaxies, since this requires angular resolutions not achievable by ground-based instruments. It will also be possible to analyze in depth blue supergiant stars formed in the stripped material, such as the one recently found [77]

The requirements for an instrument able to address the issues mentioned above are (i) a large collecting area, (ii) a wide field of view (FOV), (iii) a high spatial resolution, and (iv) the ability to perform medium spectral resolution of point-like objects as well as offering (v) integral field spectroscopy. The large collecting area is necessary to enable the observation of faint sources, since galaxies and individual stars in other galaxies are faint. To probe the extended gaseous haloes via AGN absorptions requires good spectral sensitivity, since the projected spatial density of the AGNs to be used as background sources is sufficient only at faint AGN magnitudes (one QSO per five arcmin at 21 mag.; [50]). The wide FOV is required to allow the sampling of significant numbers of galaxies in e.g. distant galaxy clusters. The high spatial resolution, of about 0".01 or better, is necessary to resolve individual stars in other galaxies and in the cores of globular clusters, as well as to reveal details of structures within galaxies such as HII regions, star clusters, etc. Field spectroscopy is necessary to quantify the stellar populations of resolved galaxies and their immediate neighborhoods.

2.2 PLANET FORMATION AND THE EMERGENCE OF LIFE

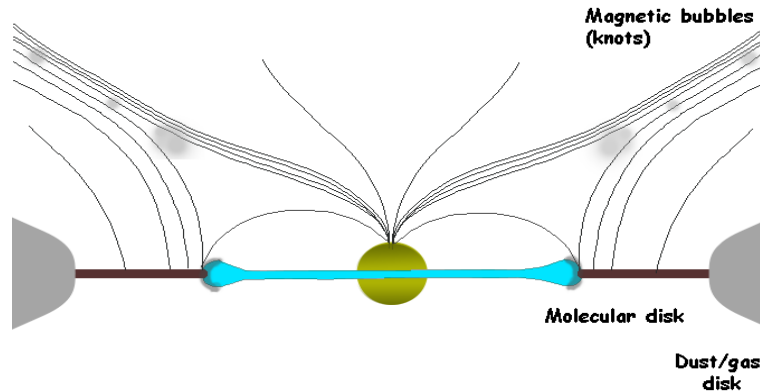
The formation and evolution of exoplanetary systems is essentially the story of the circumstellar gas and dust, initially present in the protostellar environment, and how these are governed by gravity, magnetic fields and the hard radiation from the star. Understanding formation of exoplanets, including the terrestrial ones, and of their atmospheres, calls for a deep study of the life cycle of protostellar discs from their initial conditions to the young planetary discs. UV radiation plays an essential role in this cycle, from disc ionization to chemical processing. It is also the main observational probe of the accretion mechanisms that regulate disc evolution. It also reveals the stellar dynamo that is at the seat of the global magnetic field that plays a significant role in the evolution of exoplanets' atmosphere.

2.2.1 Star-disc interaction: the accretion engine

Accretion onto magnetized structures drives hydro-magnetic processes that power the large scale optical jets and molecular outflows. These engines control the efficiency of accretion and therefore set the final stellar mass by regulating the amount of matter that falls onto the star. During this period, the disc and star form a strongly coupled system. This coupling is eventually broken and the disc dissipates. But the star's magnetic activity continues to influence the evolution of the atmospheres of any nascent planets.

The basic performance of the engine relates with the interaction between the stellar magnetic field and the ionized gas orbiting in the inner border of the disc that generates a toroidal magnetic field. The magnetic pressure from the toroidal field pushes the field lines outwards from the disc rotation axis, inflating and opening them in a butterfly-like pattern and producing a current layer between the star- and disc-dominated regions. Magnetic field dissipation in the current layer also produces high energy

radiation and particles above the disc plane. A fraction of the energy released is lost from the system through X rays, EUV, and FUV radiation that heats and ionizes the disc (especially the inner few AU).



Sketch of the accretion engine. The star acts as a magnetic rotor that interacts with the plasma orbiting around it in Keplerian orbits. The magnetic configuration is outlined from [101] as well as the layer where magnetic reconnection occurs and magnetic bubbles are generated.

The magnetic link between the star and the disc is broken and reconnected, continually ejecting blobs of plasma: the knots in the optical jets ([49,101]). The opening angle of the current layer, and its extent, depends on the stellar and disc fields, the accretion rate and the ratio of the inner disc radius and stellar rotation frequencies. The jet radiates in UV intercombination SiIII],CIII] [43,44] at 1892\AA and 1908\AA that are optically thin and can be used to probe the warm base and to study the jet collimation mechanisms. Moreover, UV instrumentation provides the best possible spatial resolution for studying any jet rotation at the base [21]. The first detection of the hot solar-like winds in T Tauri stars (TTs) has been achieved using their NV emission [46,48].

The star-disc interaction also distorts and powers the stellar magnetosphere, leading to the formation of structures extending several stellar radii [84,85] that dominate the UV radiation from the TTs. UV radiation during the T Tauri phase is typically 50 times stronger than during the main sequence evolution of solar-mass stars [57,107,47,4]. Both line strength and broadening decreases as the stars approach the main sequence [4,48] confirming that the UV excess is dominated by the extended magnetospheric emission. The correlations between the UV magnetospheric tracers seem to extend from TTs to brown dwarfs [47] providing a unique tool to study the physics and evolution of the engine for a broad range of masses. Unfortunately, the UV spectrum of just one brown dwarf [35] has been obtained because of the lack of sensitivity.

Many uncertainties remain about how star-disc interaction self-regulates and also regarding the dependence of the engine details on initial conditions e.g. the effective gravity of the star, the role of stellar radiation and magnetic field on the engine performance and the role of the ionizing radiation produced by the engine on the evolution of the disc. Despite the modeling advance so far, the real properties of the engine itself are poorly known given the lack of observations to constrain the modeling. Very important questions still open are: *How does the accretion flow proceed from the disc to the star? Is there any preferred accretion geometry as, for instance, funnel flows? What is the temperature distribution emerging from the accretion shock? What role do disc instabilities play in the whole accretion/outflow process? What are the dominant processes involved in wind acceleration? What are the relevant time-scales for mass ejection? How does this high energy environment affect the chemical properties of the disc and planetary building? Whether and how this mechanism works when radiation pressure becomes significant as for Herbig Ae/Be stars? How does the stellar magnetosphere evolve?*

A single spectrum in the UV range contains information about all the physical components - atmosphere, magnetosphere, outflows (Solar-like winds, jets), accretion flow, inner disc structure, residual gas in the young planetary system - and their evolution into exoplanetary systems. At least, a factor of 10 improvement on sensitivity is required (reaching $10^{-17} \text{ erg/s/cm}^2/\text{\AA}$ with $R \approx 20,000$) to observe the faintest components of the engine in a sample large enough to answer the above questions. This represents a factor of ten improvement over current facilities (HST).

2.2.2 The evolution of the gaseous component in the disc

The lifetime, spatial distribution, and composition of gas and dust of young (age < 30 Myr) circumstellar discs are important properties for understanding the formation and evolution of extrasolar planetary systems. Disc gas regulates planetary migration [102,5,96] and the migration timescale is sensitive to the specifics of the disc surface density distribution and dissipation timescale [6]. Moreover, the formation of giant planet cores and their accretion of gaseous envelopes occurs on timescales similar to the lifetimes of the discs around T Tauri and Herbig Ae/Be stars ($10^6 - 10^7$ yr).

UV radiation from the engine is very efficient at etching the disc surface through photoevaporative flows [53,3] and determines the lifetime of the gaseous component of discs. The dust disc clearing timescale is expected to be 2-4 Myr [51], however recent results indicate that inner molecular discs can persist to ages ~ 10 Myr in Classical TTSs [86,56,37]. Fluorescent H_2 spectra in the 912 – 1650 Å bandpass are sensitive to probe gas column densities $< 10^{-6}$ g cm $^{-2}$, making them the most sensitive tracer of tenuous gas in the protoplanetary environment. While mid-IR CO spectra or other traditional accretion diagnostics suggest that the inner gas disc has dissipated, far-UV H_2 and CO observations offer unambiguous evidence for the presence of an inner ($r < 10$ AU) molecular disc [36,38]. The penetration of UV photons from the engine into the dusty disc and the spectral distribution of the FUV radiation field have a strong influence on the chemical abundances of the disc [11]. FUV radiation photo dissociating H_2 and CO controls the chemistry of the external layers of the disc.

Spectroscopic observations of volatiles released by dust, planetesimals and comets provide an extremely powerful tool for determining the relative abundances of the vaporizing species and studying the photochemical and physical processes acting in the inner parts of the planetary discs [98,65,66].

2.2.3 Influence of UV emission on planets

The "habitability" of planets depends not only on the thermal effect of the central star but also on its magnetic activity [79], which strongly influences the chemical (and possibly biochemical) processes at the surface of the planet. Stellar UV and EUV emission is an important contributor to planetary atmosphere evaporation by photo-dissociation and is also responsible for damage to the biochemical structures necessary for a biological activity [87]. UV and EUV fluxes originate in hot plasma in the stellar atmosphere. The required temperature (from some tens of thousands to million Kelvin) is reached by the non-thermal heating of the atmosphere caused by the action of the stellar dynamo (see Sect. 2.4).

Solar-like stars show high variability in this energy band [76, 38, 40, 78]. Contrary to the corona, where the plasma is optically thin, the transition region and the chromosphere are optically thick releasing high UV fluxes. During their lives, stars lose angular momentum through torquing by magnetized stellar winds. As a result, high energy emission from solar-like stars decreases with age. As their magnetic activity intensity is connected with their rotation rate [90,80], UV emission also follows this trend [71,73]. As a consequence, planetary atmospheres receive higher UV radiation during the first stages of the planet's life. UV and EUV radiation strongly influence planetary atmospheres (e.g., [64]). A spectacular case has been revealed through planetary transit studies: the star induces massive (and irregular) planetary material ejection because of the close planet orbit (0.01 AU), the planet shows a comet-like tail, responsible for the observed variations of transit depth [81]. This example, together with the influence of magnetic activity on planetary habitability, shows how necessary it is to understand stellar dynamos to describe the orbiting planets; it is in fact insufficient to extrapolate EUV and UV fluxes from X-rays, being demonstrated that correlation between emission in the two spectral regions is not universal, but strongly dependent on the activity level of the star. For example, [92] show no correlation between UV and X-ray flux for slowly-rotating M dwarfs. Moreover, planet detection and characterization rely directly on knowledge of its stellar host (see Sect. 2.3).

2.2.4 Characterization of exoplanet atmospheres

Understanding the physical processes in exoplanetary atmospheres requires studying a large sample of systems. Many planets have been discovered through transits, for example by the Corot and Kepler missions and will be discovered with GAIA or CHEOPS. The most recent estimate [28] is that GAIA will discover some hundred transiting Hot Jupiters ($P < 5$ d) to $G < 14$, and a few thousand to $G < 16$. These will be prime targets for detecting atmospheric constituents through absorption spectroscopy, thereby characterizing the chemical and physical properties of the atmosphere. The first detections and

observation of the atmospheres of transiting extrasolar planets (HD209458b and HD189733b) have been made through UV-visible spectroscopy [99,69,67].

The observation of UV and visible absorption when a planet transits its parent star is a very powerful diagnostic technique. In fact, this method is the most effective way to detect Earth-like planets because of the strong absorption of stellar UV photons by the ozone molecule in the planetary atmosphere (see [45]). Observation of biomarkers in the high atmosphere of Earth shows that these atoms and ions are present at very high altitude (even several hundreds of kilometers) causing large absorption depths. *Electronic molecular transitions, pumped by UV photons, are several orders of magnitude stronger than the vibrational or rotational transitions observed in the infrared or radio range* provided the stellar UV radiation field is strong enough. For a typical life-supporting terrestrial planet, the ozone layer is optically thick to grazing incident UV radiation to an altitude of about 60 kilometers. With a telescope 50 times as sensitive as HST/STIS, ozone can be detected in earth-like planets orbiting stars brighter than $V \sim 10$ (easily discovered by GAIA). This magnitude corresponds to a star at a distance $d \sim 50$ pc for the latest type stars considered (K dwarf stars) and more than ~ 500 pc for the earliest stars (F dwarf stars).

2.2.4a Upper atmosphere emissions

Known planetary atmospheres radiate various non-LTE emission lines in the EUV-FUV. Among them, the giant planets are the richest and strongest emitters (Ly α , H $_2$ Lyman and Werner bands, OI1300Å, etc...). These emissions directly originate from the stellar UV flux and the precipitation of energetic particles into the upper atmosphere. Transposed to exoplanets, the detection of such direct emissions would provide crucial information on the atmospheric composition and dynamics (e.g. atmospheric escape) or the processes responsible for energy transport and deposition. Fluid and thermodynamical states could be inferred from these emissions, as homo-bases and exo-bases altitudes, temperatures, variability etc.

The main difficulty with this approach is to discern the planet emission lines from the parent star emission. [34] reported a possible observational detection of H $_2$ emissions at 1580 Å for HD209458b. There are also tantalizing hints of photo-excited molecular emission from planets orbiting M-dwarfs [38,40]. In parallel, theoretical work [74] has calculated contrast up to 10^{-2} for Ly α emission of the same planets, which can theoretically be achieved for the closest planets ($d < 20$ pc).

2.2.5 The bulk composition of exo-planets

Measuring the bulk composition of extrasolar planetary material can be done from high-resolution UV spectroscopy of debris-polluted white dwarfs [41]. Most known planet host stars will evolve into WDs, and many of their planets will survive (e.g. Mars and beyond in the solar system). The strong surface gravity of WDs causes metals to sink out of the atmosphere on time-scales much shorter than their cooling ages, leading unavoidably to pristine H/He atmospheres. Debris discs form from the tidal disruption of asteroids [58] or Kuiper belt-like objects [13], are stirred up by left-over planets [26], and can subsequently be accreted onto the white dwarf, imprinting their abundance pattern into its atmosphere.

The requirements for an instrument able to address the issues mentioned above are (i) a large collecting area, (ii) a wide field of view (FOV), (iii) a high spatial resolution, and (iv) the ability to perform high dispersion ($D=20,000-50,000$) spectroscopy in the full 920-3200 Å range, as well as offering (v) integral field spectroscopy with moderate dispersion ($D=3000$) over large fields of view ($10' \times 10'$). The large collecting area is necessary to enable the observation of faint M stars and brown dwarf. Sensitivities of 10^{-17} erg s cm $^{-2}$ Å $^{-1}$ are required to obtain good S/N profiles of the target lines. Spectral coverage in the 912 – 1150 Å bandpass is needed as the bulk of the warm/cold H $_2$ gas is only observable at $\lambda < 1120$ Å (via the Lyman and Werner ($v' - 0$) band systems). Spectropolarimetry would permit following the evolution of the dusty plasma in the circumstellar environment. Dynamical ranges above 100 and resolutions larger than 100,000 are required to separate stellar and planets contribution.

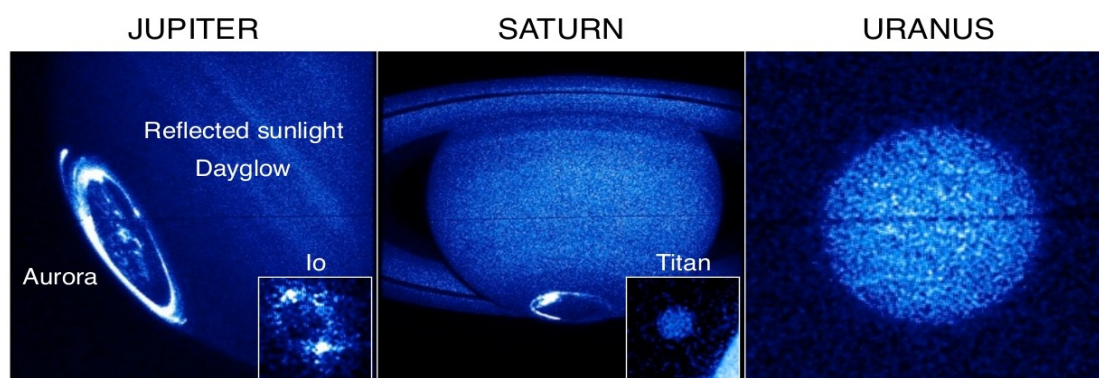
2.3 THE SOLAR SYSTEM

The various bodies in the solar system provide different, complementary pieces of the puzzle that is Solar System formation, the planets having formed by coalescence of planetesimals, of which evidence

remains in the form of asteroids and Trans-Neptunian Objects (TNOs). With its access to thousands of resonance lines and fluorescence transitions, the discovery potential of the proposed UV-visible observatory is vast, covering planetary atmospheres and magnetospheres, planetary surfaces and rings, comets, TNOs and other small solar system bodies, and interplanetary material [16]. Because UV photons interact strongly with matter, UV observations are excellent to determine the composition and structure in low-density regions of the solar system, where plasmas and atmospheres interact, and UV irradiation drives solid-phase and gaseous chemistry, the latter of which dominates the structure of planetary atmospheres above the tropopause. Further, the high spatial resolution afforded in the UV enables exploration of the entire population of solar system bodies as spatially-resolved targets. For example, an 8m diffraction-limited UV telescope will provide spatial resolution of ~ 8 and ~ 55 km at Jupiter's and Neptune's distances, respectively, at 1000 Å. This compares extremely favorably with in situ spacecraft observations; at closest approach to Jupiter Juno's UV spectrometer will provide pseudoimages with ~ 60 km spatial resolution. A non-exhaustive list of science goals is as follows:

2.3.1 Atmospheres and magnetospheres

This UV observatory will be capable of revolutionizing the study of the dynamics and composition of planetary and satellite atmospheres, particularly at the poorly understood planets of Uranus and Neptune. Observations of the abundance and distribution of species such as H, H₂, CO₂, CO, and H₂O, along with many organics and aerosols will be possible. These provide essential insight into source and loss processes, volcanism, aeronomy, atmospheric circulation and long term evolution of planetary atmospheres. These phenomena all connect to wider issues, such as historical and present habitability, terrestrial anthropogenic climate change, and the nature of the presolar nebula. Cryovolcanic and dust plumes, like those observed in situ at Enceladus are potentially within the reach of this project along with detailed studies of other small bodies -- planetary satellites and trans-Neptunian objects (TNOs). Through sensitive, high resolution imaging of planetary and satellite auroral emissions, this UV observatory will also impart a detailed understanding of all of the planets' magnetospheres, revealing the internal magnetic fields and thus internal structure and formation, along with information as to how energy and matter flow through the solar system. Short of sending dedicated spacecraft, this UV telescope is the only way to investigate the magnetospheres of the ice giants Uranus and Neptune. Jupiter's magnetosphere in particular acts as a readily-observable analogue for more distant astrophysical bodies such as exoplanets, brown dwarfs and pulsars. Importantly, this observatory will not duplicate but instead perfectly complement ESA's L1 mission JUICE.



A collage of some solar system UV targets. From left to right, these include the atmospheres, auroras and airglow of Jupiter, Io, Saturn, Titan, and Uranus.

2.3.2 Surfaces and rings

Surface spectroscopy and imaging will provide information on the ice and non-ice condensable (e.g. organic) components of surface layers, revealing interior processes and surface-atmosphere interactions, with important implications for habitability at e.g. Europa. Long term observations of albedo maps will allow the study of seasonal effects. Observations of H, ice, organics and other minor species in planetary ring systems will reveal their composition, formation- and life-times.

2.3.3 Small bodies

The main water products in cometary comas, H, O, and OH, along with the CO Cameron bands, can be uniquely observed at high spatial resolution in the UV. These observations, along with high sensitivity detections of C, S, N, D/H, and rare gases beyond the snow line, unveil the temperature and density of the pre-solar nebula in which the comets formed. Further, observations of O⁺ and CO⁺ will probe the interaction of comets with the solar wind. A major objective will be to detect comet-like activity in TNOs, Main Belt asteroids, Trojans and Centaurs, testing models of thermal evolution at large heliocentric distances. High angular resolution albedo maps and observations of binarity of thousands of Main Belt asteroids will provide information on their composition, and thus source material, while many tens of bright TNOs will be fully characterized by these UV observations.

Solar system science is largely concerned with temporal phenomena, with characteristic timescales varying from seconds to years. A highly-sensitive UV observatory placed at L2 will allow high temporal resolution data to be obtained over long, unocculted intervals, providing a revolutionary advancement over previous Earth-based and, in situ, UV platforms. With its ability to observe many solar system targets, this observatory will uniquely provide the holistic approach required to unravel the story of the solar system.

2.4 THE INTERSTELLAR MEDIUM (ISM)

The UV domain is very rich in information for absorption studies of the ISM, for both its gas and dust phases. For gas studies, the UV domain is fundamental because it includes most important gas phase transitions: from hydrogen in its different forms HI, DI, H₂, HD and from the important heavy elements like C, N, O, Si, Fe, Mg in their dominant ionization stages in neutral regions as well as in ionized and highly ionized media. This is the only way to obtain a direct census of the hydrogen fraction of the Galaxy. All of its forms are accessible in multiple resonance transitions and this provides the key link to the radio and infrared studies of the dust and diffuse gas. These diagnostic lines therefore probe gas from below 100 K to above 10⁶ K and allow to derive precise abundances, ionization, cosmic rays flux and physical parameters as temperature, pressure and density for a wide variety of media, from coronal gas to diffuse neutral and ionized clouds, and translucent or molecular clouds. In the case of the dust, the UV provides a powerful tool with the extinction curves, which presents large variations in the UV and include the largest dust feature at 2175 Å.

Extensive studies of the Galactic ISM have been performed with the previous UV facilities, in particular GHRS and STIS on board HST for wavelengths down to 1150 Å, and Copernicus and FUSE for the far-UV range. *Significant progress on this field will be achieved provided the future instrument has the potential to: (1) increase the sensitivity relative to past or existing spectroscopic facilities in order to extend the ISM studies to other galaxy discs and halos; (2) observe with the same instrument as many different species as possible to probe gas at different temperatures and densities and cover the multiphase aspect of the interstellar medium in galaxies, from the cold phases probed e.g. with CI to the hot phases responsible for the ubiquitous OVI; (3) get simultaneous access to the Lyman series of atomic hydrogen and to the Lyman and Werner bands of molecular hydrogen to derive abundances, gas-to-dust ratios, and molecular fractions; (4) get high resolution of at least 100 000 in the observations of H₂ (and other molecules CO, HD, N₂...), to resolve the velocity structure of the molecular gas, measure line widths for the various J-levels to study turbulence motion and excitation processes in molecular and translucent clouds; (5) extend the observations to high A_v targets in the Milky Way to extend the range of molecular fractions and study depletion and fractionation in denser molecular cores; (6) produce high signal to noise ratios in excess of 100 to measure weak lines like those produced by the very local ISM, from both low ionization species and high ionization species like OVI and CIV, to describe the detailed structure of the local cloud(s) and characterize the nature of the hot gas in the Local Bubble; and (7) observe large samples of targets in a wide range of environments in the Galaxy and other galaxies and couple dust and gas studies of the same sightlines.*

2.5 STELLAR PHYSICS

The UV domain is crucial in stellar physics, all stellar studies benefit from access to the UV range, and some are actually impossible without it. The intrinsic spectral distribution of hot stars peaks and the resonances lines of many species, prone to non-LTE effects, probe the highest photospheric layers, or winds (CIV, NV, etc...), or non-radiative heating in chromospheres in cool stars. Another advantage of UV observations is the extreme sensitivity of the Planck function to the presence of small amounts of hot gas in dominantly cool environments. This allows the detection and monitoring of various phenomena otherwise difficult to observe: magnetic activity, chromospheric heating, corona, starspots on cool stars, and intrinsically faint, but hot, companions of cool stars. The UV domain is also where Sun-like stars exhibit their hostility (or not) towards Earth-like life (see Sects. 2.2 and 2.3), population III stars must have shone the brightest, accretion processes convert much kinetic energy into radiation, the "Fe curtain" features respond to changes in local irradiation, flares produce emission, etc. Moreover, many light scattering and polarizing processes are stronger at UV wavelengths. Organized global magnetic fields in stars interact with their wind and environment, modify their structure and surface abundances, and contribute to the transport of angular momentum. With spectropolarimetry, one can address with unprecedented detail these important issues (from stellar magnetic fields to surface inhomogeneities, surface differential rotation to activity cycles and magnetic braking, from microscopic diffusion to turbulence, convection and circulation in stellar interiors, from abundances and pulsations in stellar atmospheres to stellar winds and accretion discs, from the early phases of stellar formation to the late stages of stellar evolution, from extended circumstellar environments to distant interstellar medium).

Measuring polarization directly in the UV wind-sensitive lines has never been attempted and would be a great leap forward in studying the nonsphericity and magnetic effects thought to be present in stellar outflows. The sampling by a space-based instrument will yield continuous time-series with short-cadence measurements. Such time-series document phenomena on stars that can be impulsive (flares, infall), periodic (pulsations, rotational migration of spots, co-rotating clouds), quasi-periodic (evolution of blobs from hot winds), or gradual (evolution of spots). Such an instrument will thus provide a very powerful and unique tool to study most aspects of stellar physics in general. In particular, it will answer the following long-standing questions, as well as new ones: *What is the incidence of magnetic fields? In which conditions does a dynamo magnetic field develop? Magnetic field dynamics and geometry: how do they affect all aspects of stellar structure and evolution? What are the properties of wind and mass loss? How does a stellar magnetic field influence mass loss, in particular what is responsible for wind clumping, the formation of a circumstellar disc or clouds, and flares? Under what conditions do OB stars become Be stars? How do their discs form, and dissipate again? What causes Luminous Blue Variable outbursts? What happens when a star reaches critical rotational velocity? What is the origin of γ Cas stars behavior? How does binarity affect stellar structure and evolution? What are the properties of the galactic white dwarf population, and what can they teach us regarding fundamental physics? How does accretion occur in X-ray binaries? What is the interplay between mass transfer and radiation pressure? Is UV/visible variability related to X-ray emission? What evolutionary channels lead to type Ia supernovae?* These questions will be answered by studying various types of stars, especially:

2.5.1 Hot stars:

Early-type (OB) stars dominate the ecology of the universe as driving agents, through their luminosities and mechanical inputs (e.g. winds, supernova explosions, novae). For that reason they all display, at least at some moment of their life time, strong variability on a wide range of timescales. This concerns, for example, Of?p stars which have very specific spectral characteristics related to their magnetic field, Be stars which are very rapidly rotating and develop decretion circumstellar disc, γ Cas stars which emit unexplained variable X-ray flux, Bp stars which host very strong fossil magnetic fields, Herbig Be stars which are the precursor of main sequence Ap/Bp stars, β Cep and Slowly Pulsating B (SPB) which pulsate, B[e] with dust and of course O stars, as well as massive binaries such as the Be X-ray binaries and those that harbor O-type subdwarf companions, etc. They are also unique targets for the study of stellar magnetospheres. Their strong, radiatively-driven winds couple to magnetic fields, generating complex and dynamic magnetospheric structures [7,27], and enhancing the shedding of rotational angular momentum via magnetic braking [106,75]. As the evolution of massive stars is particularly sensitive to rotation and mass loss [20,72], the presence of even a relatively weak magnetic field can profoundly influence the evolution of massive stars and their feedback effects, such as mechanical

energy deposition in the interstellar medium (ISM) and supernova explosions (e.g. [30]). Stellar winds from hot massive stars can be structured also on small-scales by the intrinsic “line-driven instability” (LDI). The presence and interactions between density structures on both these scales is poorly understood, and may compromise the reliability of measurements of the properties of the outflows. Spectral diagnostics such as UV resonance and visible recombination lines have different dependencies on density, and will provide crucial constraints for the further development of dynamical hot star wind models, as well as for how the resulting wind structures affect derived quantities such as mass loss and rotation, which are essential inputs for corresponding models of stellar evolution and feedback. Indeed, if mass-loss is poorly constrained, the evolution of massive stars, their fate and feedback can be completely misunderstood. Although clumping appears to be a universal feature of line-driven winds, it is not known how the LDI interacts with other processes that structure the wind. Some key questions concern possible inhibition of the lateral fragmentation of clumps, the effects on the structure within the closed field loops (the so-called “dynamical magnetosphere”), and how these different behaviors alter the interpretation of spectral diagnostics, in particular the determination of mass-loss rates. EUVO can address these issues by providing time-series at high spectroscopic resolution for OB stars.

2.5.2 Binary stars

Although their evolution is often treated in isolation, about half of all stars in the Galaxy are members of binary or multiple systems. In many cases this is not important but when it is, and that is not rare, the effects can be best studied in the UV. Magnetic fields play a central role, as they strongly affect, and are strongly affected by, the transfer of energy, mass and angular momentum between the components. However, the interplay between stellar magnetospheres and binarity are poorly understood, both from the observational and theoretical side. In higher-mass stars ($> 1.5 M_{\odot}$) the incidence of magnetic stars in binary systems provides a basic constraint on the origin of the fields, assumed to be fossil, and on whether such strong magnetic fields suppress binary formation. In low-mass stars, tidal interactions are expected to induce large-scale 3D shear and/or helical circulation in stellar interiors that can significantly perturb the stellar dynamo. Similar flows may also influence the fossil magnetic fields of higher-mass stars. Magnetically driven winds/outflows in cool and hot close binary systems have long been suspected to be responsible for their orbital evolution, while magnetospheric interactions have been proposed to enhance stellar activity. The ultraviolet is of central importance for studying the complex phenomenon of stellar magnetism under the influence of the physical processes and interactions in close binary systems.

2.5.3 White dwarfs

$>95\%$ of all stars will eventually evolve into white dwarfs, and their study is fundamental to a complete understanding of stellar evolution, with implications into galaxy evolution (through the initial-to-final mass relation) and no picture of stellar or galactic evolution can be complete without them. Detailed photospheric abundance measurements are only possible in the FUV (e.g. [8]). However, because they are intrinsically faint, only a handful of white dwarfs have been thoroughly studied with HST and FUSE. The large effective area of EUVO is indispensable to observe a representative sample of a few hundred stars spanning a wide range of ages, masses, and core/photospheric compositions. Narrow metal lines in the FUV spectra of white dwarfs are a powerful diagnostic for a range of physics, including the very sensitive search for low magnetic fields, or the detection of the coupling between scalar fields and gravitational potential [31].

2.5.4 Compact binaries

Binaries containing a white dwarf (WD), neutron star (NS), or black hole (BH) represent some of the most exotic objects in the Universe, and are ideal laboratories to study accretion and outflow processes [69], and provide insight on matter under extreme conditions. Which evolutionary path close compact binaries follow is a key but still unresolved problem. Dynamical evolution of the binary system proceeds through angular momentum loss and tidal coupling. How this occurs has implications for a wide range of other open questions. One of the fundamental questions is the nature of SN Ia progenitors, the standard candles tracing the existence of dark energy. Whether SN Ia descend from single degenerate or double degenerate binaries, or both, is still controversial. Mass accretion makes the evolution of WDs in such compact systems essentially different than isolated stars (e.g. [104]). The study of photospheric emission of these degenerates, as well as the physical and chemical conditions of the accretion flow are crucial to trace the mass transfer/accretion history and the effects on the evolution. This can be only achieved in the UV range through high resolution FUV spectroscopy of statistically significant samples of accreting

WDs spanning a wide range in stellar and binary parameters including magnetic field strengths. In this respect, the high incidence of magnetic accreting WDs compared to single WDs have lead to different but still debated proposals [95].

A large UV-visible mission is also critical to comprehending accretion disc physics in X-ray binaries and of how the disc reacts to changes in the mass transfer rate or how instabilities are driven. Rapid response to transient events, such as Novae, Dwarf Novae and X-ray transients, is of key importance to unveil changes in the physical conditions of accreted or ejected flows through the outburst evolution of the UV emission lines and continuum. Only very recently the onset of jets has also been identified in WDs accreting at high rate [61,62], suggesting that disc/jet coupling mechanism is ubiquitous in all types of binaries (BHs, NS and WDs). Also, the ratio of UV and X-ray luminosities is recently recognized as important discriminator between NS and BH binaries [52] and that FUV continuum could be affected by synchrotron emission from the jet. Furthermore, in the case of wind accretion, the detailed analysis of the wind structure and variability will improve our knowledge both in terms of how accretion takes place, and how mass loss rate is affected by photo-ionization from X-rays (see e.g. the theoretical work [49] and observational work [56] (FUSE), and [25] (HST)). When such L2 mission will be operational, GAIA will have provided accurate distances for a thousand of cataclysmic variables (CVs) and tens of X-ray binaries allowing tight observational constraints to theories of accretion and evolution. In addition UV imaging with high-temporal capabilities can efficiently allow identifying exotic binaries in Globular Clusters such as ultracompact LMXBs (UCBs). These are expected to be abundant in GC cores HST has so far, only identified three.

2.5.6 Supernovae

Ultraviolet spectrophotometry of supernovae (SNe) is an important tool to study the explosion physics and environments of SNe. However, even after 25 years of efforts, only few high-quality ultraviolet (UV) data are available – only few objects per main SN type (Ia, Ib/c, II) – that allow a characterization of the UV properties of SNe. EUVO could be of paramount importance to improve the current situation. The high-quality data will provide much needed information on the explosion physics and environments of SNe, such as a detailed characterization of the metal line blanketing, metallicity of the SN ejecta, degree of mixing of newly synthesized elements, as well as the possible interaction of the SN ejecta with material in the environment of SNe. The utility of SNe Ia as cosmological probes depends on the degree of our understanding of SN Ia physics and various systematic effects such as cosmic chemical evolution. We now know that some "twin" Type Ia supernovae which have extremely similar optical light curves and spectra, they do have different ultraviolet continua [33]. This difference in UV continua was inferred to be the result of significantly different progenitor metallicities. Early-time UV spectrophotometry with UVO of nearby SNe Ia will be crucial for understanding the detailed physics of the explosions, determining if SNe Ia have evolved (and by how much) over cosmic time, and fully utilizing the large samples high-redshift SNe Ia for precision cosmology measurements.

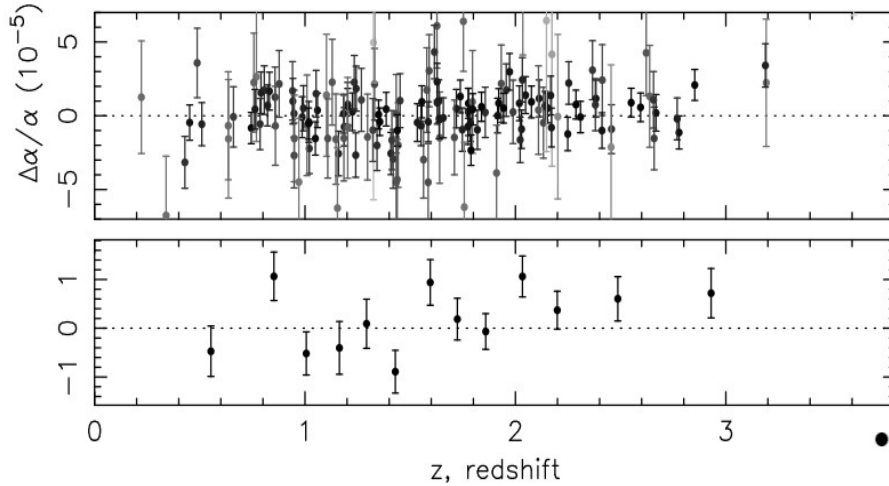
To address these issues one requires (i) a large collecting area, (ii) a wide field of view (FOV), (iii) a high spatial resolution, and (iv) high spectral resolution as well as (v) integral field spectroscopy. Furthermore, the progress achieved in stellar physics thanks to simultaneous UV and visible high-resolution spectropolarimetry will revolutionize our view of stars of all types and age but it requires an increase in sensitivity with respect to HST by a factor of 50-100 to reach the S/N required for the observation of most of the targets. High-resolution spectropolarimetry will make feasible to produce 3D maps of stars and their environment, and understand the impact of various physical processes on the life of stars. These results will have an important impact on many other domains of astrophysics as well, such as planetary science or galactic evolution.

2.5 OTHER RELEVANT PROBLEMS

Fundamental physics – testing the variation of the fine structure constant at z :

The absorption of light along the line of sight to quasars by intervening gas clouds provides a mean to measure the variation of the fine structure constant across the Universe. The fine structure constant, $\alpha = e^2/\hbar c$, is the parameter that governs the strength of electromagnetism; it couples the electromagnetic field to all charged particles in nature. The separation in wavelengths among transitions in the same multiplet depends on α . For instance, alkali-doublet (AD) type transitions, the separation is proportional to α^2 . The Many Multiplets (MM) method compares many transitions from different multiplets, from

different atoms/ions yielding uncertainties as low as $\Delta\alpha/\alpha \approx 10^{-6}$ [29,104,61]. Indications of variations of α with redshift and location in space have been obtained from measurements of the resonance UV multiplets of Fe II, Ni II, Cr II, Zn II and Mg II [106]. These multiplets are observed in the visible range for redshifts $z > 1.6$ henceforth, the current measurements have been obtained with 10-m class ground based telescopes. As shown in the figure, the measurements are very uncertain, especially because of the atmospheric effects (refraction index, sky lines, [61]). Measurements from space would be much more accurate provided that stable high resolution spectroscopy and a high collecting surface to reach $z=2$ is provided. Also, space opens the possibility of using stronger multiplets like the Lyman series of hydrogen.



$\Delta\alpha/\alpha$ measured with VLT/UVES. The bottom panel shows binned values of $\Delta\alpha/\alpha$ where approximately 12 points contribute to each bin from [61]

3. PROPOSED INSTRUMENTATION

The Hubble Space Telescope (HST) and its suite of instruments have inevitably set the standard for UV and visible astronomy from space. Conceived as a general-purpose observatory, it has been able to study a wide variety of scientific goals by providing diffraction-limited imaging and spectroscopy from low to high resolution across a wavelength range from the far ultraviolet to the near infra red. In its current configuration, the telescope is at the peak of its capability and continues to deliver exciting new discoveries. Therefore, any new facility proposed needs to reach well beyond the capabilities of HST. The scientific goals proposed for a new UV/visible observatory require dramatic increases in sensitivity, which must be driven by increases in collecting area and efficiency of the various instruments. A factor 50-100 increase in overall throughput is required to achieve the scientific advances proposed in this white paper. This will be achieved by combination of increased geometric area, through a much larger mirror than HST coupled with advances in reflective coatings, reduced number of reflections through a judicious optical design, enhance detector efficiency, and spectrograph design. However, the required increase in geometric area is likely to be larger than can be provided by a conventional monolithic design contained in the currently available launcher fairings. The baseline telescope design is as follows:

- 8m deployable segmented optical design
- wide-field diffraction limited imaging detector system, with angular resolution 0.01 arcsec
- UV spectrograph with low/medium to high resolution echelle capability, $R=20,000$ -100,000; 900-2000Å and 2000-4000Å, including a long slit or multi-slits
- Spectropolarimetry, $R=100,000$ (1000Å-7000Å)
- Integral field spectroscopy, $R=500$ -1000
- Detectors should have photon counting capability

The following sections discuss these elements in more detail and outline required technology developments.

3.2 Telescope Design

Conventional telescope designs have traditionally been used on space systems. HST is a Ritchey-Chretien design with a 2.4m aperture while the largest telescope launched into space so far (by ESA) is the 3.5m mirror of the Herschel mission. The simplest configuration to implement, a monolithic mirror with a fixed structure, is limited by the available launcher systems and fairing dimensions. For example, the Ariane 5 fairing limits us to a 4m aperture telescope yielding an increase in collecting area of ~2.8 compared to HST. Therefore, achieving the necessary factor 50-120 improvement in effective area a large increase in the telescope mirror dimensions is required coupled with significant enhancements in other areas.

Further enhancements in collecting area require technological development beyond the current state of the art. There are three main possibilities:

1. Enhanced launcher configuration with larger fairing – there are currently no relevant planned upgrades.
2. Off-axis elliptical Mirror design to the largest size acceptable by the fairing – gives only a further factor 2 increase in area, but requires a deployable structure for the secondary mirror.
3. Deployable folded mirror design – JWST has 6.5 m aperture (7.5 x HST area) and uses an Ariane 5 launch, but the system will need enhancements to fulfill the needs of a UV/visible mission. However, a folded mirror system is the only option that can deliver the necessary collecting area, requiring an aperture ~8m.

Some of the necessary technological developments are being addressed in the ESA future technology planning, as well as by NASA technology calls.

3.3 Optical Coatings

The efficiency of reflective optical coatings is of crucial importance to the efficiency of UV/visible telescopes, particularly for the shorter wavelengths where reflectivity can be low. Complex optical systems including primary/secondary mirrors, pick-off mirrors and gratings compound the problem to the power of the number of reflections in the system. For example, a reflectivity of 75% yields a net efficiency of 18% after 3 bounces. If the reflectivity could be improved to 90%, the net efficiency would be 73%, a significant, factor 4, improvement. The standard coating as used in HST is MgF_2 overcoated Al. However, its performance becomes problematic at the shorter FUV wavelengths (below 1150 Å). Alternatives such as SiC and LiF, have considerably better short wavelength performance but are typically only ~60% at the longer wavelengths. This is an area where technical development is already underway, with some promising results, but where further work is required. For example, a very thin MgF_2 coating on Al using Atomic Layer Deposition (ALD) techniques can offer efficiencies >~50% below 115nm while retaining high (90%) performance at the longer wavelengths.

3.4 Detector Systems

Separate detector systems will be required for the imaging and spectroscopic elements of the mission payload. Microchannel Plate (MCP) detectors have been the device of choice in the FUV for all recent missions, while CCDs have been used for NUV and visible bands. Each has advantages and disadvantages. CCDs are integrating devices, require cooling to reduce noise and have limited QE at the shorter wavelengths. MCPs are photon counting, have better QE at short wavelengths but have limited count rate capability and suffer gain sag over time. An ideal detector system would combine the best features of both devices - high QE, good dynamic range, long-term stability and low noise. Future detector developments may also include ICCD, sCMOS or ICMOS devices.

Back illumination, delta doping and AR-coating can improve the UV QE of CCDs (>50% @ 1250Å – L3CCD, JPL). The first devices are now being tested on sub-orbital missions. CMOS devices have similar QE performance to CCDs, and their integrated nature, with all ancillary electronics inbuilt in the device produces a compact, lower mass, lower power detector with high radiation tolerance. Radiation damage caused by cosmic rays in CCDs can cause increased dark noise, image artifacts and hot pixels. “Low Light Level” CCDs have electron multiplication, which can make their performance close to photon counting with suitable cooling (<-100C), but have higher stabilization requirements and other operating issues including ageing effects. In the long term, CMOS devices are likely to replace CCDs as general

sensors. Therefore, it is important that development work is focused on the UV performance of the former.

The performance of microchannel plate detectors is still improving; advances in MCP technology using coatings applied to borosilicate glass MCPs with atomic layer deposition (ALD) can provide higher MCP gain and detection efficiency and significantly extend detector lifetime, while readout developments employing lower noise techniques (such as capacitive division) in conjunction with adaptive digital filtering enable higher dynamic range to be achieved. ITAR restrictions on ALD coated MCPs is currently imposed only for MCPs of pore size 5 microns or smaller. A UK collaboration has recently been funded to investigate application of ALD to photon counting detector techniques.

The electronic readouts are a key element of the performance of MCP detectors. For example, a capacitive division image readout (C-DIR) with pulse digitization and scene-dependent adaptive digital filtering algorithms to combine very high resolution MCP-limited imaging at moderate count rates whilst extending the dynamic range of the detector to several Mcount/s. The combination of the low noise C-DIR readout allows operation at lower MCP gain and higher point source count rates, and the lifetime advantages conferred by ALD technology extends the useful detector operation by over an order of magnitude.

MCP performance is limited by the photocathodes efficiency. In the NUV semitransparent solar-blind CsTe photocathodes can now routinely achieve >30% QE and GaN operated in reflection mode (possibly usable for spectroscopy) has been measured at 80% at 1200Å but only 20% thus far in semi-transparent mode necessary for imaging. CsI is still likely the current photocathode of choice for FUV vacuum photocathodes, capable of >40% QE in the 400-1200Å range. But developments in III-nitride materials are likely in the medium term.

A temporal resolution of 0.01 second for data imposed by science and/or spacecraft requirements is straightforwardly achievable by all of the candidates detector technologies.

3.5 Gratings

The grating technology is well developed and grating configurations are likely to be similar to those employed in recent missions. First order holographic gratings deliver excellent efficiency (~60% peak) and low scatter. However, the highest resolving powers available are a few tens of thousands and echelle systems are still required to deliver the highest spectral resolution required. The challenge is to produce systems with the low scatter achieved by 1st order gratings. New designs of low order echelle gratings with magnifying cross dispersers show promise.

3.6 Additional elements

Integral field spectroscopy: We could dedicate a small fraction of the large focal plane for an integral field spectrometer. By limiting the short wavelength to ~180 nm this could be achieved with a coherent bundle of fused silica fibers. To go to much shorter wavelengths it would be necessary of have fibers of MgF₂ or microshutter arrays, requiring technological development.

Spectropolarimetry: Magnesium fluoride is bi-refrangent allowing the separation of an image into its two perpendicular polarizations. This would need to be inserted into two spectrographs, to produce spectra of these.

To conclude:

We present a compelling and relevant set of science investigations leading to the probing of the building blocks of life in the Universe. These investigations will be possible and enabled if the European Ultraviolet-Visible Observatory (EUVO) moves forward to its next development phase.

4. BIBLIOGRAPHY (Max. 2 pages)
- [0] *Fundamental questions in Astrophysics: guidelines for future UV observatories.*
Eds. A. I. Gómez de Castro & W. Wamsteker, Springer 2006, ISBN 1-4020-4838-6 also
published as a series in Vol 303 of *Astrophysics and Space Science*.
- [1] Ade et al. 2013, arXiv: 1303.5062
- [2] Adelberger et al. 2003 ApJ 584, 45
- [3] Alexander et al. 2006, MNRAS 369, 229, 2006
- [4] Ardila et al. 2013 arXiv:1304.3746
- [5] Armitage et al 2002, MNRAS 334,248
- [6] Armitage 2007, ApJ 665,1381
- [7] Babel & Montmerle 1997, A&A, 323, 121
- [8] Barstow et al 2003, MNRAS 341, 870
- [9] Basko et al. 1977, ApJ 215,276
- [10] Beasley et al. 2012, SPIE 8443E, 3QB
- [11] Bergin et al. 2007 - 2007, prpl.conf...,751B
- [12] Beygu et al. 2013, AJ 145, 120
- [13] Bonsor et al. 2011, MNRAS 414, 930
- [14] Boronson et al. 2011, ApJ 562, 925
- [15] Brosch et al. 2006, MNRAS 368, 864
- [16] Brosch et al. 2006, ApSS 303, 103
- [17] Cardamone et al. 2009, MNRAS 399, 1191
- [18] Cen & Ostriker 1999 ApJ 519, L109
- [19] Cen & Ostriker 2006 ApJ 650, 560
- [20] Chiosi, C. & Maeder, A. 1986, ARA&A, 24, 329
- [21] Coffey et al. 2012, ApJ 749, 136
- [22] Danforth & Shull 2008 ApJ 679, 194
- [23] Danforth et al. 2011 ApJ 679, 560
- [24] Davé et al. 1999 ApJ 511, 521
- [25] Davé et al. 2001 ApJ 552, 473
- [26] Debes et al. 2012, ApJ 747, 148
- [27] Donati et al. 2002, MNRAS, 333, 55
- [28] Dzigan & Zucker 2012, ApJL 753, L1
- [29] Dzuba et al. 1999, PRL, 82, 888
- [30] Ekström et al. 2008,IAU Symposium, V.250, 209
- [31] Flambaum & Shuryak 2008, WNMP 2007, AIP Conference Proceedings 995, 1
- [32] Fleming et al., SPIE 8145E, 245F
- [33] Foley&Kirshner 2013, ApJ 769, id. A L1
- [34] France et al. 2010, ApJ 712, 1277
- [35] France et al. 2010, ApJ 715, 596
- [36] France et al. 2011, ApJ 734, 31F
- [37] France et al. 2012a, ApJ 756, 171F
- [38] France et al. 2012b, ApJ 750L, 32F
- [39] France et al. 2012c, SPIE.8443E, 05F
- [40] France et al. 2013 , ApJ, 763, 149F
- [41] Gaensicke et al. 2006, Ap&SS 306, 177
- [42] Gaensicke et al. 2012, MNRAS, 424, 333
- [43] Gómez de Castro & Verdugo 2001, ApJ 548, 976
- [44] Gomez de Castro & Verdugo 2003, ApJ 597, 443
- [45] Gómez de Castro et al. 2006, ApSS 303, 33
- [46] Gómez de Castro & Verdugo 2007, ApJ 654, L91
- [47] Gómez de Castro & Marcos-Arenal 2012, ApJ 749, id.A 190
- [48] Gómez de Castro 2013, Arxiv:
- [49] Goodson et al. 1997, ApJ 489, 199
- [50] Green et al. 2006 ASP Conf. Series 348, 559
- [51] Hernández et al. 2007, ApJ 671, 1784
- [52] Hickson et al. 1993 Ap. Lett. Comm. 29, 1
- [53] Hynes & Robinson 2012, ApJ 749, 3

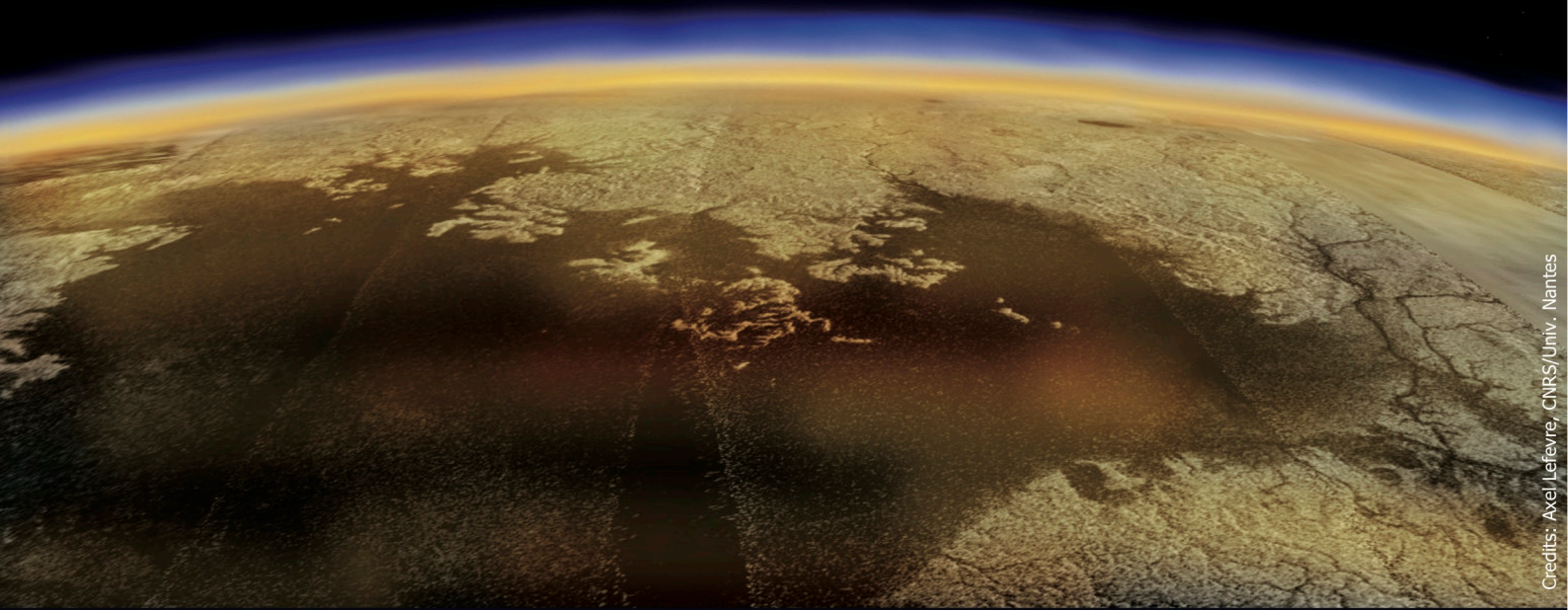
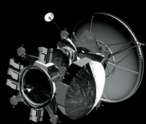
- [54] Hollenbach et al. 2005, ApJ 631, 1180
- [55] Ingleby et al. 2011, AJ 141, id.A 127
- [56] Ingleby et al. 2011b, ApJ 743, id.A 105
- [57] Johns-Krull et al. 2000, ApJ 539, 815
- [58] Jura 2003, ApJ 584, L91
- [59] Kellar et al. 2012, AJ 143, 145
- [60] Kehrig et al. 2013, accepted for publication in MNRAS
- [61] King et al. 2012, MNRAS 442, 3370
- [62] Koerding et al. 2008, Sci 320, 1318
- [63] Koerding et al. 2011, MNRAS 418, L129
- [64] Lammer, 2010, EPS 2010, Italy, <http://meetings.copernicus.org/epsc2010/>, 584
- [65] Lecavelier des Étangs et al. 2001, Nature 412, 706
- [66] Lecavelier des Étangs et al. 2003, A&A, 407, 935
- [67] Lecavelier des Étangs et al. 2012, A&A 543L, 4L
- [68] Lehner et al. 2007 ApJ 658, 680
- [69] Linsky et al. 2010, ApJ 717, 1291L
- [70] Long & Knigge 2002, ApJ 579, 752
- [71] López-Santiago et al. 2010, A&A 514, id.A 97
- [72] Maeder & Meynet 2000, ARA&A, 38, 143
- [73] Martínez-Arnaiz et al., 2011, MNRAS 414, 2629
- [74] Menager et al. 2013, Icarus, in press
- [75] Meynet et al. 2011, A&A, 525, L11
- [76] Mosser et al. 2009, A&A 506, 245
- [77] Ohyama & Hota 2013, ApJL 767, L29
- [78] Pagano 2013, Planets, Stars and Stellar Systems Vol. 4, by Oswalt, Terry D.; Barstow, Martin A., Eds. ISBN 978-94-007-5614-4.
- [79] Penz et al. 2008, A&A 477, 309
- [80] Pizzolato et al. 2003, A&A 397, 147
- [81] Rappaport et al. 2012, ApJ 752, id.A 1
- [82] Richter et al. 2006, A&A 445, 827
- [83] Robinson & Catling 2013, 757, 104
- [84] Romanova et al. 2004, ApJ 610, 920
- [85] Romanova et al. 2012, MNRAS 421, 63
- [86] Salyk et al. 2009, ApJ 699, 330
- [87] Segura et al. 2010, AsBio 10, 751S
- [88] Shull et al. 2012, ApJ 759, 23
- [89] Simcoe et al. 2006, ApJ 637, 648
- [90] Skumanich 1972, ApJ 171, 565
- [91] Stanonik et al. 2009, ApJ 696, 65
- [92] Stelzer et al. 2013, MNRAS 431, 2063
- [93] Thom & Chen 2008 ApJS 179, 37
- [94] Tilton et al. 2012 ApJ 759, 112
- [95] Tout et al. 2008, MNRAS 387, 897
- [96] Trilling et al. 2002, A&A 394, 241
- [97] Tripp et al. 2008 ApJS 177, 39
- [98] Vidal-Madjar et al. 1998, A&A 338, 694
- [99] Vidal-Madjar et al. 2003, Nature 422, 143
- [100] Vilchez & Iglesias-Paramo 1988, ApJ 508, 248
- [101] Von Rekowski & Brandenburg 2006, Astron. Nachr. 327, 53
- [102] Ward 1997, ApJL 482, L211
- [103] Wanders et al 1995, ApJ 453, L87
- [104] Webb et al. 1999, PRL, 82, 884
- [105] Webb et al. 2011, PRL, 107, 191101
- [106] Weber & Davis 1967, ApJ, 148, 217
- [107] Yang et al. 2012, MNRAS 427, 1614
- [108] Zitrin & Brosch 2008 MNRAS 390, 408
- [109] Zorotovic et al. 2011, A&A 536, 86

THE SCIENCE GOALS AND MISSION CONCEPT FOR A FUTURE EXPLORATION OF TITAN AND ENCELADUS

White Paper submitted to ESA for the Definition of the L2 and L3 Missions
in the ESA Science Programme

Spokesperson: Gabriel Tobie

Laboratoire de Planétologie et Géodynamique de Nantes
UMR-6112, CNRS, University of Nantes
2, rue de la Houssinière, 44322 Nantes cedex, France
+33 276645161, gabriel.tobie@univ-nantes.fr



Coordinators: Gabriel TOBIE (LPGNantes, FR) & Nick TEANBY (U. Bristol, GB)

Section coordinators (6): Athena COUSTENIS (LESIA, FR), Ralf JAUMANN (DLR, DE), Jürgen SCHMIDT (U. Oulu, FI), François RAULIN (LISA/UPEC, FR), Nick TEANBY (U. Bristol, GB), Gabriel TOBIE (LPGNantes, FR).

Contributors (24): Andrew COATES (MSSL-UCL, GB), Nathalie CARRASCO (LATMOS, FR), Daniel CORDIER (UTINAM, FR), Remco DE KOK (SRON, NL), Wolf D. GEPPERT (Stockholm U., SE), Tim LIVENGODD (NASA-Goddard, US), Jean-Pierre LEBRETON (LPC2E & LESIA, FR), Axel LEFEVRE (LPGNantes, FR), Kathy MANDT (SWRI, US), Francis NIMMO (UCSC, US), Conor NIXON (NASA-Goddard, US), Lucy NORMAN (UCL, GB), Robert PAPPALARDO (JPL-Caltech, US), Frank POSTBERG (Heidelberg/Stuttgart U., DE), Sébastien RODRIGUEZ (CEA-SAP, FR), Dirk SCHULZE-MAKUCH (WSU, US), Jason SODERBLOM (MIT, US), Anezina SOLOMONIDOU (LESIA, FR; U. Athens, GR), Katrin STEPHAN (DLR, DE), Ellen STOFAN (Proxemy Res., US), Elizabeth TURTLE (JHU-APL, US), Roland WAGNER (DLR, DE), Robert WEST (JPL-Caltech, US), Joseph WESTLAKE (JHU-APL, US)

Supporters (121): Achilleos N. (GB), Atkinson D. (US), Baland R.-M. (FR), Barabash S. (SE), Beauchamp P. (US), Behoukova M. (CZ), Bézard B. (FR), Bezacier L. (FR), Billebaud F. (FR), Boireau L. (FR), Bollengier O. (FR), Boudon V. (FR), Bourgeois O. (FR), Briois C. (FR), Brown R. (US), Cadek O. (CZ), Capria M. T. (IT), Carpy S. (FR), Castillo-Rogez J. (US), Charnay B. (FR), Choblet G. (FR), Coll P. (FR), Cooper J. (US), Cornet T. (FR), Cours T. (FR), D'Aversa E. (IT), Dandouras I. (FR), Daglis I. (GR), Dehant V. (BE), Dobrijevic M. (FR), Dougherty M. (GB), Durry G. (FR), Ehrenfreund P. (US), Encrenaz P. (FR), Ferri F. (IT), Filachione G. (IT), Fletcher L. (GB), Frampton R. (US), Galand M. (GB), Gautier D. (FR), Grasset O. (FR), Grindrod P. (GB), Grodent D. (BE), Gupta S. (GB), Hartogh P. (DE), Hayes A. (US), Hersant F. (FR), Hilchenbach M. (DE), Howett C. (US), Hussmann H. (DE), Iess L. (IT), Irwin P. (GB), Jackman C. (GB), Jolly A. (FR), Leese M. (GB), Kalousova K. (CZ), Karatekin O. (BE), ten Kate I. L. (NL), Khurana K. (US), Kostiuik T. (US), Langevin Y. (FR), Lavvas P. (FR), Lebonnois S. (FR), Lebleu D. (FR), Le Gall A. (FR), Leitner J. (AT), Le Mouélic S. (FR), Lilensten J. (FR), Lopes R. (US), Lopes-Moreno J. J. (ES), Lorenz R. (US), Lunine J. (US), Mahaffy P. (US), Massironi M. (IT), Marounina N. (FR), Masters A. (JP), Matson D. (US), Mitri G. (IT), Mocquet A. (FR), Monin J.-L. (FR), Monteux J. (FR), Morse A. (GB), Mousis O. (FR), Mueller-Wodarg I. (GB), Neish C. (US), Neubauer F. (DE), Nna Mvondo D. (FR), Paillou P. (FR), Patel M. (GB), Piccioni G. (IT), Poncy J. (FR), Poulet F. (FR), Rannou P. (FR), Reh K. (US), Reme H. (FR), Rengel M. (DE), Rodrigo R. (ES), Rolfe S. (GB), Schmitt B. (FR), Sittler E. (US), Sohl F. (DE), Sotin C. (US), Spencer J. (US), Spilker L. (US), Szego K. (HU), Thissen R. (FR), Tokano T. (DE), Tortora P. (IT), Tosi F. (IT), Toubanc D. (FR), Trigo-Rodriguez J. (ES), Vandaele A. (BE), Van Hoolst T. (BE), Vermeesen B. (NE), Vinatier S. (FR), Vuitton V. (FR), Wahlund J.-E. (SE), Wilquet V. (BE), Yelle R. (US), Young R. (GB), Yseboodt M. (BE).

Full list of participants with their affiliation can be found here:

<http://www.sciences.univ-nantes.fr/lpgnantes/wp-esa-titan-enceladus>

SUMMARY & MOTIVATION

The Cassini-Huygens mission, which has been in orbit around Saturn since July 2004 and released the Huygens probe that landed on Titan's surface, has revealed Titan and Enceladus to be very enigmatic objects—introducing extraordinary challenges for geologists, astrobiologists, organic chemists, and planetologists. Titan, Saturn's largest satellite, is unique in the Solar System with its extensive atmosphere made mostly of N₂, with a column density ten times that of Earth's atmosphere. The presence of a few per cent methane provides the basis for rich organic chemistry, leading to production of complex hydrocarbon compounds from the upper atmosphere down to the surface. Methane is close to its triple point on Titan, which gives rise to a methanological cycle analogous to the terrestrial hydrological cycle, characterized by cloud activity, precipitation, river networks and lakes. ***Exploring Titan in greater detail than ever possible with Cassini-Huygens offers the possibility to study an environment analogous to Earth, where methane takes on water's role, and to analyse complex chemical processes that may have prevailed on the early Earth.***

The discovery of jets of water vapour and ice grains emanating from Enceladus' south pole in 2005 was one of the major milestones of the Cassini-Huygens mission. Despite its small size (ten times smaller than Titan), Enceladus is the only icy moon in the Solar System proven to have current endogenic activity. The jets, which form a huge plume of vapour and ice grains above Enceladus' south pole, are associated with abnormally elevated heat flow along tectonic ridges, called "Tiger stripes". Sampling of the plume by Cassini's instruments revealed the presence of water vapour, organics and salt-rich ice grains, suggesting that the jet sources are connected to subsurface salt-water reservoirs. ***The surprising activity of Enceladus provides a unique opportunity to analyse materials coming from its water-rich interior, potentially containing compounds of prebiotic interest, and to study today aqueous processes that may have been important on many other icy worlds in the past.***

Based on the outstanding Cassini-Huygens discoveries, several missions to Titan and Enceladus have been proposed since 2007 to both ESA and NASA. Titan and Enceladus have also been identified as highest priority targets in the NRC Planetary Decadal Survey (2013–2022). In 2007, the Titan and Enceladus Mission (TandEM, Coustenis et al. 2008) was proposed as an L-class (large) mission in response to the first ESA's Cosmic Vision 2015–2025 Call, and accepted for further studies, with the goal of exploring Titan and Enceladus. It was merged in 2008 with NASA's Titan Explorer concept, to become the joint large (Flagship) Titan and Saturn System Mission (TSSM). TSSM was an ambitious mission dedicated to in-depth long-term exploration of Titan, with multiple flybys of Enceladus, using an orbiter and two in situ elements (hot air balloon and lake lander). ***While the mission concept considered in the present White Paper builds on TSSM heritage, it focuses on science goals that could be achieved from the combination of a Saturn-Titan orbiter and a Titan balloon.***

1- SCIENCE GOAL A: TITAN AS AN EARTH-LIKE SYSTEM

Titan is a complex world more like the Earth than any other: it is the only place besides Earth known to have a dense, predominantly nitrogen, atmosphere; it has an active climate and meteorological cycle where the working fluid—methane—behaves under Titan's conditions the way that water does on Earth; and its geology—from lakes and seas to broad river valleys and mountains—while carved in ice is, in its vast range of processes, again most like Earth. Beneath this panoply of terrestrial processes an ice crust floats atop what appears to be a liquid water ocean. ***Science Goal A seeks to understand how Titan functions as a world, in the same way that one would ask this question about Venus, Mars, and the Earth.*** How are the distinctions between Titan and other worlds in the Solar System understandable in the context of the complex interplay between geology, hydrology, meteorology, and aeronomy? Is Titan an analogue for some aspect of the Earth's history, past or future? Why is Titan endowed with an atmosphere when, for example, Jupiter's moon Ganymede, virtually identical in size and mass, is not? Although the Cassini-Huygens mission provided major advances for understanding the atmospheric and geological processes at work on Titan, many questions remain unanswered—addressing these questions require future missions designed to explore these worlds.

1.1 Titan's atmosphere:

Meteorology and methane cycle: Titan is the only body in the Solar System besides Earth with an active 'hydrologic' cycle, featuring methane rather than water as the condensable fluid in clouds, rain, and surface reservoirs (lakes). Titan has an obliquity of 26.7° (similar to Earth) giving pronounced seasonal change during its 29.5-year orbit around the Sun. Cassini imaging shows that Titan's

tropospheric clouds range from mid-latitude streaks, equatorial bands and patches, and summer polar convective outbursts, to a long-lived high-altitude winter polar cap (Rodriguez et al., 2009, 2011; Le Mouélic et al., 2012). Polar cloud activity disappears as equinox approaches, but measurements are intermittent. Occasional equinoctial tropical methane monsoons have been speculated and recent observations of an equatorial arrow-shaped cloud (Turtle et al., 2011, Fig. 1) suggest an inter-tropical convergence zone following solar insolation maximum (Mitchell et al., 2011). Such storms and subsequent rainfall could explain the formation of equatorial fluvial surface erosion and intricate valley networks like those seen around the Huygens probe landing site (Tomasko et al., 2005).

Despite these major advances, Cassini's observations are limited by incomplete time coverage, due to the Saturn-centric orbit, which leads to a sporadic time series dependent on widely spaced Titan flybys/encounters. This makes it difficult to accurately assess global trends and seasonality in Titan's rapidly changing methane cycle accurately. Continuous measurement of cloud distribution, characteristics, and evolution is now essential to constrain energy budgets, surface methane sources, and tropospheric circulation. It is unclear how often it rains on Titan, which determines erosion rates and how fluids are transferred around the globe. Investigating Titan's active meteorology will reveal the controlling factors that link surface and atmospheric interactions.

Global dynamics, circulation, and seasonal change: Titan provides a giant natural laboratory for testing Earth-based climate and general circulation models under different physical conditions. Much of Titan's general circulation, however, remains to be constrained—particularly above 500 km, below 100 km, within polar vortices, and in equatorial regions (Flasar et al., 2009). Measurements of circulation in these regions are important for constraining how strongly tropospheric and stratospheric circulation are coupled, whether the tropopause wind minimum and near surface flow reversal observed by Huygens (Bird et al., 2005) are global features, and how thermal energy is redistributed in the upper atmosphere. A Titan orbiter would provide continuous coverage for remote sensing instruments and a regular series of radio occultations that would provide major advances. By comparing such measurements with legacy data from Voyager and Cassini, long-term climate trends could also be investigated.

Studying the seasonal variation of circulation patterns has been limited by the coverage available from each flyby, which is non-uniform and globally incomplete, but significant progress has been made. In addition to winds derived by Huygens probe radio tracking (Bird et al., 2005), there have also been cloud-tracking attempts, but these have been severely limited by Cassini's short flyby durations. Indirect measurements of the middle-atmospheric zonal winds have been derived from temperature fields via the thermal wind equation and vertical winds have been probed using chemical tracers and adiabatic heating (Coustenis et al., 2010; Teanby et al., 2012). Maps of atmospheric temperature and composition also show that Titan's atmospheric rotation axis is different from that of a solid body (Achterberg et al. 2008). The cause of this is currently unclear, but could be linked to thermal tides. Gravity waves appear to be an important and controlling feature of Titan's atmosphere and a major contributor to the super-rotation, but have only been directly profiled at a single point and a single season by the Huygens probe, so at present are very poorly constrained. Titan's detached haze varies in altitude from about 300 km to 500 km and is apparently synchronized to seasonal cycles (West et al., 2011). The nature of this feature is still under debate, but the haze is clearly an important tracer of atmospheric dynamics in Titan's upper stratosphere. The vertical distribution of haze in the troposphere is also unknown and could provide nuclei for condensation. Further progress in all dynamical aspects of Titan's atmosphere now requires high temporal resolution monitoring from an orbiter.

Temperature structure: Titan's temperature structure and its evolution over seasonal timescales are essential for understanding climatic evolution, global circulation, photochemistry, and condensation processes. The chemical composition of Titan's atmosphere is similar to Earth's nitrogen-dominated atmosphere and both planets feature a distinct stratosphere. Titan's atmosphere is unique within the Solar System because it is so cold and extends to such high altitude, with evidence that upper atmospheric temperature is influenced by both magnetospheric plasma (external influence) and atmospheric waves (internal), causing it to change rapidly. Yet it appears relatively stable. Atmospheric escape or irreversible photochemical conversion are processes that could eliminate Titan's current atmosphere, which is either replenished by processes not yet fully understood or else is being explored in a temporary state. A better understanding and more data on the chemistry of Titan's atmosphere and its interaction with the surface will enable us to solve this question.

Competition between absorbed ultraviolet and emitted infrared radiation creates Titan's pronounced Earth-like stratopause, which is not present on Mars or Venus, making Titan especially relevant for comparison with Earth. Although the single Huygens atmospheric profile suggests waves could be important above 500 km (Fulchignoni et al., 2005), there are serious gaps in our knowledge that Cassini will not be able to address. We have little information for the altitude ranges 0–100km and 500–950km. These regions are important because many of Titan's trace species condense around 100 km, whereas the 500–950-km region links the bulk neutral atmosphere to photochemical source regions. Additionally the location of the homopause on Titan has long been an issue of debate and has major implications for atmospheric escape rate (Strobel et al., 2009). Thermospheric temperatures strongly influence escape and are important for determining how the system as a whole operates and how or if equilibrium is maintained. Furthermore, a currently inaccessible region below 200 km within the winter polar vortex is a potential site for exotic chemistry on nitrile/hydrocarbon ices and could have parallels with Earth's Antarctic polar chemistry and polar stratospheric clouds (Flasar et al., 2005).

Complex chemistry and haze formation: A mission to Titan is the most effective way to study complex organic, inorganic and ionic chemistry at all altitudes, from formation of complex hydrocarbon species high in the atmosphere down through the bulk atmosphere. Currently formation of complex molecules, ions, and haze is poorly constrained. The present lack of constraints on the aerosol chemical composition precludes clear benchmarks for further synthetic organic solids studies in the laboratory. A mission dedicated to Titan is essential to determine haze composition, how its formation in the ionosphere links to other levels, if its composition changes with altitude, how it affects climate, and its role in the methane cycle and surface alteration.

Titan's atmosphere is rich in organic compounds sourced from a highly active photochemical cycle that begins in the ionosphere (~1000 km) and influences the entire atmospheric column (Lavvas et al., 2008). Discovery of the extent of the chemical complexity of Titan's ionosphere was one of Cassini's major breakthroughs and encompasses neutral species, positive ions, and negative ions (Coates et al., 2007). Cassini found unexpected negative ions up to 13,800 amu/q and positive ions up to ~1000 amu, implying that linked neutral-cation-anion chemistry could play a key role in haze formation (Lavvas et al., 2013). Some amount of nitrogen inclusion occurs in these compounds, but it is unclear how far into the chemical chains nitrogen prevails. Ion structures are at present unconstrained and could be chains, rings or even fullerenes, which may play a role transporting oxygen to the surface. Furthermore, although Cassini's instruments have not yet detected molecules more complex than benzene (C₆H₆) below 500km, recent laboratory work by Gudipati et al. (2013) showed that complex chemistry may be important throughout the entire atmospheric column, including both upper and lower atmospheric regions. Similar processes could have occurred in early Earth's atmosphere and studying Titan would help probe the mechanisms involved.

1.2 Titan's geology:

Titan's dense atmosphere is opaque at most visible and near-infrared wavelengths and the surface is only visible using reflected sunlight at specific windows in the near infrared and at RADAR wavelengths. Prior to Cassini's arrival at Saturn in 2004, bright and dark features were observed in near-infrared images acquired by the Hubble Space Telescope and earth-based telescopes (e.g., Coustenis et al., 2005). But the lack of spatial resolution precluded any geological interpretation. Observations performed by the Cassini RADAR, the Visual and Infrared Mapping Spectrometer (VIMS), and the Imaging Science Subsystem (ISS), have revealed a remarkably diverse Earth-like surface in terms of landforms and geologic features, indicating that Titan shares many characteristics with the Earth (Jaumann et al., 2009; Stephan et al. 2013). Titan's landscapes are shaped by a variety of surficial processes including impact cratering, aeolian, fluvial and lacustrine processes, and also endogenic processes including cryovolcanism and tectonism.

Impact craters: A remarkable characteristic of Titan's surface is the relative paucity of impact craters—one of the many attributes it shares with the Earth—which indicates a relatively young and active surface (Jaumann et al., 2009; Wood et al., 2010; Neish and Lorenz, 2012). Wood et al. (2010) and Neish and Lorenz (2012) list a total of 60 possible impact craters using Cassini RADAR data (currently covering ~33% of the surface) ranging from 3 to 445 km in diameter. Titan's craters appear in some ways morphologically different from those on airless icy satellites, perhaps due to effects of the atmosphere or subsurface volatiles. Soderblom et al. (2010), for example, report an apparent fluidized-ejecta blanket, similar in morphology to the bright crater outflows of Venus. With so few

preserved craters, the age of Titan's surface remains uncertain with estimates depending both on the cratering chronology model used and the sample set selected; estimates range from ~ 200 Ma to ~ 1 Ga, depending on which crater scaling function is used (e.g., Neish and Lorenz 2012).

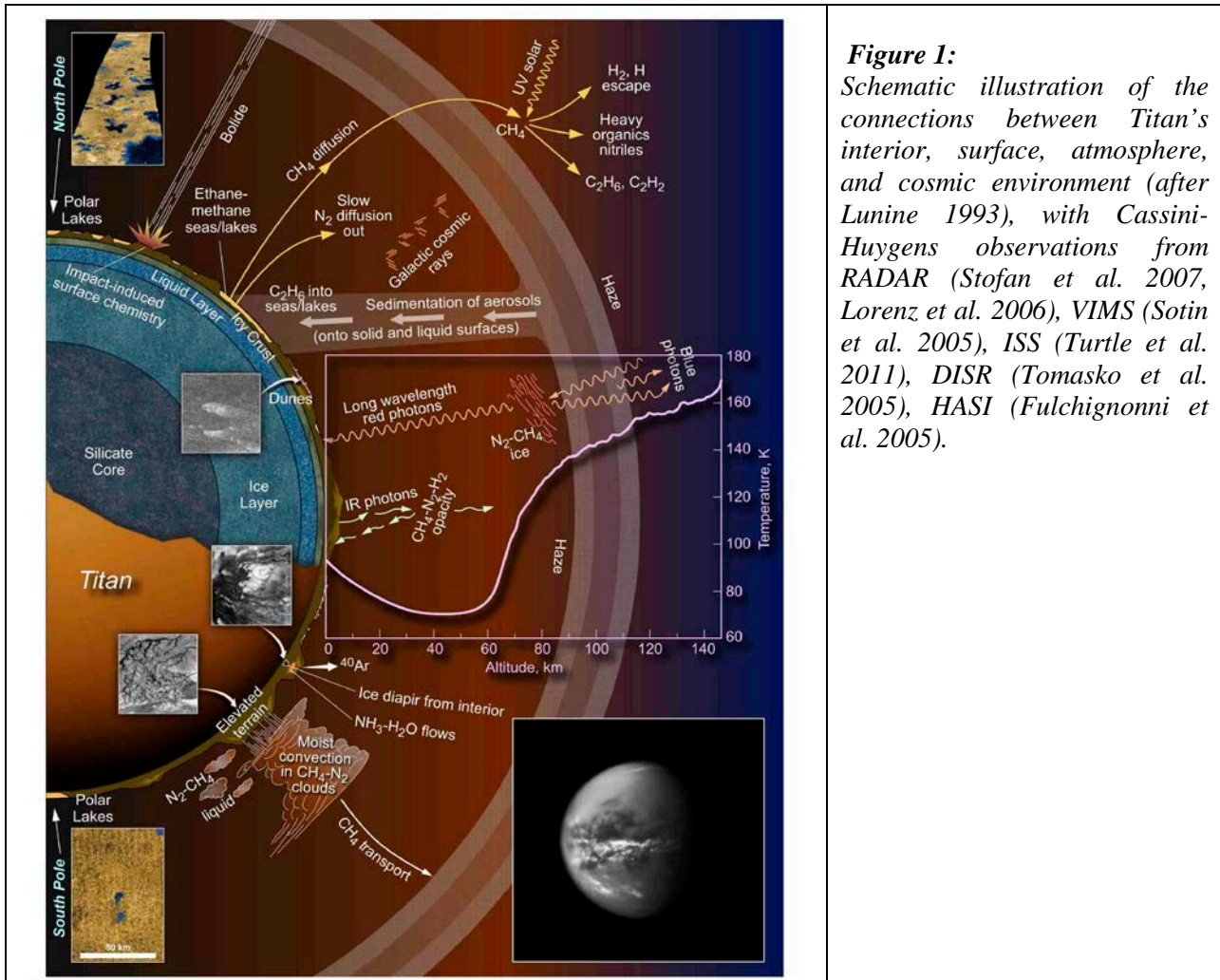


Figure 1: Schematic illustration of the connections between Titan's interior, surface, atmosphere, and cosmic environment (after Lunine 1993), with Cassini-Huygens observations from RADAR (Stofan et al. 2007, Lorenz et al. 2006), VIMS (Sotin et al. 2005), ISS (Turtle et al. 2011), DISR (Tomasko et al. 2005), HASI (Fulchignoni et al. 2005).

Aeolian features and processes: Aeolian activity on Titan has proven to be one of the major forces at work, as is especially apparent at low latitudes. Almost half the terrain within 30° of the equator is covered in dark (presumably organic-rich) streaks or dunes (e.g., Lorenz et al., 2006; Radebaugh et al., 2008). In a few of the best-imaged regions, these dunes are hundreds of kilometres long and ~150-m high. Almost all appear to be linear dunes, a type common in the Arabian, Sahara, and Namib deserts on Earth, but very rare on Mars. These types of dunes typically form in long-lived bidirectional wind regimes. A tidal wind origin has been proposed for Titan, but seasonal wind changes may also play a role. While it has not been demonstrated that these dunes are presently active, they are certainly young relative to other geologic features (cf. Radebaugh et al., 2008). Interestingly, dune morphologies suggest westerly surface winds, at odds with the Huygens wind measurements.

Fluvial features and processes: Fluvial surface modification was very evident at the Huygens landing site (Tomasko et al., 2005; Lorenz et al., 2008; Jaumann et al., 2008, 2009). Not only were steeply incised channels a few kilometres long and ~30-m across observed in the nearby bright highland (Perron et al., 2006; Jaumann et al., 2008), but the view from the probe after landing showed rounded cobbles characteristic of transport in a low-viscosity fluid (Tomasko et al., 2005). Radar-bright channels have been observed at low and mid-latitudes (Lorenz et al., 2008, Langhans et al., 2012), while channels incised to depths of several hundred meters are exposed elsewhere. At high latitudes radar-dark, meandering channels suggest a lower-energy environment where deposition of fine-grained sediment occurs. Whether formation of these larger channels—some of which exceed a kilometre in width—and the large-scale flow features near the landing site (Soderblom et al. 2007; Jaumann et al., 2009) requires a different climate regime remains to be determined. The flow of methane rivers in an unsaturated atmosphere on Titan is analogous to the problem of ephemeral water flow on Mars and terrestrial deserts: determining whether the rivers dry out, freeze solid, or drain into sub-surface alkanifers or ephemeral lakes and seas requires measurement of presently unknown meteorological factors.

Lacustrine features and processes: Extremely radar-dark features at Titan's high latitudes are consistent with liquid-filled lakes and seas ranging in size from less than 10 km² to at least 100000 km² (Stofan et al., 2007). A specular reflection observed in VIMS data also indicates surface liquids (Stephan et al., 2010). Although ethane has been detected as a component of the liquid (Brown et al., 2008), the composition remains largely uncertain (Cordier et al., 2012). Empty lakebeds have been detected (Stofan et al., 2007, Hayes et al. 2008) and the existence of evaporite deposits is suspected (Barnes et al., 2011). The morphology of boundaries between some lakes and their surroundings resembles a terrain flooded by liquids, with the dark material appearing to flood valleys between brighter hilly terrain and in some cases occupying networks of channels that feed into or out of the lakes. Other lakes (e.g., many of the smaller lakes at high northern latitudes and possibly Ontario Lacus in the south) appear to be formed by dissolution (e.g., Cornet et al., 2012). The Huygens landing site is littered with 1–10-cm-scale mostly rounded pebbles, implying they were tumbled and deposited by liquids feeding into a now dry lake bed from dendritic valley systems seen in the Huygens DISR images (Keller et al., 2008). Knowing the depths of the lakes is of high importance, both to constrain the total amount of liquid they contain, as well as to understand the underlying geological processes and 'methanological' cycling that formed them.

Endogenic activity: Cryovolcanism is a process of particular interest at Titan, especially because of the astrobiological potential of liquid water erupting onto photochemically produced organics (e.g., Fortes et al., 2006; Poch et al., 2012). Radiogenic heating in Titan's interior, possibly augmented by tidal heating, can provide enough heat to drive a substantial resurfacing rate (e.g., Tobie et al., 2006). Kinetically, cryovolcanism is much easier in the Saturnian system, where ammonia can facilitate the generation and rise of cryofluids through an ice crust, than in the Galilean satellites (e.g. Fortes et al., 2007). Several candidate sites of cryovolcanism have been identified in Cassini near-infrared VIMS and RADAR data (e.g., Lopes et al., 2013). Evidence for active volcanism, however, is still debated (cf. Moore and Pappalardo, 2011), and the role of cryovolcanism on Titan is an important factor for understanding exchange processes between atmosphere, surface and interior. It thus needs further scrutiny.

The role tectonism plays on Titan is also not well understood. A number of large-scale linear features are seen optically (Porco et al., 2005). Features on Titan called 'morphotectonics' (Solomonidou et al., 2013) are parts of the landscape morphology correlated to tectonics that are/were subsequently subjected to exogenous processes, surficial and/or atmospheric. Such features include mountains (e.g., Radebaugh et al., 2007), ridges (e.g., Mitri et al., 2010), faults (e.g. Radebaugh et al., 2011), and canyons (e.g. Lopes et al., 2010). Radar imagery of some of these features has not helped in their interpretation and is not yet sufficiently widespread to evaluate tectonic patterns, although some linear mountain ranges (e.g., Radebaugh et al., 2007) have been detected, several forming a chevron pattern near the equator. Near-infrared imagery by Cassini VIMS has also shown long ridges (e.g., Soderblom et al., 2007; Jaumann et al., 2009). An outstanding mystery is the nature of the large bright terrain Xanadu and its adjoining counterpart Tsegihi. These areas are distinct optically, and they have unusual radar properties. SAR imagery shows Xanadu to be extremely rugged, much like the Himalayas on Earth, although mountain-forming processes on Titan have not been robustly identified and may differ from place to place (e.g., Radebaugh et al., 2007; Jaumann et al., 2009).

Evidence for a global internal ocean on Titan: A series of geophysical measurements (gravity field, Iess et al., 2012; electric field, Béghin et al., 2012; obliquity, Baland et al., 2011; and shape, Nimmo and Bills 2010) performed by Cassini-Huygens indicate the presence of a global water ocean, likely salt-rich, a few 10s to > 100 km below the surface. Measured tidal fluctuations in the gravity field are consistent with the existence of a decoupling water layer below the ice shell (Iess et al., 2012). Such an ocean, with an elevated concentration of ionic solutes, may also explain the electric field perturbation observed by Huygens and interpreted as a Schumann resonance (Béghin et al., 2012). The salt enrichment as well as the ⁴⁰Ar atmospheric abundance (Niemann et al., 2010), suggest an efficient leaching process and prolonged water-rock interactions. The chemical exchanges associated with water-rock interactions may be quantified by accurately measuring the ratio between radiogenic and non-radiogenic isotopes in noble gases (Ar, Ne, Kr, Xe) in Titan's atmosphere (Tobie et al., 2012). Further tidal monitoring from gravity, topography and rotation data along with magnetic and electric field measurements would provide key constraints on the physical properties of the ocean (depth, density, electric conductivity) as well as the ice shell (thickness, viscosity structure).

1.3 Summary of science questions, investigations and key measurements relevant for Goal A:

	<i>Saturn-Titan Orbiter</i>	<i>Titan Balloon</i>
A: Titan as an Earth-like system		
How does Titan's methane cycle vary with season?	Cloud distribution, lake changes	Rain, surface evaporation, detailed cloud activity
How does Titan's global circulation vary with season?	Cloud tracking, Doppler & thermal winds	Balloon tracking, tropospheric winds
What is Titan's atmospheric temperature structure and how does this influence atmospheric escape, photochemistry, and haze production?	<i>In situ</i> upper atmosphere, remote sensing, occultations	Tropospheric temperatures, condensation processes
How are hazes distributed globally and seasonally? What causes the detached haze layer?	Imaging, spectroscopy, middle atmosphere distribution/composition	Tropospheric hazes distribution and composition
What is the composition of Titan's atmosphere? How does it change over seasonal timescales?	Global spectroscopy, <i>in situ</i> upper atmosphere	High precision <i>in situ</i> gas and isotopes
How old is Titan's surface? What erosional processes are currently active on Titan?	High resolution imaging of craters, aeolian, fluvial and glacial features, surface changes	Very high resolution surface imaging: morphology and activity
What are the properties of Titan's lakes (composition, waves)? How do they vary over seasonal and geological timescales?	Distribution, seasonal change, lake depths	Shoreline imaging, lake clouds, dry lakebeds
What is the composition of Titan's surface and how does it interact with the atmosphere and subsurface?	High-resolution global spectroscopy, hydrocarbon deposits	Very high spatial resolution spectroscopy, deposits
Are cryovolcanic and tectonic processes currently active on Titan? Have these endogenic processes been active in the past?	Cryovolcanic and tectonic features	Very high spatial resolution imaging: morphology, activity
What is the origin of Titan's atmosphere? How has it evolved since its formation? What is the resupply process of methane?	Isotopic ratio, noble gas, atmospheric escape	Isotopic ratios, noble gases
What is Titan's internal structure? What are the properties of any internal ocean and of the icy shell?	Gravity, topography, spin state, magnetic field	Electric field

Table 1.1: List of key questions associated to Goal A and of the physico-chemical quantities to be measured to address them.

Table 1.2: Measurement requirements to address the science questions of Goal A.

<i>Saturn-Titan Orbiter</i>	<i>Titan Balloon</i>
A-1. Detect seasonally driven surface changes in the methane hydrological cycle, in particular lake-extent.	A-12. Determine profiles of T, P, CH ₄ , C ₂ H ₆ and other hydrocarbon mole fraction.
A-2. Map the formation and dissipation of clouds, and determine their altitude.	A-13. Track cloud motions and determine particle size and properties of clouds and haze.
A-3. Determine temperature, wind fields and the abundances of the gaseous + solid constituents in the stratosphere and agnostosphere (500-950 km) versus altitude and latitude, with a goal of detecting seasonally driven changes.	A-14. Determine wind directions in the troposphere and the interaction with the surface in dune fields.
A-4. Collect molecular species (ion and neutral) from one pole to the equator, with an altitude goal of 600 km for <i>in situ</i> orbiter measurements at certain points, covering lower altitudes with remote techniques.	A-15. Acquire regional geological maps at 2.5 m resolution and measure regional topography.
A-5. Determine exchange of energy and escape of major volatile species, including H ₂ , methane and N ₂ , by comprehensive longitudinal sampling.	A-16. Determine the surface abundance of water ice, hydrates and hydrocarbon compounds, at 5-m resolution.
A-6. Map at least 80% of the surface to 50-m resolution, in one near-infrared band.	A-17. Perform subsurface sounding (with vertical resolution < 10 m).
A-7. Map the spatial distribution of simple hydrocarbons and important geologic materials.	A-18. Search for methane source and possible cryovolcanic activity.
A-8. Determine the topography by altimetry over 80% of the surface with 10-m vertical resolution.	A-19. Determine electric and magnetic perturbations.
A-9. Perform the sub-surface sounding (lakes, dunes, crustal layering) with 10-m vertical resolution.	
A-10. Determine Titan's gravity field, and its time-variation, with an accuracy of 10 ⁻⁹ m.s ⁻² at an altitude of 1500 km, and to degree and order 6.	
A-11. Characterize magnetic induction and magnetospheric interactions.	

Table 1.1 summarizes the key questions and corresponding investigations that should be addressed by a future L-class mission to characterize Titan as an Earth-like system. The measurements, both from orbit and in situ from a balloon, required to address these scientific objectives are listed in Table 1.2. Further details on the mission concepts and relevant instruments are provided in Section 4.

2- SCIENCE GOAL B: ENCELADUS AS AN ACTIVE CRYOVOLCANIC MOON

The detection of jets of water vapour and ice particles emanating from the south polar terrain of Enceladus is one of the major discoveries of the Cassini-Huygens mission. This surprising activity has been studied by a suite of instruments onboard the Cassini spacecraft, analysing the plume structure and composition of the vapour and icy grain components (also called dust in the following), their mass ratio, the speed and size distributions of the constituents, the interaction with the Saturnian corotational plasma, as well as the replenishment of the magnetosphere and E ring region with fresh plasma and dust particles. *Science goal B seeks to further characterize the present-day activity of Enceladus, to understand what processes power it and how it affects the Saturnian environment.*

Enceladus is now the only icy world in the Solar System proven to have current endogenic activity (Triton's geysers are believed to be solar-driven). This cryovolcanic activity offers a unique possibility to sample fresh material emerging from subsurface liquid water and to understand how exchanges with the interior controls surface activity. It provides us with an opportunity to study today, phenomena that may have been important in the past throughout the outer Solar System, when tidal effects and/or higher radiogenic heat fluxes could have powered eruptions, melting, and aqueous chemistry in a number of icy bodies.

2.1 Enceladus' plume activity:

Plume characteristics: About 200 kg/s of vapour is ejected from Enceladus' south pole at speeds exceeding 500 m/s (Hansen et al., 2008), which is well above the escape velocity of 240 m/s. The gas is emitted in a broad, vertically extended plume with embedded, collimated and supersonic jets (Waite et al., 2006; Hansen et al., 2008, 2011). The dust plume also exhibits a broad component and localized jets (Porco et al., 2013), but it has a relatively small scale-height (Porco et al., 2006; Spahn et al., 2006; Schmidt et al., 2008), corresponding to slower mean ejection speeds on the order of 100 m/s. Schmidt et al. (2008) infer a dust production rate of 5–10% of the vapour production, although later photometric studies indicate a more massive dust plume (Hedman et al., 2009; Ingersoll and Ewald, 2011). A precise determination of the dust-to-vapour ratio, and variability therein, is now crucial to better understand the physical mechanisms responsible for the activity.

In contrast to the gas plume, only a small fraction (1–5%) of ejected icy dust exceeds the escape velocity of Enceladus and feeds the E ring. Most grains fall back on the surface in a characteristic global “snow” pattern (Kempf et al., 2010; Schenk et al., 2010). The size distribution of this dust was constrained from *in situ* measurements (Spahn et al., 2006) and infrared spectroscopy (Hedman et al., 2009) to roughly follow a power law (exponent -4), extending from the submicron range up to a few microns. Estimating accurately, both the fraction of particles falling back to the surface and the thickness of surface deposit, will provide essential information on the duration of plume activity.

Gas and grain composition: *In situ* measurements by Cassini INMS (Waite et al., 2006, 2009) showed that plume gas consists primarily of water vapour and about ~5–10% other volatiles. The main volatile species are CO₂, NH₃ and a mixture of organic gases (Waite et al., 2009). Amongst the latter are lightweight molecules like methane, acetylene and propane, but recent measurements also indicate even higher molecular weight compounds with masses exceeding 100 amu and aromatic organics (Waite et al., 2011). A molecule with mass 28, which could be attributed to N₂, CO or C₂H₄, was also identified, but due to the lack of resolution, the ratio CO/N₂/C₂H₄ can not be constrained by Cassini. This information is, however, essential in establishing the origin of the volatiles.

Analysing the composition of particles in the E ring and directly in the plume with Cassini CDA, Postberg et al. (2009, 2011) found that nearly all grains contain at least small amounts of sodium (roughly on ppm level), while other grains show much larger fractions of sodium and potassium salts like NaCl, NaHCO₃ and KCl. The conclusion was that these salt-rich grains (~1 % salinity) must directly disperse from salt water. Moreover, the composition inferred by CDA matches the prediction of Zolotov (2007) for the composition of a subsurface ocean that is, or was, in contact with a pristine

rocky core. Nanometre-sized silicate inclusions in E ring ice grains (Hsu et al., 2011) further support this finding. As with the gas phase, the presence of organic compounds is also conjectured for the icy solids (Hillier et al., 2007, Postberg et al., 2008), but their precise nature is currently unconstrained.

In the plume, salt-rich particles were found to be more abundant close to jet sources. Postberg et al. (2011) concluded that these must be larger grains, ejected at lower speeds. As a consequence, the overwhelming part of the dust mass ejected into the plume is salt-rich whereas the small and fast salt-poor grains dominate (by number) the dust fraction that escapes into the E ring. A third type of dust particle was observed by Cassini's plasma instrument (Jones et al., 2009). To be detectable by this instrument, the grains must not be larger than a few nanometres, if singly charged. The locations where these small particles are detected are closely associated with the strongest jets in the plume. Precise determination of the different particle populations and their correlation with the jets is crucial to better understand the source of the jets and their interaction with the Saturnian environment.

Plume interaction with the magnetosphere: Enceladus is the main source of material in Saturn's magnetosphere, playing a similar role to Io in the Jovian system. On the one hand, the gas plume constitutes an obstacle for the corotational Kronian plasma. The deflected plasma forms a system of currents that lead to measurable deviations in the planetary dipolar magnetic field and the corotational electric field (Dougherty et al., 2006; Kriegel et al., 2009, 2011; Jia et al., 2010) and charge exchange collisions lead to an effective deceleration of the corotational plasma. On the other hand, the plume gas feeds a neutral torus around the orbit of Enceladus (Burger et al., 2007; Fleshman et al., 2009). Electron impacts and photoionization ionize neutrals in the plume and torus, thus replenishing the magnetospheric plasma (Tokar et al., 2006, 2008, 2009; Fleshman et al., 2010). The possible importance of dust-charging processes for the electromagnetic field close to the plume has been emphasized (Simon et al., 2011; Kriegel et al., 2011), and the presence of a dusty plasma was conjectured for the plume (Wahlund et al., 2009; Shafiq et al., 2011). Such conclusions are subject to controversial debate and a future mission provides a unique opportunity to verify and quantify the related processes and settle these issues.

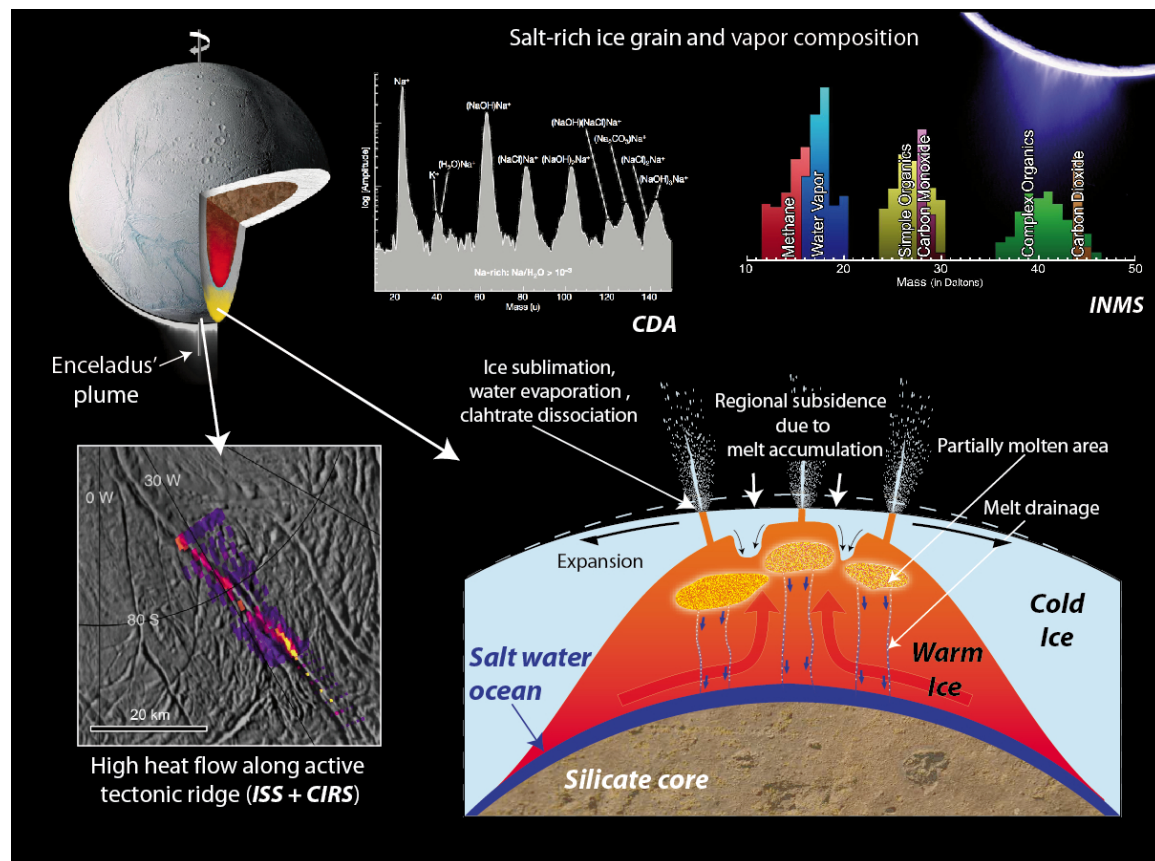


Figure 2: Schematic illustration of connections between Enceladus' plume, surface and interior. Ice shell structure sketch from Tobie et al. (2008) combined with Cassini observations from ISS (Porco et al., 2006), CIRS (Spencer et al., 2011), CDA (Postberg et al., 2009), and INMS (Waite et al., 2009).

2.2 Link with Enceladus' surface and interior

Plume source and surface activity: Qualitatively, a consistent picture based on the presence of liquid water on Enceladus seems to emerge. Only in this way can the elevated salinity of the dust particles be understood (Postberg et al., 2009, 2011). Salt-rich particles could form by direct dispersion from liquid, possibly when bubbles of exsolved gases burst at the liquid's surface (Matson et al., 2012). This scenario could also resolve the problem with the large dust/vapour ratio (Ingersoll and Ewald, 2011) and the observation of fairly large (and massive) particles in the lower parts of the plume (Hedman et al., 2009). Additional mass could condense on these particles when they are transported upwards in the supersaturated vents below the ice crust. In contrast, salt-poor (Postberg et al., 2009, 2011) and nano-sized grains (Jones et al., 2009) might form by direct homogeneous condensation from the gas phase (Schmidt et al., 2008). Water vapour in the plume would then directly evaporate from liquid. Some non-water volatile compounds in the plume gas (Waite et al., 2006, 2009; Hansen et al., 2011) could be released at a quasi-steady rate from the warm ice close to the liquid, or in de-pressurized zones close to the cracks.

While locations of jet sources are relatively well constrained from Cassini ISS data (Porco et al., 2013), the possible time variability of activity still remains an open question (Hurford et al., 2012, Gosmeyer et al. 2013). Stellar occultations from UVIS indicate little or no variability of the vapour production rate over a time-span of five years (Hansen et al., 2011). Moreover, when viewed at different Enceladus's orbital true anomalies, the small observed changes appear to contradict the predictions of tidally-driven eruption models (Hurford et al., 2012). On the other hand, recent analysis of multiple VIMS observations of the plume has been reported to be consistent with such models (Hedman et al., 2012). Generally, limited spatial and temporal resolution of remote Cassini observations as well as the uncertain phase-function of the plume hamper the determination of possible variations with orbital true anomaly. Multiple, dedicated close flybys by a future spacecraft performed at different orbital true anomalies will permit the detection of correlations between eruption activity and tidal cycles, as well as comparison with activity observed by Cassini.

Evidence for subsurface salt-water reservoirs on Enceladus: The detection of salt-rich ice grains in the plume (Postberg et al. 2011) clearly indicates the existence of a subsurface salt-water reservoir on Enceladus. The low K/Na ratio in salt-bearing ice grains (Postberg et al. 2009) further indicates that water-rock interactions at the origin of the salt enrichment occurred at relatively low temperature (Zolotov, 2007; Zolotov et al., 2011). Such enrichment suggests efficient leaching processes and prolonged water-rock interactions. The involved chemical exchanges may be quantified by measuring accurately the ratio between radiogenic and non-radiogenic isotopes in noble gases (Ar, Ne, Kr, Xe) in Enceladus's plume and by determining more precisely the composition of organics, salts, and other minerals contained in sampled ice grains. The size and composition of the internal ocean—if any—must also be addressed. Monitoring tides and rotation (via measurements of altimetry, gravity, surface tracking) as well as magnetic signals may provide essential information on the ocean extent, density and electric conductivity, thus constraining its composition.

Geodynamical evolution of Enceladus: Enceladus' icy surface reveals a wide variety of tectonic structures that record a long history of tectonic deformation (Spencer et al. 2009). Ancient tectonically modified plains identified outside the active south-polar region (Crow-Willard and Pappalardo, 2010) suggest a complex geological history with multiple episodes of enhanced activity. Long-wavelength topography, as well as heterogeneity in crater distribution and tectonic activity, probably reflect strong temporal and spatial variations in ice shell thermal structure (Schenk and McKinnon, 2009; Kirchoff and Schenk, 2009). As indicated by the huge emitted heat flow (Howett et al., 2011), tidal interaction dominates the moon's evolution. Variations of endogenic activity are expected due to coupling with the orbital evolution. However, it is still unknown how activity varies on geological timescales. Surface and sub-surface mapping of Enceladus will permit a better understanding of its long-term evolution.

2.3 Summary of science questions, investigations and key measurements relevant for Goal B:

Table 2.1 summarizes the different key questions and corresponding investigations that should be addressed by a future L-class mission to characterize Enceladus as an active cryovolcanic moon. The measurements required to address these scientific objectives are listed in Table 2.2. Further details on the mission concepts and relevant instruments are provided in Section 4.

	Saturn-Titan Orbiter	Titan Balloon
B: Enceladus as an active cryovolcanic moon		
What is the composition of Enceladus' plume and what implications does this have for origin of the Saturn system icy moons?	In situ gas & ice grain sampling, occultations: organic compounds, noble gases, isotopic ratios	-
What are the characteristics of the plume source region and origin of the plume salts?	Thermal/visible imaging, in situ ice grain sampling, subsurface sounding	-
What dust-plasma interactions occur within the plume? How does the plume interact with Saturn's magnetosphere?	In situ sampling, occultation, magnetic field, plasma	-
What processes drive the surface and plume activities and is this a long-lived or transient phenomenon?	Heat flow, tectonic morphology and distribution, change in plume activity	-
What are the internal structure and properties of any internal ocean? How is this coupled to the ice shell and the rocky core?	Gravity, topography, spin state, magnetic field, orbital dynamics	-

Table 2.1: List of key questions associated to Goal B and of the physico-chemical quantities that will be measured to address them

Saturn-Titan Orbiter
<p>B-1. Determine the spatial distribution, and possible time variations, of gas compounds in the plume, from in situ sampling and occultation, with at least $M/\Delta M \sim 10000$ and a detection limit at least 1000-x lower than Cassini.</p> <p>B-2. Determine density, as well as velocity and size distribution of the ice grains and their spatial and temporal variations with at least 0.5-km spatial resolution.</p> <p>B-3. Determine the mass spectra of ice grains from 1 amu to 500 amu with at least 10-x higher mass resolution and 100-x better spatial resolution than Cassini.</p> <p>B-4. Map surface features at global scale with at least 0.5-km spatial resolution.</p> <p>B-5. Map surface composition (water ice, frost, non-water compounds) at 1 km spatial resolution at global scale, and down to 300-m spatial resolution on regional scales.</p> <p>B-6. Map surface features at 1-m spatial resolution for selected candidate locations, in particular around the identified jet sources.</p> <p>B-7. Acquire regional topography maps of Enceladus' surface with a spatial resolution up to 0.1 km and a vertical resolution of ~ 10 m.</p> <p>B-8. Map the surface temperature distribution in active regions with a precision of 1 K and a spatial resolution of 100 m.</p> <p>B-9. Sound the subsurface up to 50 km in depth, at 10-m vertical resolution over the active south pole region.</p> <p>B-10. Monitor possible time variations in activity of the jet sources.</p> <p>B-11. Determine degree-two gravity field and harmonic amplitudes at precisions of 10^{-7} of Enceladus' surface gravity.</p> <p>B-12. Monitor time variations of the gravity field, spin state and magnetic field.</p> <p>B-13. Measure global plasma and magnetic field structure in the vicinity of Enceladus.</p>

Table 2.2: Measurement requirements to address the science questions of Goal B

3- SCIENCE GOAL C: CHEMISTRY OF TITAN AND ENCELADUS—CLUES FOR THE ORIGIN OF LIFE

Both Titan and Enceladus possess several, if not all, of the key components for habitability: internal liquid water, organic material, energy sources, and a stable environment. Complex organics discovered in Titan's upper atmosphere indicate that a very rich organic chemistry is occurring on Titan. How these organic compounds formed, and how they evolve once at the surface and buried in the subsurface remain open questions. Complex organics are also strongly indicated in Enceladus' plume, though not precisely identified. The presence of salt water as a plume source further increases the astrobiological potential of Enceladus. Titan and Enceladus offer an opportunity to study analogous prebiotic processes that may have led to the emergence of life on Earth. *Goal C seeks to determine the degree of chemical complexity on the two moons, to analyse complex chemical processes that may have prevailed on the early Earth, and to detect compounds of prebiotic interest.*

3.1 Similarities of Titan and Enceladus with the early Earth:

Retracing the processes that allowed the emergence of life on Earth around 4 Ga ago is a difficult challenge since most traces of the environmental conditions at that time have been erased. It is, therefore, crucial for astrobiologists to find extraterrestrial planetary bodies with similarities to our planet, providing a way to study some of the processes that occurred on the primitive Earth, when prebiotic chemistry was active. Although Titan is much colder than the Earth, and has formed in a different environment, it nevertheless presents—perhaps more than any other object in the Solar System—striking analogies with our planet. A major example is Titan's atmosphere, which is composed of the same main constituent, nitrogen, and has a similar structure with a surface pressure of 1.5 bar. Methane's complex cycle on Titan mimics that of water on the Earth and generates, with nitrogen, a large inventory of organic molecules leading to an intense prebiotic chemistry, such as hydrogen cyanide (HCN) and cyanoacetylene (HC_3N) (Raulin et al., 2012). Moreover, Titan is the only planetary body, other than the Earth with long-standing bodies of liquid on its surface, albeit hydrocarbons instead of water. The degree of complexity that can be reached from organic chemistry in the absence of permanent liquid water bodies on Titan's surface, however, has yet to be determined.

Analogies also concern potential habitats. Although quite speculative, Titan lakes could harbour very exotic life (McKay and Smith, 2005; Schulze-Makuch and Grinspoon, 2005), using energy provided by the reduction of hydrocarbons into methane, cell membranes made of reversed vesicles (Norman and Fortes 2011) and no liquid water. Another place is the likely internal liquid water reservoir mixed with some ammonia. Models of Titan's formation even suggest that, initially, this subsurface ocean was in direct contact with the atmosphere and with the internal bedrock (e.g., Tobie et al. 2006, Lunine et al. 2009), offering interesting analogies with the primitive Earth, and the potential implication of hydrothermal vents in terrestrial-like prebiotic chemistry. It cannot be excluded that life may have emerged in this environment and may have been able to adapt and persist since the current conditions are not incompatible with life as we know it on Earth (Fortes, 2000). Thus, it seems essential to confirm the presence of this ocean and determine some of its properties. With the likely presence of subsurface salt-water reservoirs, Enceladus also offers interesting analogies with terrestrial oceans and subglacial lakes. The co-existence of organic compounds, salts, liquid water and energy sources on this small moon provides all necessary ingredients for the emergence of life by chemoautotrophic pathways (McKay et al, 2008)—a generally held model for the origin of life on Earth in deep sea vents. In this model, life on Earth began in deep sea hot springs where chemical energy was available from a mix of H, S, and Fe compounds. The fact that the branches of the tree of life that are closest to the common ancestor are thermophilic has been used to argue a thermophilic origin of life—although other explanations are possible. *In situ* sampling of the plume provides a unique opportunity to search for the specific molecules associated with such systems, including H_2 , H_2S , FeS , etc., and to study processes analogous to those involved with the origin of life on Earth.

3.2 Origin and early evolution of volatile compounds on Titan and Enceladus:

A preliminary requirement for assessment of the astrobiological potential of Titan and Enceladus is to constrain the origin(s) of volatile compounds and to determine how their inventory evolved since satellite accretion. The present-day composition of Titan's atmosphere, as revealed by Cassini-Huygens, results from a combination of complex processes including internal outgassing, photochemistry, escape and surface interactions. The detection of a significant amount of ^{40}Ar (the decay product of ^{40}K) by Cassini-Huygens (Niemann et al., 2005, 2010; Waite et al., 2005) indicated that a few percent of the initial inventory was outgassed from the interior. The chemical exchanges with the surface and the interior as well as the initial composition, however, still remain unconstrained (e.g. Tobie et al. 2013). In contrast, the analysis of Enceladus' plumes provides a unique opportunity to observe eruptive processes in real time and to constrain the composition of the building blocks of the Saturnian system (Waite et al. 2009). Comparison between Titan and Enceladus thus enables us to differentiate what was inherited during formation from what was acquired during their evolution.

The isotopic ratios in different gas compounds observed on Titan and Enceladus constitute crucial constraints to assess their origin and evolution. Cassini-Huygens and ground-based measurements provided isotopic ratios of H, C, N, and O in N_2 , CO, CH_4 , HCN and C_2 hydrocarbons at various altitudes in Titan's atmosphere (e.g. Mandt et al., 2012; Nixon et al., 2012). The measured $^{15}\text{N}/^{14}\text{N}$ ratio is enigmatic because it is about 60% higher than the terrestrial value (Niemann et al., 2010), suggesting an abnormally high fractionation. In contrast, $^{13}\text{C}/^{12}\text{C}$ in methane implies little to no

fractionation, suggesting that methane has been present in the atmosphere for less than a billion years (Mandt et al., 2012). In the absence of a proper initial reference value, however, it is impossible to retrieve information on fractionation processes with confidence. Precise isotopic ratios in the photochemical by-products of CH₄ and N₂ on Titan are also lacking. Except for D/H in H₂O on Enceladus (with large error bars, Waite et al., 2009), no information is yet available for the isotopic ratio in Enceladus' plume gas. Simultaneous precise determination of isotopic ratios in N, H, C and O-bearing species in Enceladus' plume and Titan's atmosphere will permit a better determination of the initial reference ratio and a quantification of the fractionation process due to atmospheric escape and photochemistry. In situ sampling of the plasma and energetic particle environment surrounding Titan is also required to provide a better understanding of present escape processes.

3.3 Titan complex prebiotic-like chemistry:

In Titan's atmosphere, the coupling between CH₄ and N₂ chemistries produces many organics in the gas and particulate phases, especially hydrocarbons, nitriles and complex refractory organics. The latter seem to be well modelled by the solid products, commonly called "tholins", formed in laboratory simulation experiments. Water and oxygen ions coming from a magnetospheric source linked to Enceladus plumes are also involved in this atmospheric chemistry (e.g., Sittler et al. 2009). Could these water-oxygen compounds then be locked up into aerosols? Several organic compounds have already been detected in Titan's stratosphere, including hydrocarbons and nitriles (Coustenis et al., 2007, 2010). Direct analysis of the ionosphere by the INMS instrument during the closest Cassini flybys of Titan shows the presence of many organic species at very high altitudes (1100–1300 km): the INMS and CAPS measurements strongly suggest that high-molecular-weight species (up to several 1000 amu) are present in the ionosphere (Waite et al., 2007). This unexpected discovery revolutionizes the understanding of the organic processes occurring in Titan's atmosphere, indicating that ionospheric chemistry plays a key role in the formation of complex organic compounds in Titan's environment. It is essential to determine ionosphere ion and neutral composition with sufficient mass range and resolution to study a wide range of organically relevant compounds.

The presence of water vapour and benzene has been unambiguously confirmed by the CIRS instrument, which also detected isotopomers of several organics (Coustenis et al., 2010; Nixon et al., 2012). The GCMS data collected during the descent of the Huygens probe show that the middle and lower stratosphere and the troposphere are poor in volatile organic species, with the exception of methane (Niemann et al., 2005; 2010). Condensation of such species on aerosol particles is a probable explanation for these atmospheric characteristics. The Huygens ACP instrument carried out the first *in situ* chemical analysis of these particles. The results show that they are made of nitrogen-rich refractory organics, which release HCN and NH₃ during pyrolysis, supporting the tholin hypothesis (Israel et al., 2005; Coll et al., 2012). These measurements suggest that the aerosol particles are made of a refractory organic nucleus, covered with condensed volatile compounds. However, neither the nature and abundances of the condensates, nor the elemental composition and molecular structure of the refractory part of the aerosols have been determined. Moreover, the chirality of its complex organic part is unknown.

The nitrogen content of the aerosols means they are of immediate astrobiological interest following their production in the upper atmosphere (Hörst et al., 2012). Once deposited on Titan's surface, aerosols and their complex organic content produced by atmospheric chemistry may also follow a chemical evolution of astrobiological interest. Laboratory experiments show that, once in contact with liquid water, tholins can release many compounds of biological importance, such as amino acids and purines (Poch et al., 2012). Such processes could be particularly favourable if liquid water is brought to the surface by cryovolcanism (Lopes et al. 2007) or cratering events (Artemevia and Lunine, 2003). Thus one can envision the possible presence of such compounds on Titan's surface or near subsurface. Long-term chemical evolution is impossible to mimic experimentally in the laboratory. It is, therefore, crucial to be able to perform a detailed chemical analysis (at the elemental, molecular, isotopic and chiral levels) of the various types of surface zones, particularly those where cryovolcanism and impact ejecta (or melt sheets) are or have been present.

3.4 Enceladus' prebiotic aqueous processes:

The jets emanating from Enceladus' south pole are probably the most accessible samples from an extra-terrestrial liquid water environment in the Solar System. In addition to water ice, jets include

CO₂ and several organics such as methane, propane, acetylene, and even higher molecular weight compounds with masses exceeding 100 amu, present in the gas and ice grains (Waite et al., 2009). Most of the erupted ice grains contain significant amounts of sodium and potassium salts (about 1%) indicating that salt water plays an important part as a plume source (Postberg et al., 2009, 2011), which suggests contact with Enceladus' rocky core. The ice grains also carry tiny silicate particles that may have previously floated in the liquid (Hsu et al. 2011). The total heat emission at the south polar Tiger Stripes exceeds 15 GW, and in some of the hot spots where jets emanate, the surface temperatures are estimated to exceed 200 K (Spencer et al. 2011). Such enormous heat output, associated with liquid water in contact with rocks, favours prebiotic processes, providing both an energy source and mineral surfaces for catalysing chemical reactions.

The low molecular weight organics detected by Cassini may be just one part of a suite of organics present in the plume and on the surface. Studies of the nature of these organics could tell us whether or not they are biogenic. The molecular species likely to be produced by such a prebiotic or biotic chemistry—such as amino acids, heterocyclic bases, lipidic compounds and sugars—could be detected in the plume of Enceladus using *in situ* techniques. It is also crucial to confirm the presence of liquid water reservoirs and to constrain their composition, both by remote sensing and *in situ* measurements.

3.5 Summary of science questions, investigations and key measurements relevant for Goal C:

Table 3.1 summarizes the different key questions and the corresponding investigations that should be addressed by a future L-class mission to characterize the complex chemistry of Titan and Enceladus and their astrobiological potential. The measurements, both from orbit and *in situ* from a balloon, required to address these scientific objectives are listed in Table 3.2. Further details on the mission concepts and relevant instruments are provided in Section 4.

	<i>Saturn-Titan Orbiter</i>	<i>Titan Balloon</i>
C: Chemistry of Titan and Enceladus—clues to the origin of life		
What are the compositions of the heavy ions and neutrals found in Titan's upper atmosphere?	In situ plasma, ion, and mass spectrometry	-
What is the composition of Titan's haze? Are there variations in composition with altitude, latitude and/or season?	Spectroscopy/in situ of haze forming and intermediate regions	In situ analysis of aerosols in the troposphere
How do organic compounds evolve on Titan's surface? Do these compounds interact with liquid water in cryovolcanic or impact sites?	Global spectroscopy, subsurface sounding	Very high spatial resolution spectral imaging, surface sampling, subsurface sounding
What is the nature of Enceladus' chemistry? Does the plume contain complex molecules of astrobiological interest?	<i>in situ</i> gas & grain sampling, surf. mapping	-
Do water reservoirs exist at shallow depths on Enceladus? How does/did liquid water interact with rocky and/or organic material on Enceladus?	Surface spectral-mapping, geophysics, <i>in situ</i> gas & grain sampl.	-

Table 3.1: List of key questions associated to Goal C and of the physico-chemical quantities to be measured to address them.

4- MISSION CONCEPT

4.1 Previous mission concepts for post-Cassini-Huygens exploration of Titan and Enceladus

Future exploration of the Saturnian system with a focus on Titan and Enceladus has been considered for quite some time, almost since the first years of the Cassini-Huygens mission. Early discoveries by Cassini-Huygens at Titan and Enceladus (discussed above) demonstrated the need for further exploration of the two satellites with a dedicated orbiter, and a balloon for *in situ* exploration of Titan, with newer instrumentation specifically adapted for the environments revealed by Cassini-Huygens, and possibly at different seasonal periods. To place our proposed mission concept in this context, previously proposed mission concepts are briefly outlined below.

<i>Saturn-Titan Orbiter</i>	<i>Titan Balloon</i>
<p>C-1. Perform chemical analysis of the ions and neutral, including heavy species (up to several 1000 amu) in Titan's upper atmosphere.</p> <p>C-2. Perform chemical analysis of the haze particles in Titan's upper atmosphere: search for variations with latitude, altitude and time</p> <p>C-3. Determine the isotopic composition in major C, H, N, O-bearing species in Titan's atmosphere and Enceladus' plume.</p> <p>C-4. Quantify the different isotopes of noble gases (Ar, Ne, Kr, Xe) in Titan's atmosphere and Enceladus' plume.</p> <p>C-5. Determine the infrared spectra of Titan's surface: search for organics of astrobiological interest, and potential correlation with cryovolcanism or impact sites.</p> <p>C-6. Determine the nature of organics and salts contained in the icy grains of Enceladus' plume.</p> <p>C-7. Search for organics of astrobiological interest in the plume and on the surface near the jet sources.</p> <p>C-8. Perform chiral analysis of organic compounds and search for potential enantiomeric excess</p> <p>C-9. Search for near surface water reservoir on Enceladus</p>	<p>C-10. Perform chemical analysis of the haze particles throughout the descent, and determine spatial and temporal variations in the troposphere.</p> <p>C-11. Determine the location and the composition of complex organics with a few meter resolution.</p> <p>C-12. Identify ammonia, sulfate, inorganic salts and compounds containing phosphorous and other potentially reactive inorganic agents</p> <p>C-13. Determine the thickness of organic deposit (liquid and solid) from subsurface sounding.</p> <p>C-14. Measure noble gases and C, H, O, N isotopes in gas phases and aerosols.</p>

Table 3.2: Measurement requirements to address the science questions of Goal C

The Titan explorer (Leary et al., 2008) and the Titan and Enceladus Mission (TandEM, Coustenis et al., 2009) concepts had been selected respectively by NASA and ESA for studies before they were merged into the joint large (Flagship) Titan and Saturn System Mission (TSSM) concept, which was extensively studied in 2008 (K. Reh and J. Lunine et al.–NASA, and C. Erd, J.-P. Lebreton and A. Coustenis et al.–ESA, TSSM NASA/ESA Joint Summary Report, 2009). TSSM aimed at an in-depth long-term exploration of Titan's atmospheric and surface environment and *in situ* measurements in one of Titan's lakes with goals to explore Titan as an Earth-like System, to examine Titan's organic inventory and to explore Enceladus and the coupling and interaction of the two moons with Saturn's magnetosphere. To achieve these goals, a dedicated orbiter would carry two in situ elements: the Titan montgolfière (hot air balloon) and the Titan Lake Lander, each of which would provide complementary data and analyses directly in the atmosphere and on the surface of Titan, and sound its interior. The mission would launch in the 2023–2025 timeframe on a trajectory using Solar Electric Propulsion (SEP), as well as gravity assists, to arrive ~9 years later for a 4-year mission in the Saturn system. Soon after arrival at Saturn, the montgolfière and Lake Lander would be delivered to Titan. The three TSSM elements would operate as follows:

- The orbiter, powered by MMRTGs (Multi-Mission Radioisotope Thermal Generators), would perform 7 close-up Enceladus flybys and then enter into orbit around Titan for 2 years of dedicated observations.
- The montgolfière would study both Titan's atmosphere and surface from above the equator at low altitude (~10 km) for at least 6 months using MMRTGs.
- The Lake Lander would perform the first extraterrestrial oceanographic experiment by landing in one of the Titan's seas, the Kraken Mare, located at approximately 75° N.

This mission was ranked second in the final decision by the agencies and was not considered for further study. It has, however, inspired several other proposed concepts for smaller size missions:

- Titan Aerial Explorer (TAE) was an M3 candidate for ESA's Cosmic Vision call (Hall et al., 2011). TAE was a pressurised balloon, which was planned to fly in the lower atmosphere of Titan at an altitude of 8 km for 3 to 6 months over Titan's equatorial latitudes, with direct to Earth transmission and no need for an orbiter to relay data.
- The Aerial Vehicle for *in situ* and Airborne Titan Reconnaissance (AVIATR) was an alternative idea to the Titan balloon. In Titan's low gravity and a dense atmosphere, an ASRG (Advanced Stirling Radioisotope Generator) powered airplane could fly more easily than on Earth and could sample directly the atmosphere over large swaths of Titan's surface (Barnes et al., 2012).
- The Titan Mare Explorer (TiME), a Discovery candidate, was a probe focusing on exploring Titan's lakes by landing in and floating across Ligeia mare. This lander was designed to study the chemical composition, wave and geological characteristics of the lakes (Stofan et al.,

2010). A similar idea was the Titan Lake Probe, which included a submarine concept (Waite et al., 2010).

- Another Discovery candidate was the Journey to Enceladus and Titan (JET), a single Saturn orbiter with only two instruments and radio science that would explore the plume of Enceladus and the atmosphere and surface of Titan (Sotin et al., 2011).
- A seismic network has also been considered as part of the geophysical payload of such missions (Lorenz et al., 2009), resembling a geophone array widely used on the Earth, capable of detecting ground motions caused by natural or controlled sources.

4.2 A new ESA-led L-class mission concept for the exploration of Titan and Enceladus

Following the ESA recommendation in the present call for the definition of the L2 and L3 missions regarding limited international cooperation, we propose a mission concept with only two elements (Saturn-Titan Orbiter and Titan Balloon) led by ESA, while the TSSM project relied on an important collaboration with NASA in the same spirit as Cassini-Huygens. Note that in a complementary White Paper dedicated to the exploration of Titan and its lakes, led by G. Mitri, an alternative mission concept relying on the combination of an orbiter and a lake probe is considered.

Mission scenario and elements: For L2 and L3 launch opportunities, the duration of the cruise from Earth to Saturn is estimated to 8-10 years. Once arrived at Saturn, the Saturn-Titan orbiter will deliver the Titan balloon, perform a Saturn Tour Phase of about 2 years with multiple flybys of Titan and Enceladus (and possibly of other moons), and finally be captured around Titan at the end of the Saturn Tour Phase in an elliptical orbit (700 km periapsis to 15000 km apoapsis) followed by a two month aero-breaking phase. This aero-breaking phase will enable the exploration of a poorly known, but chemically critical, part of the atmosphere (700–800 km), with *in situ* atmospheric sampling at altitudes much lower than possible with Cassini. Following the aero-breaking phase, the orbiter will be placed into a circular 1500-km, near-polar orbit, for the orbital science phase. This orbit will allow detailed mapping of all latitudes with very high temporal resolutions. The complete global coverage will provide a substantial increase in our understanding of Titan's climatic system and allow global access to all type of surface terrain, atmospheric phenomena, and upper atmosphere interactions.

The Saturn Tour Phase will be optimized for Enceladus science via numerous flybys targeted over Enceladus' southern plumes and geological features, or potentially other ancient active regions elsewhere on the moon. Additionally, the Saturn Tour Phase will allow direct *in situ* study of the possible transport of (organic) material between Enceladus and Titan, and indirectly to other parts of the Saturnian system.

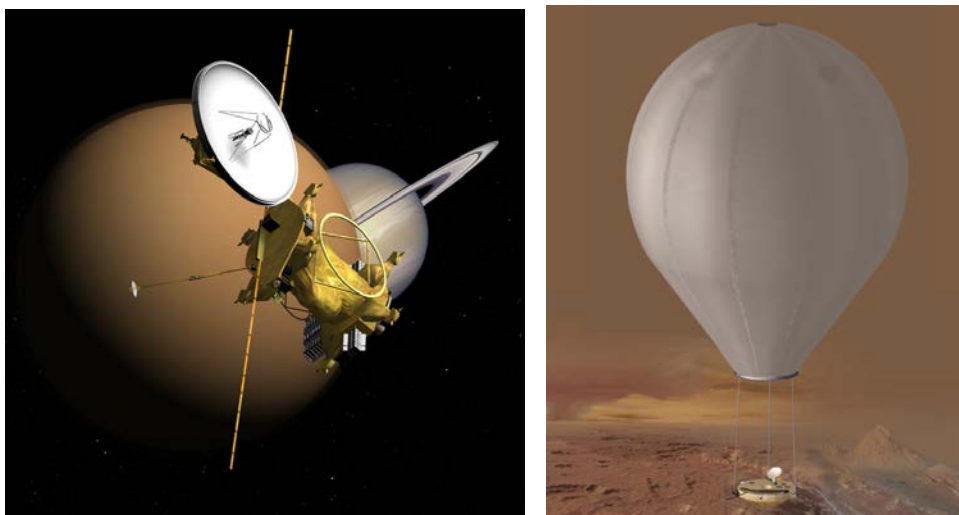


Figure 3: Concepts of orbiter and hot air balloon considered for TSSM

The Titan balloon would be deployed during the first Titan flyby. Data would be transmitted to the orbiter via a steerable high gain antenna, for relay to Earth. Direct-to-Earth transmission may also be considered, which would be more convenient during the Saturn Tour Phase. The balloon would be built on ESA's *in-situ* heritage established with the Huygens probe. A balloon provides an ideal platform for studying Titan's lower atmosphere in detail (e.g. Lorenz, 2008). Penetrating the thick atmosphere to sound the troposphere and surface from orbit is extremely difficult otherwise. The

balloon would be able to sample multiple altitudes in the 1–10-km range and by using Titan’s winds and global circulation pattern could systematically cover many different latitudes and terrain types. Extremely high-resolution surface imaging could be performed, and the chemical composition of the aerosols and atmospheric gases could be directly sampled. Such measurements would be invaluable for interpreting orbital data, studying evolution of the atmosphere, and determining haze composition and the extent of the complex organic chemistry. Titan’s low gravity and thick atmosphere make it an ideal candidate for a balloon-based mission.

<i>Saturn-Titan Orbiter</i>	<i>Titan Balloon</i>
1. High-resolution Imager (2, 2.7, 5–6 μm) and Spectrometer (0.85–2.4/ 4.8–5.8 μm) [A,B,C]	1. Visual Imaging System (two wide angle stereo cameras and one narrow angle camera) [A,C]
2. Penetrating Radar and Altimeter (> 20 MHz) [A,B,C]	2. Imaging Spectrometer (1–5.6 μm) [A,C]
3. Thermal Infrared Spectrometer (7–333 μm) [A,B,C]	3. Atmospheric Structure Instrument and Meteorological Package [A]
4. High Resolution Mass spectrometer (up to 10000 amu) [A,B,C]	4. Electric Environment Package [A,C]
5. Icy grain and organic Dust analyser [A,B,C]	5. Radar sounder (> 150 MHz) [A,C]
6. Plasma suite [A,B,C]	6. Gas Chromatograph Mass Spectrometer (1–600 amu) [A,C]
7. Magnetometer [A,B,C]	7. Radio science using spacecraft telecom system [A, C]
8. Radio Science Experiment [A,B, C]	8. Magnetometer [A, C]
9. Sub-Millimetre Heterodyne [A, B,C]	
10. UV Spectrometer [A, B,C]	

Table 4.1: Tentative instrument payload to address the three mission goals A, B and C.

Strawman instrument payload: Table 4.1 presents a tentative payload that would address the required measurements presented in Tables 1.2, 2.2 and 3.2 for the science goals A, B, C. The proposed instruments will benefit from the heritage of previously successful missions such as Cassini–Huygens, Rosetta, Venus and Mars Express, as well as new missions currently under study (such as JUICE, ExoMars, etc.).

Critical issues and technological developments: For an ESA-led mission to Saturn system, a critical issue concerns the power source. Beyond Jupiter, solar power is inefficient and radioisotope power sources are the only alternative. In the TSSM concept, MMRTGs or ASRGs using ^{238}Pu were considered and were to be provided by NASA. Within Europe the radioisotope ^{241}Am is considered a feasible alternative to ^{238}Pu and can provide a heat source for small-scale radioisotope thermoelectric generators (RTGs) and radioisotope heating units (RHUs) (Sarsfield et al., 2012). ^{241}Am exists in an isotopically pure state within stored civil plutonium at reprocessing sites within the UK and France—about 1000 kg of ^{241}Am exist in the civil PuO_2 stockpile of the UK and France. A study is underway to design a process that will chemically separate ^{241}Am (Sarsfield et al., 2012). The development of ^{241}Am -based RTGs is under consideration by ESA and should be available at high TRL before the proposed L2/L3 launch windows.

Following the TSSM pre-selection, a feasibility study by CNES and JPL was initiated in order to optimize the design of a hot air balloon under Titan’s conditions. The assessment was based on ^{238}Pu -RTGs, which, in addition to providing electric power, were the heat source for generating buoyancy of the balloon. The use of ^{241}Am -RTGs, which provide 20% less decay heat per unit mass, will require further assessment of the feasibility. A pressurized air balloon, as proposed in the TAE project, may also be considered as an alternative. A detailed comparison between the different approaches will be needed to determine the best option for *in situ* exploration of Titan’s atmosphere. Instrumenting the balloon heat shield with a simple seismometer and possibly other simple instruments that would sit at the surface after landing was also considered in TSSM. Such option would require further study to evaluate their feasibility and utility. Finally, supports from national agencies will be essential in developing the new generation of highly capable instrumentations, as well as in pursuing experimental and modelling efforts initiated with Cassini-Huygens, in order to be ready for this next rendezvous with Titan and Enceladus.

REFERENCES

- Achterberg, R. K. et al. (2008), *Icarus* 197(2), 549-555.
- Artemieva N. and Lunine J. (2003), *Icarus* 164, 471–480.
- Baland, R. M. et al. (2011), *Astron. Astrophys.*, 530, id.A141, 6 pp.
- Barnes, J. W. et al., 2012. *Experim. Astronomy* 33, 55-127.
- Barnes, J. W. et al. (2011), *Icarus*, 216, 136-140.
- Béghin, C. et al. (2012), *Icarus*, 218, 1028-1042.
- Bird, M. K. et al. (2005), *Nature* 438(7069), 800-802.
- Brown, R. H. et al. (2008), *Nature*, 2008, 454, 607-610
- Burger, M. H. et al. (2007), *J. Geophys. Res.*, 112 (A6), CiteID A06219.
- Coates, A. J. et al. (2007), *Geophys. Res. Lett.* 34(22), L22103.
- Coll P. et al. (2012), *Planet. Space Sci.*, 10.1016/j.pss.2012.07.006.
- Cordier, D. et al (2012), *Planet. Space Sci.*, 61, 99-107.
- Cornet, T. et al. (2012), *Icarus*, 218, 788-806.
- Crow-Willard, E.N. and R.T. Pappalardo (2010), LPSC 41, Abstract #2715.
- Coustenis, A. et al. (2005), *Icarus*, 177, 89-105.
- Coustenis, A. et al. (2007), *Icarus* 189, 35–62.
- Coustenis, A. et al. (2009), *Experimental Astronomy*, 23,893-946.
- Coustenis, A. et al. (2010), *Icarus* 207, 461-476.
- Dougherty, M. et al. (2006), *Science*, 311, 1406-1409.
- Crow-Willard, E.N. and R.T. Pappalardo (2010). 41st LPSC, Abstract #2715.
- Flasar, F. M. et al. (2005) *Science* 308, 975-978.
- Flasar F. et al. (2009) In: *Titan from Cassini-Huygens*. Brown R, Lebreton J, Waite J, editors. Springer.
- Fleshman, B. L. et al. (2010), *Geophys. Res. Lett.*, 37, CiteID L03202.
- Fortes A. D. (2000), *Icarus*, 146, 444–452.
- Fortes, A.D., and Grindrod, P.M. (2006), *Icarus*, 182, 550-558.
- Fortes, A.D. et al. (2007), *Icarus*, 188, 139-153.
- Fulchignoni, M. et al. (2005), *Nature* 438, 785-791.
- Gosmeyer, C. et al. (2013), *AAS Meeting #221*, #353.16.
- Hall et al. (2011), *IPPW8*, Portsmouth, VA, USA.
- Hansen, C. et al. (2008), *Nature*, 456, 477-479.
- Hansen, C. et al. (2011), *Geophys. Res. Lett.*, 38, CiteID L11202.
- Hayes A. et al. (2008), *Geophys. Res. Lett.*, 35, CiteID L09204.
- Hedman, M.M. et al. (2009), *Astrophys. J.*, 693, 1749-1762.
- Hedman, M.M. et al. (2012), *AGU Fall Meeting*, abstract p32a-12.
- Hillier, J. K et al. (2007), 377, 1588–1596.
- Hörst, S.M. et al. (2012), *Astrobiology*, 12, 9, doi: 10.1089/ast.2011.0623.
- Howett, C. J. A. et al. (2011), *J. Geophys. Res.*, 116, E3, CiteID E03003.
- Hsu, H.-W. et al. (2011), *J. Geophys. Res.* 116 (A8), CiteID A08213.
- Hurford, T. A. et al. (2012), *Icarus*, 220, 896-903.
- Iess, L. et al. (2012), *Science*, 337, 457-460.
- Ingersoll, A. P. and Ewald, S. P. (2011), *Icarus*, 216, 492-506.
- Israel G. et al. (2005), *Nature*, 438, 796–799.
- Jaumann, R. et al. (2008), *Icarus*, 197, 526-538.
- Jaumann, R. et al. (2009). In: *Titan from Cassini-Huygens*, eds Brown et al. p. 75 – 140, Springer.
- Jia, Y.-D. et al. (2010), *Geophys. Res. Lett.*, 37, CiteID L09203.
- Jones, G. H. et al. (2009), *Geophys. Res. Lett.*, 36, CiteID L16204.
- Keller, H. U. et al. (2008), *Planet. Space Sci.* 56, 728 – 752.
- Kempf, S. et al. (2011), *Icarus*, 206, 446-457.
- Kirchoff, M. R. and Schenk, P. (2009), *Icarus*, 202, 656-668.
- Kriegel, H. et al. (2009), *Planet. Space Sci.*, 57, 2113-2122.
- Kriegel, H. et al. (2011), *J. Geophys. Res.*, 116 (A10), CiteID A10223.
- Langhans, M.H. et al. (2012). *Planet. Space Sci.*, 60, 34-51.
- Lavvas, P. et al. (2013), *PNAS*, 110(8), 2729-2734.
- Lavvas, P. P. et al. (2008), *Planet. Space Sci.* 56, 1-66.
- Leary et al. (2008), http://www.lpi.usra.edu/opag/Titan_Explorer_Public_Report.pdf.
- Le Mouelic, S. et al. (2012), *Planet. Space Sci.* 60, 86-92.
- Lopes R. M. C. et al. (2007), *Icarus*, 186, 395–412.
- Lopes, R.M.C. et al. (2010), *Icarus*, 205, 540-558.
- Lopes, R.M.C. et al. (2013), *J. Geophys. Res.*, 118, 1-20.
- Lorenz, R. D. et al. (2006), *Science*, 312, 724 – 727.
- Lorenz, R.D. et al. (2008), *Planet. Space Sci.*, doi:10.016/j.pss.2008.02.009.
- Lorenz, R. D. (2008), *J. British Interplanet. Soc.*, 61, 2-13.
- Lorenz, R. D. et al. (2009), *White paper, Solar System Decadal Survey*.
- Lunine J. (1993), *Rev. Geophysics*, 31, 133-149.
- McKay C. P. and Smith H. D. (2005), *Icarus*, 178, 274–276.
- Lunine, J. et al. (2009). In: *Titan from Cassini-Huygens*, eds Brown et al. p. 35–59, Springer.
- Mandt, K. E. et al. (2012), *Astrophys. J.*, 749, id.160, 14 pp.
- Matson, D. L. et al. (2012), *Icarus*, 221, 53-62.
- McKay C. P. et al. (2008), *Astrobiology*, 8, 909–919.
- Mitchell, J. L. et al. (2011), *Nature Geoscience* 4,

- 589-592.
- Mitri, G. et al. (2010), *J. Geophys. Res.* 115, E10002.
- Moore, J. M., and Pappalardo, R.T. (2011), *Icarus*, 212, 790 – 806.
- Neish, C. D., and R. D. Lorenz (2012), *Planet. Space Sci.*, 60, 26 – 33.
- Niemann, H.B. et al. (2005), *Nature*, 438, 779–784.
- Niemann, H.B. et al. (2010), *J Geophys. Res.- Planets*, 115, E1206.
- Nimmo, F. and Bills, B. (2010), *Icarus*, 208, 896-904.
- Nixon, C. et al. (2012), *Astrophys. J.*, 749, id. 159.
- Norman, L. H. and Fortes, A. D. 2011. *Astronomy & Geophysics*, 52, 39-42.
- Perron, J.T. et al. (2006), *J. Geophys. Res.*, 111, E11001, doi:10.1029/2005JE002602.
- Poch, O. et al. (2012). *Planet. Space Sci.*, 61, 114-123.
- Porco, C.C. et al. (2005), *Nature*, 434, 159 – 168.
- Porco, C. C. et al. (2006), 311, 1393-1401.
- Porco, C. C. (2008), *Scientific American*, Dec.
- Porco, C. C. et al. (2013), 44th LPSC, abstract n° 1719.
- Postberg, F. et al. (2008), *Icarus*, 193, 438-454.
- Postberg, F. et al. (2009). *Nature*, 459, 1098–1101.
- Postberg, F. et al. (2011). *Nature*, 474, 620–622.
- Radebaugh, J. et al. (2007), *Icarus* 192, 77 – 91.
- Radebaugh, J. et al. (2008), *Icarus*, 194, 690 – 703.
- Radebaugh, J. et al. (2011), *Icarus*, 211, 672-685.
- Raulin, F. *Chem. Soc. Rev.*, 41, 5380 - 5393
- Rodriguez, S. et al. (2011), *Icarus*, 216, 89-110.
- Rodriguez, S. et al. (2009), *Nature*, 459, 678-682.
- Sarsfield, M. J. et al. (2012), *European Nuclear Conference*, Manchester, UK, 9-12 Dec.
- Schenk, P. and McKinnon, W. B. (2009), *Geophys. Res. Lett.*, 36, CiteID L16202.
- Schenk, P. et al. (2010), *Icarus*, 211, 740-757.
- Schmidt, J. et al. (2008), *Nature*, 451, 685-687.
- Schulze-Makuch, D. and Grinspoon, D. H. (2005), *Astrobiology*, 5, 560-567.
- Simon, S. et al. (2011), *J. Geophys. Res.*, 116 (A4), CiteID A04221.
- Sittler, E. C. et al. (2009), *Planet. Space Sci.*, 57, 1547-1557.
- Shafiq, M. et al. (2011), *Planet. Space Sci.*, 59, 17-25.
- Soderblom, J. M. et al. (2010), *Icarus* 208, 905-912.
- Soderblom, L, et al. (2007), *Planet. Space Sci.*, 55, 2015–2024.
- Solomonidou, A. et al. (2013), *Plan. Space Sci.*, 77, 104-117.
- Sotin, C. et al. (2005), *Nature*, 435, 786-789.
- Sotin, C. et al. (2011), 42nd LPSC, The Woodlands, TX, p. 1326.
- Spahn, F. et al. (2006), *Science*, 311, 1416-1418.
- Spencer, J. R. et al. (2009), In: *Saturn from Cassini-Huygens*, eds. Dougherty et al., p. 683-724, Springer.
- Spencer, J. R. et al. (2011), 42nd LPSC, The Woodlands, TX. *LPI Contribution No. 1608*, p. 2553
- Stephan, K., R. et al. (2010), *Geophys. Res. Lett.* 37, L07104, doi:10.1029/2009GL042312.
- Stephan, K. et al. (2013). In: *The Science of Solar System Ices*, edited by M. S. Gutipati and J. Castillo-Rogez. *Astrophys. and Space Sci. Libr.* 356, p. 279 – 367, Springer
- Stofan, E.R. et al. (2007), *Nature*, 445, doi:10.1038/nature05438.
- Stofan, E. R. et al., 2010. *LPSC*, Woodlands, TX, USA.
- Strobel, D. F. (2009), *Icarus* 202(2), 632-641.
- Teanby, N. A. et al. (2012), *Nature* 491(7426), 732-735.
- Tobie, G. et al. (2006), *Nature*, 440, 61-64.
- Tobie G. et al. (2008), *Icarus*, 196, 642-652.
- Tobie, G. et al. (2012), *Astrophys. J.*, 752, id. 125
- Tobie G. et al. (2013), In: *Titan: Interior, Surface, Atmosphere, and Space Environment*, eds Müller-Wodarg et al., Cambridge University Press, p. 29-62.
- Tokar, R. L. et al. (2006), *Science*, 311, 1409-1412.
- Tokar, R. L. et al. (2008), *Geophys. Res. Lett.*, 35, CiteID L14202.
- Tokar, R. L. et al. (2009), *Geophys. Res. Lett.*, 36, CiteID L13203.
- Tomasko, M.G. et al. (2005), *Nature*, 438(8), 765–778.
- TSSM NASA/ESA Joint Summary Report (2009), ESA-SRE(2008)3, JPL D-48442, NASA Task Order NMO710851.
- Turtle, E. P. et al. (2011), *Science* 331, 1414-1417.
- Wahlund, J.-E. et al. (2009), *Planet. Space Sci.*, 57, 1795-1806.
- Waite, J. H. et al. (2005), *Science*, 308, 982-986.
- Waite, J. H. et al. (2006), *Science*, 311, 1419-1422.
- Waite, J. H. et al. (2007), *Science*, 316, 870–875.
- Waite, J. H. et al. (2009), *Nature*, 460, 487-490.
- Waite, J. H. et al. (2010), *EGU*, Vienna, Austria.
- Waite, J. H. et al. (2011), 42nd LPSC, The Woodlands, abstract n°1608.
- West, R. A. et al. (2011), *Geophys. Res. Lett.*, 38, L06204.
- Wood, C.A. et al. (2010), *Icarus*, 206, 334-34.
- Zolotov, M. Y. (2007), *Geophys. Res. Lett.* 34, CiteID L23203.
- Zolotov, M. Y. et al. (2011), *DPS-EPSC*, Nantes, France, p. 1330

THE GRAVITATIONAL UNIVERSE

A science theme addressed by the *eLISA* mission observing the entire Universe



Prof. Dr. Karsten Danzmann
Albert Einstein Institute Hannover
MPI for Gravitational Physics and
Leibniz Universität Hannover
Callinstr. 38
30167 Hannover
Germany

karsten.danzmann@aei.mpg.de

Tel.: +49 511 762 2229
Fax: +49 511 762 2784

Detailed information at
<http://elisascience.org/whitepaper>

*The last century has seen enormous progress in our understanding of the Universe. We know the life cycles of stars, the structure of galaxies, the remnants of the big bang, and have a general understanding of how the Universe evolved. We have come remarkably far using electromagnetic radiation as our tool for observing the Universe. However, gravity is the engine behind many of the processes in the Universe, and much of its action is dark. Opening a gravitational window on the Universe will let us go further than any alternative. Gravity has its own messenger: Gravitational waves, ripples in the fabric of spacetime. They travel essentially undisturbed and let us peer deep into the formation of the first seed black holes, exploring redshifts as large as $z \sim 20$, prior to the epoch of cosmic re-ionisation. Exquisite and unprecedented measurements of black hole masses and spins will make it possible to trace the history of black holes across all stages of galaxy evolution, and at the same time constrain any deviation from the Kerr metric of General Relativity. *eLISA* will be the first ever mission to study the entire Universe with gravitational waves. *eLISA* is an all-sky monitor and will offer a wide view of a dynamic cosmos using gravitational waves as new and unique messengers to unveil The Gravitational Universe. It provides the closest ever view of the early processes at TeV energies, has guaranteed sources in the form of verification binaries in the Milky Way, and can probe the entire Universe, from its smallest scales around singularities and black holes, all the way to cosmological dimensions.*

AUTHORS

Pau Amaro Seoane, Sofiane Aoudia, Gerard Auger, Stanislav Babak, Enrico Barausse, Massimo Bassan, Volker Beckmann, Pierre Binétruy, Johanna Bogenstahl, Camille Bonvin, Daniele Bortoluzzi, Julien Brossard, Iouri Bykov, Chiara Caprini, Antonella Cavalleri, Monica Colpi, Giuseppe Congedo, Karsten Danzmann, Luciano Di Fiore, Marc Diaz Aguilo, Ingo Diepholz, Rita Dolesi, Massimo Dotti, Carlos F. Sopuerta, Luigi Ferraioli, Valerio Ferroni, Noemi Finetti, Ewan Fitzsimons, Jonathan Gair, Domenico Giardini, Ferran Gibert, Catia Grigani, Paul Groot, Hubert Halloin, Gerhard Heinzl, Martin Hewitson, Allan Hornstrup, David Hoyland, Mauro Hueller, Philippe Jetzer, Nikolaos Karnesis, Christian Killow, Andrzej Krolak, Ivan Lloro, Davor Mance, Thomas Marsh, Ignacio Mateos, Lucio Mayer, Joseph Moerschell, Gijs Nelemans, Miquel Nofrarias, Frank Ohme, Michael Perreux-Lloyd, Antoine Petiteau, Eric Plagnol, Edward Porter, Pierre Prat, Jens Reiche, David Robertson, Elena Maria Rossi, Stephan Rosswog, Ashley Rüter, B.S. Sathyaprakash, Bernard Schutz, Alberto Sesana, Benjamin Sheard, Ruggero Stanga, Tim Sumner, Takamitsu Tanaka, Michael Tröbs, Hai-Bo Tu, Daniele Vetrugno, Stefano Vitale, Marta Volonteri, Gudrun Wanner, Henry Ward, Peter Wass, William Joseph Weber, Peter Zweifel

Affiliations can be found in a detailed list of authors at <http://elisascience.org/authors>

CONTRIBUTORS

Heather Audley, John Baker, Simon Barke, Matthew Benacquista, Peter L. Bender, Emanuele Berti, Nils Christopher Brause, Sasha Buchman, Jordan Camp, Massimo Cerdonio, Giacomo Ciani, John Conklin, Neil Cornish, Glenn de Vine, Dan DeBra, Marc Dewi Freitag, Germán Fernández Barranco, Filippo Galeazzi, Antonio Garcia, Oliver Gerberding, Lluís Gesa, Felipe Guzman Cervantes, Zoltan Haiman, Craig Hogan, Daniel Holz, C.D. Hoyle, Scott Hughes, Vicky Kalogera, Mukremil Kilic, William Klipstein, Evgenia Kochkina, Natalia Korsakova, Shane Larson, Maike Lieser, Tyson Littenberg, Jeffrey Livas, Piero Madau, Peiman Maghami, Christoph Mahrtdt, David McClelland, Kirk McKenzie, Sean McWilliams, Stephen Merkowitz, Cole Miller, Shawn Mitryk, Soumya Mohanty, Anneke Monsky, Guido Mueller, Vitali Müller, Daniele Nicolodi, Samaya Nissanke, Kenji Numata, Markus Otto, E. Sterl Phinney, Scott Pollack, Alix Preston, Thomas Prince, Douglas Richstone, Louis Rubbo, Josep Sanjuan, Stephan Schlamminger, Daniel Schütze, Daniel Shaddock, Sweta Shah, Aaron Spector, Robert Spero, Robin Stebbins, Gunnar Stede, Frank Steier, Ke-Xun Sun, Andrew Sutton, David Tanner, Ira Thorpe, Massimo Tinto, Michele Vallisneri, Vinzenz Wand, Yan Wang, Brent Ware, Yinan Yu, Nicolas Yunes

Affiliations can be found in a detailed list of contributors at <http://elisascience.org/contributors>

SUPPORTERS

Among the, roughly, 1000 scientific supporters of the Gravitational Universe science theme, are

GERARDUS 'T HOOFT *Utrecht University (Netherlands)*, **BARRY BARISH** *Caltech (United States)*, **CLAUDE COHEN-TANNOUDJI** *College de France (France)*, **NEIL GEHRELS** *NASA Goddard Space Flight Center (United States)*, **GABRIELA GONZALEZ** *LIGO Scientific Collaboration Spokesperson, LSU (United States)*, **DOUGLAS GOUGH** *Institute of Astronomy, University of Cambridge (United Kingdom)*, **STEPHEN HAWKING** *University of Cambridge, DAMTP (United Kingdom)*, **STEVEN KAHN** *Stanford University/SLAC National Accelerator Laboratory (United States)*, **MARK KASEVICH** *Stanford University, Physics Dept. (United States)*, **MICHAEL KRAMER** *Max-Planck-Institut fuer Radioastronomie (Germany)*, **ABRAHAM LOEB** *Harvard University (United States)*, **PIERO MADAU** *University of California, Santa Cruz (United States)*, **LUCIANO MAIANI** *Università di Roma La Sapienza (Italy)*, **JOHN MATHER** *NASA Goddard Space Flight Center (United States)*, **DAVID MERRITT** *Rochester Institute of Technology (United States)*, **VIATCHESLAV MUKHANOV** *LMU München (Germany)*, **GIORGIO PARISI** *Università di Roma la Sapienza (Italy)*, **STUART SHAPIRO** *University of Illinois at Urbana-Champaign (United States)*, **GEORGE SMOOT** *Université Paris Diderot (France)*, **SAUL TEUKOLSKY** *Cornell University (United States)*, **KIP THORNE** *California Institute of Technology (United States)*, **GABRIELE VENEZIANO** *Collège de France (France)*, **JEAN-YVES VINET** *Virgo Collaboration Spokesperson, OCA Nice (France)*, **RAINER WEISS** *MIT (United States)*, **CLIFFORD WILL** *University of Florida (United States)*, **EDWARD WITTEN** *Institute for Advanced Study, Princeton (United States)*, **ARNOLD WOLFENDALE** *Durham University (United Kingdom)*, and **SHING-TUNG YAU** *Harvard University (United States)*.

A complete list of supporters can be found at <http://elisascience.org/supporters>

INTRODUCTION

In the early years of this millennium, our view of the Universe has been comfortably consolidated in some aspects, but also profoundly changed in others. In 2003, the double pulsar PSR J0737-3039 was discovered [1–2], and General Relativity passed all of the most stringent precision tests in the weak-field limit.

In 2013, the toughest test on General Relativity was performed through the observation of PSR J0348+0432, a tightly-orbiting pair of a newly discovered pulsar and its white-dwarf companion. Given the extreme conditions of this system, some scientists thought that Einstein's equations might not accurately predict the amount of gravitational radiation emitted, but General Relativity passed with flying colours [3]. However, no test of General Relativity could be considered complete without probing the strong-field regime of gravitational physics, where mass motions are close to the speed of light, c , and gravitational potentials close to c^2 : Around a black hole, a central singularity protected by an event horizon, relativistic gravity is extreme. Exploring the physics of the inspiral of a small compact object skimming the event horizon of a large black hole, or the physics of a black hole-black hole collision, is probing gravity in the relativistic strong-field limit.

2013 marked an important date: ESA's Planck mission confirmed the Λ -CDM paradigm of cosmology at an unprecedented level of accuracy, offering the most precise all-sky image of the distribution of dark matter across the entire history of the Universe [4], further confirming that the minuscule quantum fluctuations which formed at the epoch of inflation were able to grow hierarchically from the small to the large scale under the effect of the gravity of dark matter, and eventually evolved into the galaxies we observe today with their billions of stars and central black holes [5]. In the Λ -CDM model, the first black holes, called seeds, formed in dark matter halos from the dissipative collapse of baryons [6–7], and as halos clustered and merged, so did their embedded black holes [8–9]. As a consequence, binary black holes invariably form, driven by galaxy collisions and mergers, and trace the evolution of cosmic structures.

In 2003 an ongoing merger between two hard X-ray galactic nuclei, with the characteristics of an Active Galactic Nucleus (AGN), was discovered just 150 Mpc away in the ultra-luminous infrared galaxy NGC 6240 (see Figure 1) [10]. Recent optical surveys have shown evidence of dual AGN [11] in today's Universe; even Andromeda and our Milky Way are due to collide and both house a central black hole!

Galaxy mergers were even more frequent in the past. The ensuing coalescence of massive black holes offers a new and unique tool, not only to test theories of gravity and the black hole hypothesis itself, but to explore the Universe from the onset of the cosmic dawn to the present.

In 2000, we discovered that dormant black holes are ubiq-

uitous in nearby galaxies and that there are relationships between the black hole mass and the stellar mass of the host galaxy [12–13]. This gave rise to the concept that black holes and galaxies evolve jointly. Black holes trace galaxies and affect their evolution; likewise, galaxies trace black holes and affect their growth. Coalescing binary black holes pinpoint the places and times where galaxies merge, revealing physical details of their aggregation.

This new millennium also witnessed the discovery of numerous ultra-compact binary systems containing white dwarfs and/or neutron stars in close orbit, in the Milky Way [14]. These are excellent laboratories for exploring the extremes of stellar evolution in binary systems. They will transform into Type Ia supernovae or into merging binaries which will soon be detected by the ground-based gravitational wave detectors LIGO and VIRGO.

Stellar mass black holes with a pulsar as companion remain elusive, as they are very rare systems. Tracing these almost dark ultra-compact binaries of all flavours with gravitational waves all over the Milky Way will reveal how binary stars formed and evolved in the disc and halo of the Galaxy. Some of these are already known through electromagnetic observations and serve as guaranteed verification sources.

All of these advances were possible using only our first 'sense' for observing the Universe, electromagnetic radiation, tracing electromagnetic interactions of baryonic matter in the Universe. However, almost all of the Universe remains electromagnetically dark. On astronomical scales gravitation is the real engine of the Universe. By 'listening' to gravity we will be able to see further than ever before.

We can 'listen' to the Universe by directly observing gravitational waves, ripples in the fabric of spacetime travelling at the speed of light, which only weakly interact with matter and travel largely undisturbed over cosmological distances. Their signature is a fractional squeezing of spacetime perpendicular to the direction of propagation, with an amplitude $h = \Delta L/L$ on the order of 10^{-20} . Laser interferometry is a standard tool for such measurements and has been under development for over 30 years now [15].

Electromagnetic observations of the Universe, plus theoretical modelling, suggest that the richest part of the gravitational wave spectrum falls into the frequency range accessible to a space interferometer, from about 0.1 mHz to 100 mHz. In this band, important first-hand information can be gathered to tell us how binary stars formed in our Milky Way, and to test the history of the Universe out to redshifts of order $z \sim 20$, probing gravity in the dynamical strong-field regime and on the TeV energy scale of the early Universe [16]. *eLISA* will be the first ever mission to survey the entire Universe with gravitational waves, addressing the science theme *The Gravitational Universe*. The Next Gravitational Observatory (NGO) mission concept studied by ESA for the L1 mission selection is used as a strawman mission concept for *eLISA* [15]. ■

*In Sections I, II, and III, we will describe the *eLISA* observational capabilities in terms of an observatory sensitivity that is shown in Figure 12 as a U-shaped curve. The curve is based upon the sensitivity model of the strawman mission concept presented in Section IV. For the purpose of the discussion in Sections I, II, and III, this is just a baseline requirement. The question of how this requirement can be met, or exceeded, in an actual mission is addressed in Section IV.*

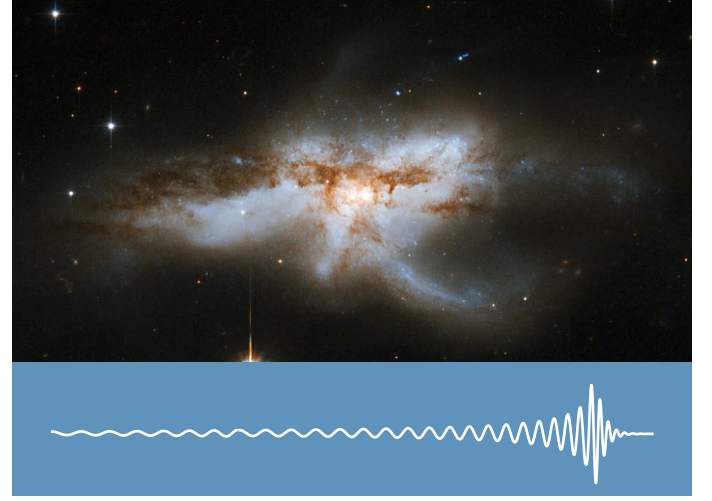


Figure 1: Merging galaxy NGC 6240 and a representative waveform of the expected gravitational waves from the coalescence of two supermassive black holes. Observations with NASA's Chandra X-ray observatory have disclosed two giant black holes inside NGC 6240. They will drift toward one another and eventually merge into a larger black hole. *Credit: NASA, ESA, the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration, and A. Evans (University of Virginia, Charlottesville/NRAO/Stony Brook University).*

I. ASTROPHYSICAL BLACK HOLES

As we will discuss in more detail in this section, *eLISA* observations will probe massive black holes over a wide, almost unexplored, range of redshift and mass, covering essentially all important epochs of their evolutionary history. *eLISA* probes are coalescing massive binary black holes, which are among the loudest sources of gravitational waves in the Universe. They are expected to appear at the ‘cosmic dawn’, around a redshift of $z \sim 11$ or more, when the first galaxies started to form. Coalescing binary black holes at a redshift as remote as $z \sim 20$ can be detected with *eLISA*, if they exist. *eLISA* will also explore black holes through ‘cosmic high noon’ (a term introduced by [17]), at redshifts of $z \sim 3$ to $z \sim 1.5$, when the star formation rate in the Universe and the activity of Quasi Stellar Objects (QSOs) and AGNs was highest. In the ‘late cosmos’, at $z < 1$, *eLISA* will continue to trace binary mergers, but it will also detect new sources, the Extreme Mass Ratio Inspirals (EMRIs), i.e., the slow inspiral and merger of stellar mass black holes into large, massive black holes at the centres of galaxies. These are excellent probes for investigating galactic nuclei during the AGN decline.

I.1 Massive binary black holes

The exploration of the sky across the electromagnetic spectrum has progressively revealed the Universe at the time of the cosmic dawn. The most distant star collapsing into a stellar mass black hole, the Gamma Ray Burst GRB 090429B, exploded when the Universe was 520 Myr old (at a redshift of $z \sim 9.4$), confirming that massive stars were born and died very early on in the life of the Universe [18]. The most distant known galaxy, MACS0647-JD, at a redshift of $z \sim 10.7$, was already in place when the Universe was about 420 Myr old [19], and ULAS J1120+0641 holds the record for being the most distant known QSO, thus the most distant supermassive ($\sim 10^9 M_\odot$) accreting black hole at redshift $z \sim 7.08$, about 770 Myr after the big bang [20]. Such observations clearly show that stars, black holes, and galaxies, the key, ubiquitous components of the Universe, were present before the end of the reionisation phase around $z \sim 6$ [21]. These are the brightest sources, probing only the peak of an underlying distribution of smaller objects: the less luminous pre-galactic discs, and the less massive stars and black holes, of which little is known. Even the brightest QSOs fade away in the optical regime

due to the presence of neutral hydrogen in the intergalactic medium (the Gunn-Peterson trough, [22]), and the search for the deepest sources using X-rays may be hindered by intrinsic obscuration, confusion due to crowding and the unresolved background light [23].

The entire zoo of objects, which in the past formed the small building blocks of the largest ones we see today, is so far pretty much unexplored. These primitive objects started to form at the onset of the cosmic dawn, around $z \sim 20 - 30$, according to current cosmological models [24]. In fact, simulations indicate that the very early pre-galactic, gas-rich discs had low masses, small luminosities and were very metal-poor. At an epoch of $z \sim 20$ to 30, the earliest stars may have had masses exceeding $100 M_\odot$, ending their lives as comparable stellar mass black holes, providing the seeds that would later grow into supermassive black holes [6, 25]. However, as larger, more massive and metal enriched galactic discs progressively formed, other paths for black hole seed formation became viable (see [26] for a review). Global gravitational instabilities in gaseous discs may have led to the formation of quasi-stars of $10^3 - 10^4 M_\odot$ that later collapsed into seed black holes [7]. Further alternatives arise in the form of the collapse of massive stars formed in run-away stellar collisions in young, dense star clusters [27] or the collapse of unstable self-gravitating gas clouds in the nuclei of gas-rich galaxy mergers at later epochs [28]. Thus, the initial mass of seed black holes remains one of the largest uncertainties in the present theory of black hole formation, as the mechanism is still unknown, and the electromagnetic horizon too small for the direct detection of the formation of individual seeds.

Most of the investigations of galaxies and AGNs in the electromagnetic Universe, in terms of richness of sources, focus on a later epoch: the cosmic high noon, a period around $z \sim 1.5 - 3$. This epoch features several critical

transformations in galaxy evolution, since around that time both the luminous QSOs and the star formation rate were at their peak [29–30]. Galaxy mergers and accretion along filaments during cosmic high noon were likely to be the driving force behind the processes of star formation, black hole fueling, and galaxy growth. This turned star-forming discs into larger discs or quenched spheroidal systems hosting supermassive black holes of billions of solar masses [31–33]. In this framework, massive black hole binaries inevitably form in large numbers, over a variety of mass scales, driven by frequent galaxy mergers [8–9, 34]. Signs of galaxy mergers with dual black holes at wide separations (on the order of kpc) come from observations of dual AGNs in optical and X-ray surveys, while observations of binary black holes with sub-pc scale separations remain uncertain and only candidates exist at present [35]. Studies of the dynamics of black holes in merging galaxies have shown that black hole coalescences trace the merger of the dense baryonic cores better than the mergers of dark halos, as their dynamics are sensitive to gas and star content and feed-back [36–37].

Theoretical models developed in the context of the Λ -CDM paradigm [38–41] have been successful in reproducing properties of the observed evolution of galaxies and AGNs, such as the colour distribution of galaxies, the local mass density and mass function of supermassive black holes, and the QSO luminosity function at several wavelengths out to $z \sim 6$. Information about the underlying population of inactive, less massive and intrinsically fainter black holes, which grew through accretion and mergers across all cosmic epochs, is still lacking and difficult to gather.

The Gravitational Universe proposes a unique, new way to probe both cosmic dawn and high noon, to address a number of unanswered questions:

- When did the first black holes form in pre-galactic halos, and what is their initial mass and spin?
- What is the mechanism of black hole formation in galactic nuclei, and how do black holes evolve over cosmic time due to accretion and mergers?
- What is the role of black hole mergers in galaxy formation?

eLISA will study the evolution of merging massive black holes across cosmic ages, measuring their mass, spin and redshift over a wide, as yet unexplored, range. Black holes with masses between $10^4 M_\odot$ and $10^7 M_\odot$ will be detected by *eLISA*, exploring for the first time the low-mass end of the massive black hole population, at cosmic times as early as $z \sim 10$, and beyond.

eLISA discovery domain

Coalescing black hole binaries enter the *eLISA* sensitivity band from the low frequency end, sweeping to higher frequencies as the inspiral gets faster and faster, as shown in Figure 13. Eventually they merge, with the formation

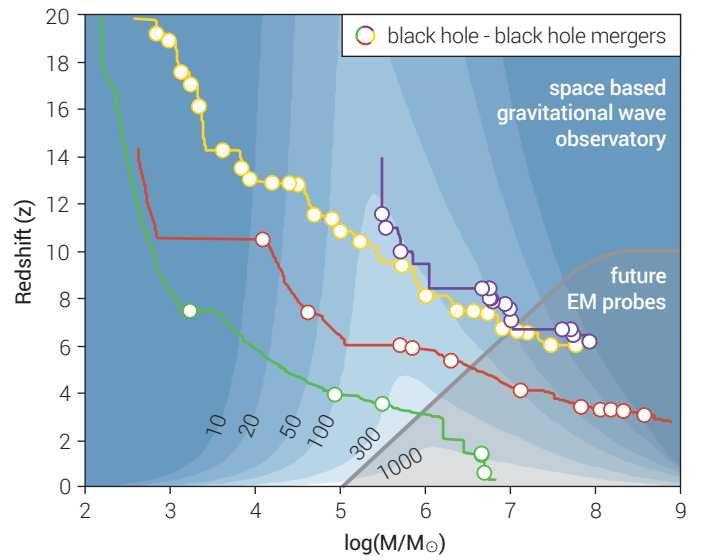


Figure 2: Constant-contour levels of the sky and polarisation angle-averaged SNR for *eLISA*, for equal mass non-spinning binaries as a function of their total rest frame mass, M , and cosmological redshift, z . The tracks represent the mass-redshift evolution of selected supermassive black holes: two possible evolutionary paths for a black hole powering a $z \sim 6$ QSO (starting from a massive seed, blue curve, or from a Pop III seed from a collapsed metal-free star, yellow curve); a typical $10^9 M_\odot$ black hole in a giant elliptical galaxy (red curve); and a Milky Way-like black hole (green curve). Circles mark black hole-black hole mergers occurring along the way. These were obtained using state of the art semi-analytical merger tree models [65]. The grey transparent area in the bottom right corner roughly identifies the parameter space for which massive black holes might power phenomena that will likely be observable by future electromagnetic probes.

of a common event horizon, followed by the ringdown phase during which residual deformation is radiated away and a rotating (Kerr) black hole remnant is formed. The waveform detected by *eLISA* is a measure of the amplitude of the strain in space as a function of time in the rest frame of the detector. This waveform carries information about the masses and spins of the two black holes prior to coalescence, the inclination of the binary plane relative to the line of sight, the luminosity distance and sky location, among other parameters [42]. Complete waveforms have been designed by combining Post Newtonian expansion waveforms for the early inspiral phase with an analytical description of the merger and ringdown phase, calibrated against highly accurate, fully general relativistic numerical simulations of black hole coalescence [43–44]. The first figure of merit of the *eLISA* performance is the signal-to-noise ratio (SNR) of a massive black hole binary coalescence with parameters in the relevant astrophysical range. Figure 2 shows *eLISA* SNRs for equal mass, non-spinning coalescing binaries. Here we compute the SNR as a function of the total mass, M , and of the redshift, z , averaging over all possible source sky locations and wave polarisations, assuming two-year observations. The plot highlights the extraordinary capabilities of the instrument in covering almost all of the mass-redshift parameter space needed to trace black hole evolution. Binaries with $10^4 M_\odot < M < 10^7 M_\odot$ can be detected out to $z \sim 20$ with an $\text{SNR} \geq 10$, if they exist. Figure 2 shows that virtually all massive black holes in the Universe were loud *eLISA* sources at some point in their evolution.

Figure 3 shows error distributions in the source parameter estimation for events collected and extracted from a meta-catalogue of ~ 1500 simulated sources. The catalogue is constructed by combining predicted merger distributions from a number of cosmological models encompassing a broad range of plausible massive black hole evolution scenarios [45]. Uncertainties are evaluated using the Fisher Information Matrix approximation, which gives an estimate of the errors on the inferred parameters. Figure 3 illustrates that individual redshifted masses can be measured with unprecedented precision, with an error of $0.1\% - 1\%$ (we recall that observations can only determine the redshifted source masses, i.e., the product of mass and $(1+z)$). The spin of the primary black hole can be measured with very high accuracy, with $0.01 - 0.1$ absolute uncertainty.

Current theoretical models predict coalescence rates in the range $10 - 100$ per year [46–48]. For more than 10% of these, mostly occurring at a redshift of $z < 5$, the distance can be determined to better than a few percent and the sky location determined to better than a few degrees, which makes these sources suitable targets for coincident searches of electromagnetic counterparts (see Section V).

Astrophysical impact

eLISA will allow us to survey the vast majority of all coalescing massive black hole binaries throughout the whole Universe. This will expose an unseen population of objects which will potentially carry precious information about the entire black hole population. It will provide both the widest and deepest survey of the sky ever, since gravitational wave detectors are essentially omni-directional by nature, and thus act as full-sky monitors. As highlighted in Figure 2, the range of black hole redshifts and masses that will be explored is complementary to the space explored by

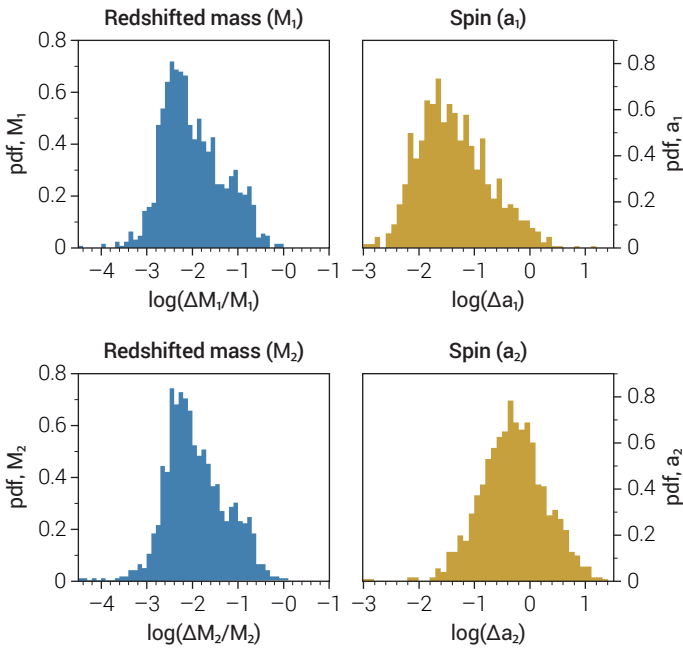


Figure 3: *eLISA* parameter estimation accuracy for massive black hole binaries – probability density functions. Left panels show errors on the redshifted mass, and right panels on the spins. Top panels refer to the primary black hole, bottom panels to the secondary black hole.

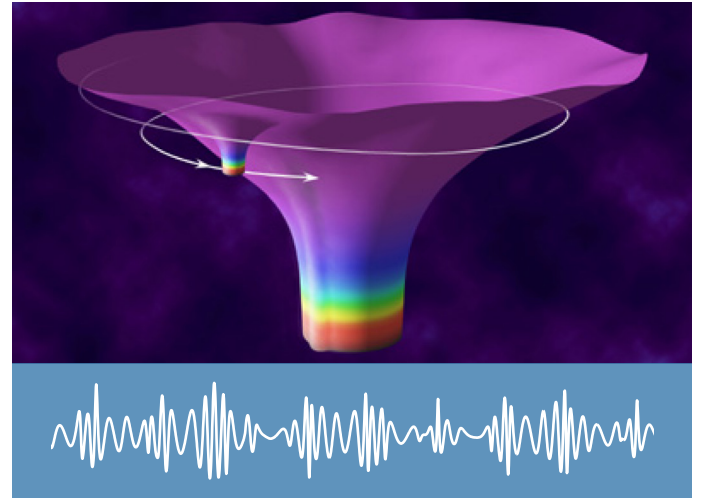


Figure 4: An artist's impression of the spacetime of an extreme-mass-ratio inspiral and a representative waveform of the expected gravitational waves. A smaller black hole orbits around a supermassive black hole. Credit: NASA.

electromagnetic observations (see Figure 2).

eLISA will create the first catalogue of merging black holes, which will enable us to investigate the link between the growing seed population and the rich population of active supermassive black holes evolving during cosmic dawn and high noon. In doing this, we will probe the light end of the mass function at the largest redshifts and investigate the role of early black holes in cosmic re-ionisation and the heating of the intergalactic medium [49].

Black hole coalescence events will illuminate the physical process of black hole feeding. While the mass distribution carries information about the seeds, the spin distribution charts the properties of the accretion flows, whether they are chaotic or coherent [50]. Gravitational wave observations alone will be able to distinguish between the different evolution scenarios [46].

1.11 Extreme Mass Ratio Inspirals (EMRIs)

The present Universe is in a phase in which both the star formation rate and AGN activity are declining. In this late cosmos we observe quiescent massive black holes at the centres of galaxies within a volume of about 0.02 Gpc^3 [51]. The current census comprises about 75 massive black holes out to $z < 0.03$. The black hole of $4 \times 10^6 M_\odot$ at the Galactic Centre is the most spectacular example [52–53]. Thanks to its proximity, a young stellar population has been revealed precisely where no young stars were predicted to form, as star-forming clouds are expected to be tidally disrupted there [54]. This indicates our lack of understanding about the origin of stellar populations around black holes, and in particular stellar dynamics, even in our own Galaxy. By probing the dynamics of intrinsically dark, relic stars in the nearest environs of a massive black hole, *eLISA* will offer the deepest view of galactic nuclei, exploring phenomena inaccessible to electromagnetic observations [55–56]. The probes used are the so-called EMRIs: a compact star (either a neutron star or a stellar mass black hole) captured

in a highly relativistic orbit around the massive black hole and spiralling through the strongest field regions a few Schwarzschild radii from the event horizon before plunging into it (Figure 4).

Key questions can be addressed in the study of galactic nuclei with EMRIs:

- What is the mass distribution of stellar remnants at the galactic centres and what is the role of mass segregation and relaxation in determining the nature of the stellar populations around the nuclear black holes in galaxies?
- Are massive black holes as light as $\sim 10^5 M_\odot$ inhabiting the cores of low mass galaxies? Are they seed black hole relics? What are their properties?

eLISA will observe EMRI events, exploring the deepest regions of galactic nuclei, those near the horizons of black holes with masses close to the mass of the black hole at our Galactic Centre, out to redshifts as large as $z \sim 0.7$.

Stellar mass black holes are expected to dominate the observed EMRI rate for *eLISA*, as mass segregation by two-body relaxation tends to concentrate the heaviest stars near the central, massive black hole [16, 57–58], and a stellar mass black hole inspiral has a higher SNR, so it can be detected out to farther distances. EMRIs can be tracked around a central black hole for up to $10^4 - 10^5$ cycles on complex relativistic orbits (see Figure 5). As a consequence, the waveform carries an enormous amount

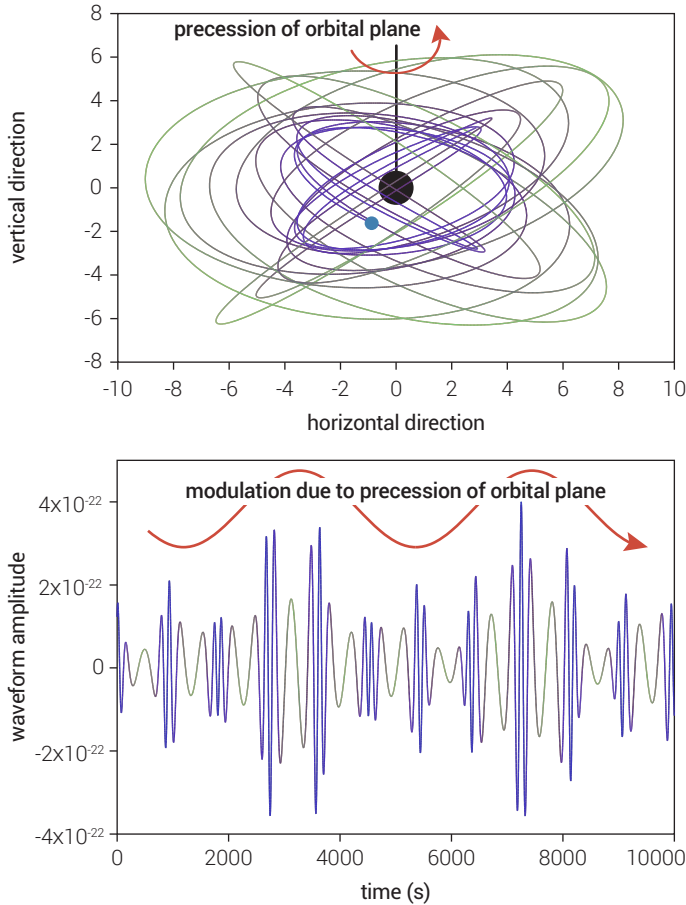


Figure 5: EMRI orbit and signal. In the top panel we see the geometrical shape of the ornate relativistic EMRI orbit. The lower panel shows the corresponding gravitational wave amplitude as a function of time.

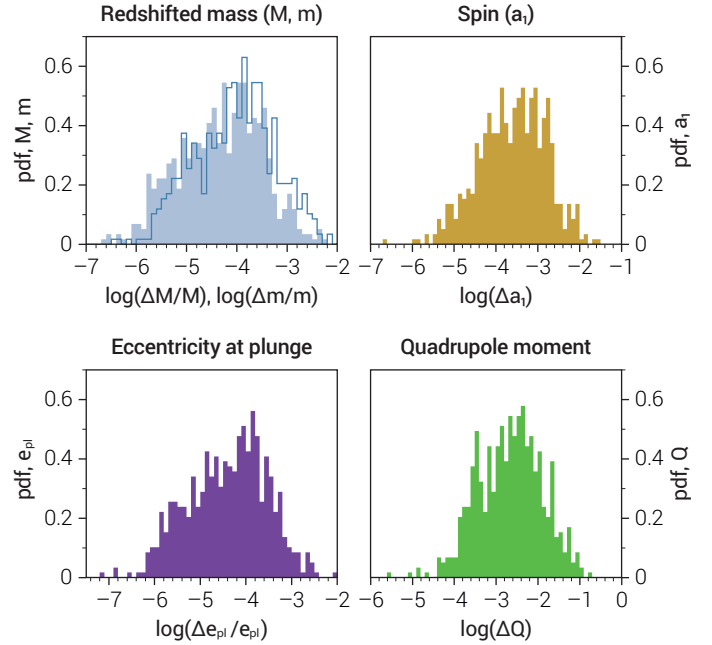


Figure 6: eLISA parameter estimation accuracy for EMRIs. Top left panel: estimation of the redshifted masses (filled: massive black hole; solid line: stellar black hole); top right panel: spin of the massive black hole; bottom left panel: eccentricity at plunge; bottom right panel: minimum measurable deviation of the quadrupole moment of the massive black hole from the Kerr value.

of information [59]. Through observations of dark components alone, *eLISA* will detect EMRIs with an $\text{SNR} > 20$ in the mass interval for the central black hole between $10^4 M_\odot < M < 5 \times 10^6 M_\odot$ out to redshift $z \sim 0.7$ (see Figure 13), covering a co-moving volume of 70 Gpc^3 , a much larger volume than observations of dormant galactic nuclei today. The estimated detection rates, based on the best available models of the black hole population and of the EMRI rate per individual galaxy [60], are about 50 events for a 2 year mission, with a factor of 2 uncertainty from the waveform modelling and lack of knowledge about the system parameters, and an additional uncertainty of at least an order of magnitude stemming from the uncertain dynamics of dense stellar nuclei [61–62]. As shown in Figure 6, the masses of both black holes are, in most cases, measured to better than one part in 10^4 , the eccentricity at plunge is determined to a 10^{-4} accuracy, and the spin of the primary black hole to better than 10^{-3} . The deviation of the quadrupole moment of the massive black hole with respect to the Kerr metric value is determined to better than 0.01, enabling unprecedented tests of General Relativity to be performed (see Section II).

Astrophysical impact

In *The Gravitational Universe*, EMRIs are exquisite probes for testing stellar mass black hole populations in galactic nuclei. With *eLISA* we will learn about the mass spectrum of stellar mass black holes, which is largely unconstrained both theoretically and observationally. The measurement of even a few EMRIs will give astrophysicists a totally new way of probing dense stellar systems, allowing us to determine the mechanisms that govern stellar dynamics in the galactic nuclei [58].

Measurements of a handful of events will suffice to constrain the low end of the massive black hole mass function, in an interval of masses where electromagnetic observations are poor, incomplete or even missing [63]. By 2028 the Large Synoptic Survey Telescope (LSST) will have observed a large number of tidal disruption events [64], which will also teach us a lot about black holes and stellar populations in galactic centres. However, these events will typically be on the higher end of the black hole's mass function, and will not reveal masses with the same precision as *eLISA*, whose observations will give us information about the nature and the occupation fraction of massive black holes in low mass galaxies (as yet unconstrained) [65]. This will provide additional information about the origin of massive black holes, complementary to that gathered via observations of high z massive black hole mergers. EMRIs will also provide the most precise measurements of Milky Way type massive black hole spins, and make it possible for us to investigate the spin distribution of single, massive black holes up to $z \sim 0.7$. EMRIs can occur around black holes in all galaxies, regardless of their nature, i.e., whether they are active or not. As such, spin measurements will not be affected by observational uncertainties in the spectra of the AGN.

Detection rates will tell us about the density of stellar mass black holes in the vicinity of the central black hole, constraining the effectiveness of the mass segregation processes [66]. The measured eccentricity and orbital inclination distributions can be linked directly to the preferred channel of EMRI formation, giving us important clues about the efficiency of dynamical relaxation, and the frequency of binary formation and breakup in dense nuclei [67]. ■

II. THE LAWS OF NATURE

This section covers tests of strong gravity and cosmology, two regimes where *The Gravitational Universe* will offer major advances.

II.1 High precision measurements of strong gravity

Einstein's General Relativity is one of the pillars of modern cosmology. The beauty of General Relativity is that it is a falsifiable theory: once the underlying mass distribution is identified to be a black hole binary system with fixed masses and spins, the theory has no further adjustable parameters. Thus even a single experiment incompatible with a prediction of the theory would lead to its invalidation, at least in the physical regime of applicability of the experiment.

The Gravitational Universe will explore relativistic gravity in the strong-field, non-linear regime. It seems unlikely that any other methods will achieve the sensitivity of *eLISA* to deviations of strong-field gravity by 2028 (see Section V). Unlike the ground-based instruments, *eLISA* will have

sufficient sensitivity to observe even small corrections to Einstein gravity.

The strong-field realm of gravity theories can be probed near the event horizon of Kerr black holes or in other large-curvature environments (e.g., in the early Universe). Gravity can be thought of as strong in the sense that gravitational potentials are a significant fraction of c^2 or in the sense that the curvature tensor (or tidal force) is of very large magnitude. Testing strong gravity takes on different meanings depending upon which notion of 'strong gravity' is being used. In any case, both are very important to test with high precision. The measurement of the anisotropy of the Cosmic Microwave Background (CMB) by ESA's Planck satellite directly constrains properties of the quantum fluctuations during inflation which are thought to be the seeds of structure formation. This connection between physics on the smallest and largest scales is strong motivation for testing General Relativity to the highest possible accuracy, and in particular in the strong-field regime, where deviations could hint at a formulation of a quantum theory of gravity.

The nature of gravity in the strong-field limit is, so far, largely unconstrained, leaving open several questions:

- Does gravity travel at the speed of light ?
- Does the graviton have mass?
- How does gravitational information propagate: Are there more than two transverse modes of propagation?
- Does gravity couple to other dynamical fields, such as, massless or massive scalars?
- What is the structure of spacetime just outside astrophysical black holes? Do their spacetimes have horizons?
- Are astrophysical black holes fully described by the Kerr metric, as predicted by General Relativity?

An outstanding way to answer these questions and learn about the fundamental nature of gravity is by observing the vibrations of the fabric of spacetime itself, for which coalescing binary black holes and EMRIs are ideal probes.

Exploring relativistic gravity with binary black hole mergers in the strong-field, dynamical sector

The coalescence of a massive black hole binary with mass ratio above one tenth generates a gravitational wave signal strong enough to allow detection of tiny deviations from the predictions of General Relativity. The signal comprises three parts—inspiral, merger and ringdown—each of which probes strong-field gravity. The inspiral phase is well understood theoretically: It can last several months in-band, and it could be observed with an SNR of tens to hundreds by *eLISA*. The non-linear structure of General Relativity, and possible deviations from it, are encoded in the phase and amplitude of the gravitational waves. Any effect that leads to a cumulative dephasing of a significant fraction of a wave cycle over the inspiral phase can be de-

tected through matched filtering. Thousands of inspiral wave cycles will be observed, making it possible to detect even very small deviations in the inspiral rate predicted by General Relativity [68–71]. The propagation of gravitational waves can be probed through the dispersion of the inspiral signal. In General Relativity, gravitational waves travel with the speed of light (the graviton is massless) and interact very weakly with matter. Alternative theories with a massive graviton predict an additional frequency dependent phase shift of the observed waveform due to dispersion that depends on the graviton’s mass, m_g , and the distance to the binary. An *eLISA*-like detector could set a bound around $m_g < 4 \times 10^{-30}$ eV [72], improving current Solar System bounds on the graviton mass, $m_g < 4 \times 10^{-22}$ eV, by several orders of magnitude. Statistical analysis of an ensemble of observations of black hole coalescences could also be used to place stringent constraints on theories with an evolving gravitational constant [73] and theories with Lorentz-violating modifications to General Relativity [74].

The inspiral is followed by a dynamical coalescence that produces a burst of gravitational waves. This is a brief event, comprising a few cycles and lasting about $5 \times 10^3 \text{ sec} (M/10^6 M_\odot)$, yet it is very energetic, releasing $10^{59} (M/10^6 M_\odot)$ ergs of energy, corresponding to 10^{22} times the power of the Sun.

After the merger, the asymmetric remnant black hole settles down to a stationary and axisymmetric state through the emission of quasi-normal mode (QNM) radiation. In General Relativity, astrophysical black holes are expected to be described by the Kerr metric and characterised by only two parameters: mass and spin (the ‘no-hair’ theorem). Each QNM is an exponentially damped sinusoid with a characteristic frequency and damping time that depends only on these two parameters [75–77]. A measurement of two QNMs will therefore provide a strong-field verification that the final massive object is consistent with being a Kerr black hole [78–79]. The QNM spectrum of a black hole also has unique features which allow it to be distinguished from other (exotic) compact objects [80–83]. *eLISA* will observe ringdown signals with sufficient SNR to carry out these tests [84–86].

Exploring relativistic gravity with EMRIs in the strong-field, stationary, non-linear regime

EMRIs will provide a precise tool to probe the structure of spacetime surrounding massive black holes. The inspiralling compact object can generate hundreds of thousands of gravitational wave cycles while it is within ten Schwarzschild radii of the central black hole. These waveform cycles trace the orbit that the object follows, which in turn maps out details of the underlying spacetime structure, in a way similar to how stellar orbits have been used to precisely characterise the supermassive black hole at the centre of the Milky Way [87–88].

As seen in Figure 6, EMRI observations will not only pro-

vide very precise measurements of the ‘standard’ parameters of the system, but will provide strong constraints on departures of the central massive object from the Kerr black hole of General Relativity [89]. The no-hair theorem tells us that the stationary axisymmetric spacetime around it should be completely determined by its mass and spin parameter. The gravitational wave signal from an EMRI occurring in a ‘bumpy’ black hole spacetime in which the multipole moments differ from their Kerr values would show distinctive, detectable signatures [89–94]. Figure 6 shows that 10 % deviations in the mass quadrupole moment, Q , from the Kerr value would be detectable for any EMRI observed with an SNR greater than 20. For typical systems, 0.1 % deviations will be detectable, and for the best systems, 0.01 % deviations will be detectable [59].

An observed inconsistency with the Kerr multipole structure might indicate a surprisingly strong environmental perturbation, the discovery of a new type of exotic compact object consistent with General Relativity, or a failure in General Relativity itself, but these possibilities will be observationally distinguishable (for a review of different hypotheses see [95–97]). These deviations could exhibit themselves in the following ways: For a boson star, the EMRI signal would not shut off after the last stable orbit [98]. For horizonless objects such as gravastars, the orbiting body would resonantly excite the modes of the (putative) membrane replacing the black hole horizon [99], and for certain non-Kerr axisymmetric geometries, orbits could become ergodic [100] or experience extended resonances [101].

Alternatives that will be testable with *eLISA* observations include the dynamical Chern-Simons theory [102–105], scalar-tensor theories (with observable effects in neutron star-black hole systems where the Neutron Star (NS) carries scalar charge [106]), Randall-Sundrum-inspired braneworld models [107–108] and theories with axions that give rise to ‘floating orbits’ [109]. Generic alternatives could also be constrained using phenomenologically parametrised models [110].

Cosmography

The Gravitational Universe will use black hole binary mergers as ‘standard sirens’ to extract information on the expansion of the Universe, by measuring the expansion history with completely different techniques to electromagnetic probes. The term standard siren for gravitational wave sources, refers to a source that has its absolute luminosity encoded in its signal shape, analogous to a standard candle (like a Type Ia supernova) for electromagnetic sources. Black hole coalescences could serve as standard sirens for cosmography [111–112] by providing absolute and direct measurements of the luminosity distance, $D_L(z)$. When coupled with independent measurements of redshift, z , (for example, from associated transient electromagnetic sources), these standard siren sources put points on the distance vs. redshift curve, and directly constrain

the evolution history of the Universe. Several mechanisms have been proposed that will provide an electromagnetic counterpart to massive black hole coalescences detectable by *eLISA* [113], however, confident identification might be viable only for low redshift events. Although rare, events at redshift $z \leq 1-2$ are so loud that the baseline *eLISA* mission has the capability of localising them to within 10 square degrees, perhaps even 1 square degree, when information from the late merger of the black holes is included in the measurement model. This pins down these events on the sky well enough to allow searches for electromagnetic counterparts to the merger using wide area surveys such as LSST that will be active in 2028. With an associated counterpart, *eLISA* observations will allow 1 % measurements of $D_L(z)$ for 60 % of the sources, offering the prospect of ultra-precise determination of points on the distance-redshift curve that are completely independent of all existing constraints from Type Ia supernova, the CMB, etc. It is to be emphasised that there is no distance ladder in these measurements, since the luminosity distance is measured directly. This is possible because these sources are fundamentally understood, and General Relativity calibrates the distances. Weaker statistical constraints could also be derived in the absence of electromagnetic counterparts, using mergers at low redshift ($z < 2$) [114] or EMRIs [115–116]. EMRI observations could provide an independent measurement of H_0 to a precision of a few percent.

Impact on science

The Gravitational Universe will permit unprecedented measurements of General Relativity in the strong-field regime. *eLISA* will map the spacetime around astrophysical black holes, yielding a battery of precision tests of General Relativity in an entirely new regime. These have the potential to uncover hints about the nature of quantum gravity, as well as enabling measurements of the properties of the Universe on the largest scales.

II.II Cosmology on the TeV energy scale – a fossil background of gravitational waves

Several processes occurring at very high energies in the primordial Universe can produce a stochastic background of gravitational waves. The detection of this relic radiation would have a profound impact both on cosmology and on high energy physics. Any fossil radiation of gravitational waves, if not washed away by inflation and later phase transitions, would have decoupled from matter and energy at the Planck scale. It can therefore directly probe cosmological epochs before the decoupling of the cosmic microwave background, currently our closest view of the big bang (Figure 7). The characteristic frequency of the gravitational waves is set by the horizon scale and therefore by the temperature of the Universe at the time of production. The *eLISA* frequency band of 0.1 mHz to 100 mHz corresponds to 0.1 to 100 TeV energy scales in the early Universe, at which new physics is expected to become visible. The

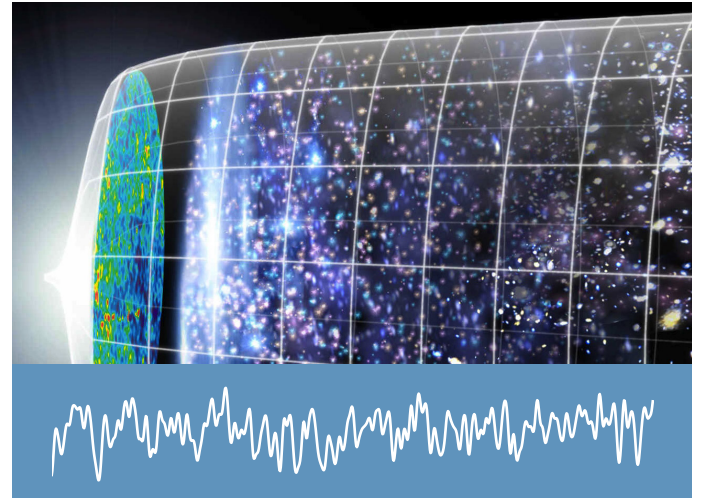


Figure 7: Evolution timeline (big bang and the early Universe) and a typical random waveform of the expected gravitational waves. Gravitational waves are the only way to see beyond the cosmic microwave background. Credit: NASA / WMAP science team.

Large Hadron Collider (LHC) has been built to investigate the physics operating at this energy scale, and in 2012 the experiment produced with the remarkable discovery of the Higgs boson which has completed the particle spectrum of the Standard Model. It is the final confirmation that spontaneous symmetry breaking is the mechanism at play in electroweak physics, and is the first example of a fundamental scalar field playing a role in a phase transition that took place in the very early Universe. These findings further motivate the search for a cosmic background of gravitational waves. *eLISA* would have the sensitivity to detect a relic background created by new physics active at TeV energies if more than a modest fraction, $\Omega_{\text{GW}} \sim 10^{-5}$ of the energy density of the Universe is, converted to gravitational radiation at the time of production.

A gravitational wave detector in space has the potential to revolutionise our understanding of the physics of the infant Universe by exploring the microphysical behaviour of matter and energy through the direct detection of gravitational waves produced at this epoch, rather than by observing collisions of elementary particles.

Discovery space

Abundant evidence suggests that the physical vacuum has not always been in its current state, and in many theories beyond the Standard Model, the conversion between vacuum states corresponds to a first-order phase transition. As the Universe expands and its temperature drops below the critical temperature, bubbles of a new phase form, expand, and collide, generating relativistic bulk flows, whose energy then dissipates in a turbulent cascade. The corresponding acceleration of matter radiates gravitational waves on a scale not far below the horizon scale [117–120]. *eLISA* could detect these gravitational waves, thus probing the Higgs field self couplings and potential, and the possible presence of supersymmetry, or of conformal dynamics at TeV scales. In general, since the Hubble length at the TeV scale is about 1 mm, the current threshold at which

the effects of extra dimensions might appear happens to be about the same for experimental gravity in the laboratory and for the cosmological regime accessible to *eLISA*, thus allowing *eLISA* to probe the dynamics of warped sub-millimetre extra dimensions, present in the context of some string theory scenarios [121–122].

In some braneworld scenarios the Planck scale itself is not far above the TeV scale. Consequently, the reheating temperature would be in the TeV range and *eLISA* could probe inflationary reheating. After inflation the internal potential energy of the inflaton is converted into a thermal mix of relativistic particles, which can generate gravitational waves with an energy of about 10^{-3} or more of the total energy density [123–125]. *eLISA* could also probe gravitational waves produced directly by the amplification of quantum vacuum fluctuations during inflation, in the context of some unconventional inflationary models, such as pre-big bang or bouncing brane scenarios [126–128].

Phase transitions often lead to the formation of one-dimensional topological defects known as cosmic strings. Fundamental strings also arise as objects in string theory and, although formed on submicroscopic scales, it has been realised that these strings could be stretched to astronomical size by cosmic expansion [129–130]. Cosmic strings interact and form loops which decay into gravitational waves; *eLISA* will be the most sensitive probe for these objects, offering the possibility of detecting direct evidence of fundamental strings. The spectrum from cosmic strings is distinguishably different from that of phase transitions or any other predicted source [130]: It has nearly constant energy per logarithmic frequency interval over many decades at high frequencies, offering the possibility of simultaneous detection by *eLISA* and ground-based interferometers. Moreover, if strings are not too light, occasional distinctive gravitational wave bursts might be observed from kinks or cusps on string loops. If detected, these individual bursts will provide irrefutable evidence for a cosmic string source. ■

III. ULTRA-COMPACT BINARIES IN THE MILKY WAY

Only a minority of the stars in the Universe are companionless, the majority being part of binary or multiple star systems (Figure 8). About half of the binaries form with sufficiently small orbital separations to interact and evolve into compact systems, often with white dwarfs, neutron stars, or possibly stellar mass black holes as components. The shortest period systems, known as ultra-compact binaries, are important sources of gravitational waves in the mHz frequency range [131]. These binaries are the outcome of one or more common-envelope phases that occur when one star evolves to the giant or supergiant stage. Unstable mass exchange leads to a short-lived phase in which

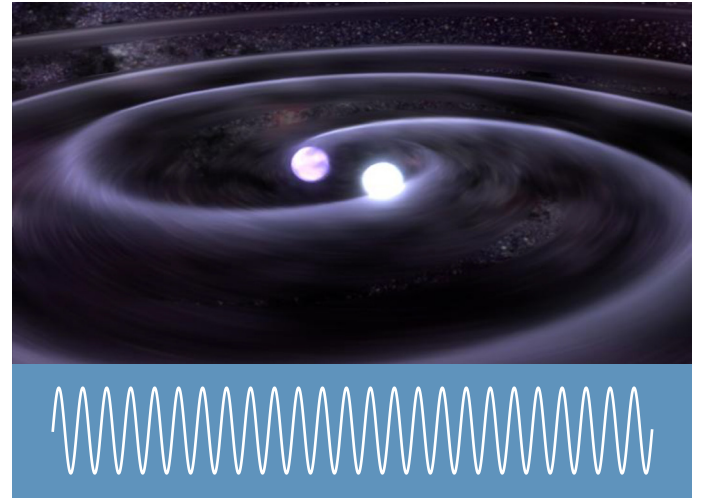


Figure 8: Illustration of a compact binary star system and a representative waveform of the expected gravitational waves. Two stars orbiting each other in a death grip are destined to merge all the while flooding space with gravitational waves. Credit: GSFC/D.Berry.

the stellar core and the companion spiral towards each other, transferring angular momentum to the extended envelope of the giant that is blown away, leaving a tight binary system behind. This a necessary step in the formation of X-ray binaries, binary pulsars and double white dwarf binaries, observed in a variety of states and configurations [132]. Although the Milky Way is full of these sources, only a tiny fraction are currently observed and studied in-depth through observations from radio to X-rays.

After the common-envelope phase, the compact stars in the binary are well separated, but evolve with shorter and shorter orbital periods due to the angular momentum loss via gravitational waves, until eventually the stars undergo another mass exchange once they reach orbital periods of minutes or less. Depending on the nature of the compact objects, they either merge or survive. Double neutron star binaries ultimately merge in bursts of high-frequency gravitational waves, potentially emitting a burst of electromagnetic radiation in the form of a short Gamma-Ray Burst. Although predicted to exist, no neutron star-stellar mass black hole binary or black hole-black hole binary has yet been detected. For binaries with a white dwarf component, mass loss is delicately balanced against loss of angular momentum, the outcome of which is unclear.

Almost always, the loss of energy through gravitational waves is so strong that it causes the system to merge and possibly to end in a Type Ia or sub-luminous supernova explosion [133–134] or in (rapidly spinning) neutron stars that may have millisecond radio pulsar or magnetar properties [135]. In the remaining small fraction of cases, the mass transfer stabilises the system and long-lived interacting binaries are formed (as in the AM CVn systems if the companion is a white dwarf, or ultra-compact X-ray binaries if it is a neutron star). The physics of this evolutionary junction is rich and diverse, it involves tides, mass transfer, highly super-Eddington accretion, and mass ejection. *eLISA* will provide the data necessary for us to quantify the roles that these various physical processes play.

Currently fewer than 50 ultra-compact binaries are known, only two of which have periods less than 10 minutes [14]; *eLISA* will discover several thousand of these. These systems are relatively short-lived and electromagnetically faint, but several known verification binaries are strong enough gravitational wave sources that they will be detected within weeks by *eLISA*. The discovery of many new ultra-compact binaries is one of the main objectives of *eLISA*, as it will provide a quantitative and homogeneous study of their populations and the astrophysics governing their formation. It also adds an additional facet to the knowledge of the Milky Way's structure: The distribution of sources in the thin disc, thick disc, halo and its globular clusters (known breeding grounds for the formation of compact binaries via dynamical exchanges) [136]. These detections will enable us to address a number of key questions:

- *How many ultra-compact binaries exist in the Milky Way?*
- *What is the merger rate of white dwarfs, neutron stars and stellar mass black holes in the Milky Way (thus better constraining the rate of the explosive events associated with these sources)?*
- *What does that imply for, or how does that compare to, their merger rates in the Universe?*
- *What happens at the moment a white dwarf starts mass exchange with another white dwarf or neutron star, and what does it tell us about the explosion mechanism of type Ia supernovae?*
- *What is the spatial distribution of ultra-compact binaries, and what can we learn about the structure of the Milky Way as a whole?*

To answer these questions *eLISA* will observe thousands of individual sources and for the first time capture the signal from a foreground of sources, the *sound* of millions of tight binaries.

Discovery space

The vast majority of ultra-compact binaries will form an unresolved foreground signal in *eLISA* [137] as shown in Figure 13. Its average level is comparable to the instrument noise, but due to its strong modulation during the year (by more than a factor of two) it can be detected. The overall strength can be used to learn about the distribution of the sources in the Galaxy, as the different Galactic components (thin disc, thick disc, halo) contribute differently to the modulation [138]. Their relative amplitudes can be used to set upper limits on the, as yet completely unknown, halo population [139].

For a two-year *eLISA* mission, several thousand binaries are expected to be detected individually with an $\text{SNR} > 7$ [140] and their periods (below one hour and typically 5 – 10 min) determined from the periodicity of the gravitational wave signal. For many systems it will be possible to measure the

first time derivative of the frequency, and thus determine the chirp mass (a combination of the masses of the two stars that can be used to distinguish white dwarf, neutron star and black hole binaries) and the distance to the source. For more than 100 sources concentrated around the inner Galaxy, we expect distance estimates with accuracies better than 1 %, enabling us to make a direct measurement of the distance to the Galactic Centre.

The number of ultra-compact binaries with neutron star or black hole components is still highly uncertain [141]. With *eLISA* operating as an all sky monitor, these systems can be observed throughout the Milky Way, providing a complete sample of binaries at the shortest periods (below 30 minutes), including ones containing stellar mass black holes, if they exist. The number of sources detected at mHz frequencies is directly related, via the gravitational wave orbital decay time scale, both to the number of systems formed at larger separations, and to the number of mergers. Therefore, the thousands of *eLISA* detections will also probe the formation of these binaries and their coalescence rates. Using *eLISA*, the sky position of about fifteen hundred sources will be determined to better than ten square degrees and more than half will have their distance determined to better than 20 % [142]. Several hundred of these may be found with current and future wide-field instruments such as the Visible and Infrared Survey Telescope for Astronomy (VIRCAM) and LSST. A few dozen of the highest SNR sources will have error boxes that are significantly smaller and can be realistically found with small field of view (several arcmin) cameras such as the Multi-adaptive Optics Imaging Camera for Deep Observations (MICADO) on the European Extremely Large Telescope (E-ELT) or the James Webb Space Telescope (JWST).

Astrophysical impact

The large number of ultra-compact binaries discovered, and the fact that the sample is complete at the shortest periods, will help determine the total number of systems of all types as well as their merger rates. The systems containing neutron stars and/or black holes will be observed about a million years prior to coalescence, a phase that is still unexplored.

The highest SNR systems will allow a study of the complex physics of white dwarf mergers or of how systems survive as interacting binaries. Recent detailed simulations [143] have cast doubt on the theory that the actual merger would be a truly dynamical process taking only one or two orbits, and instead show that the merger would take place over many orbits, possibly allowing *eLISA* to observe some mergers directly.

eLISA will detect ultra-compact binaries beyond the Galactic Centre and in the Milky Way's halo using observations which are unaffected by dust obscuration, providing an independent probe of the components and formation history of the Milky Way. ■

IV. A STRAWMAN MISSION FOR THE *eLISA* SPACE GRAVITATIONAL WAVE OBSERVATORY

All of the above scientific objectives can be addressed by a single L-class mission consisting of 3 drag-free spacecraft forming a triangular constellation with arm lengths of one million km and laser interferometry between “free-falling” test masses. The interferometers measure the variations in light travel time along the arms due to the tidal deformation of spacetime by gravitational waves. Compared to the Earth-based gravitational wave observatories like LIGO and VIRGO, *eLISA* addresses the much richer frequency range between 0.1 mHz and 1 Hz, which is inaccessible on Earth due to arm length limitations and terrestrial gravity gradient noise.

The Next Gravitational wave Observatory (NGO) mission studied for the L1 selection [15] is an *eLISA* strawman mission concept. It enables the ambitious science program described here, and has been evaluated by ESA as both technically feasible and compatible with the L2 cost target. Its foundation is mature and solid, based on decades of development for LISA, including a mission formulation study, and the extensive heritage of flight hardware and ground preparation for the upcoming LISA Pathfinder geodesic explorer mission, which will directly test most of the *eLISA* performance and validate the *eLISA* instrumental noise model [144–145].

Mission design

The NGO mission has three spacecraft, one ‘mother’ at the vertex and two ‘daughters’ at the ends, which form a single Michelson interferometer configuration (Figure 9). The spacecraft follow independent heliocentric orbits without any station-keeping and form a nearly equilateral triangle in a plane that is inclined by 60° to the ecliptic. The constellation follows the Earth at a distance between 10° and

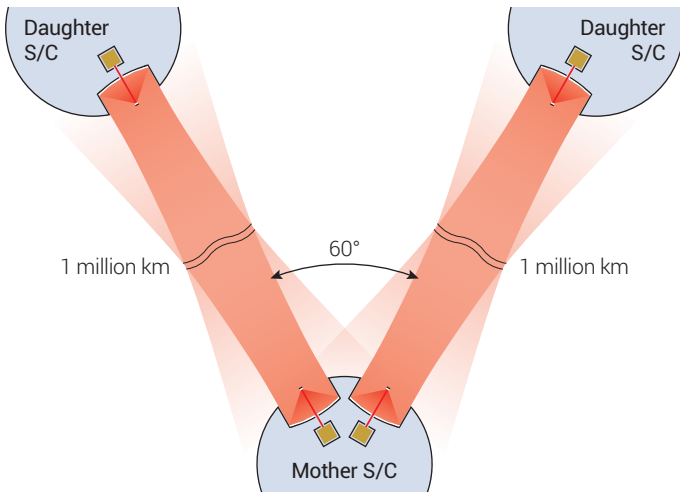


Figure 9: *eLISA* configuration (not to scale). One mother and two daughter spacecraft exchanging laser light form a two-arm Michelson interferometer. There are four identical payloads, one at the end of each arm, as shown in Figure 11.

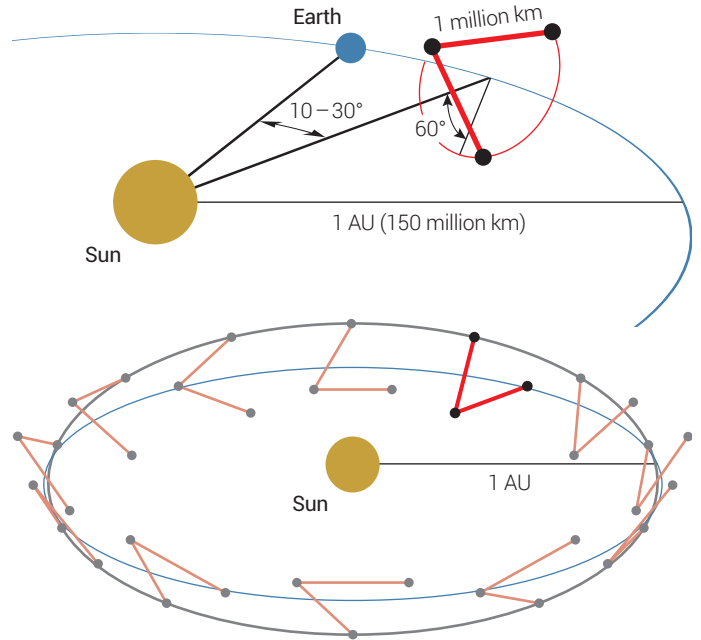


Figure 10: *eLISA* Orbits. The three *eLISA*-NGO spacecraft follow the Earth as an almost stiff triangle, purely due to celestial mechanics.

30°, as shown in Figure 10. Celestial mechanics causes the triangle to rotate almost rigidly about its centre as it orbits around the sun, with variations of arm length and opening angle at the percent level.

The payload consists of four identical units, two on the mother spacecraft and one on each daughter spacecraft (Figure 11). Each unit contains a Gravitational Reference Sensor (GRS) with an embedded free-falling test mass that acts both as the end point of the optical length measurement, and as a geodesic reference test particle. A telescope with 20 cm diameter transmits light from a 2 W laser at 1064 nm along the arm and also receives a small fraction of the light sent from the far spacecraft. Laser interferometry is performed on an optical bench placed between the telescope and the GRS.

On the optical bench, the received light from the distant spacecraft is interfered with the local laser source to pro-

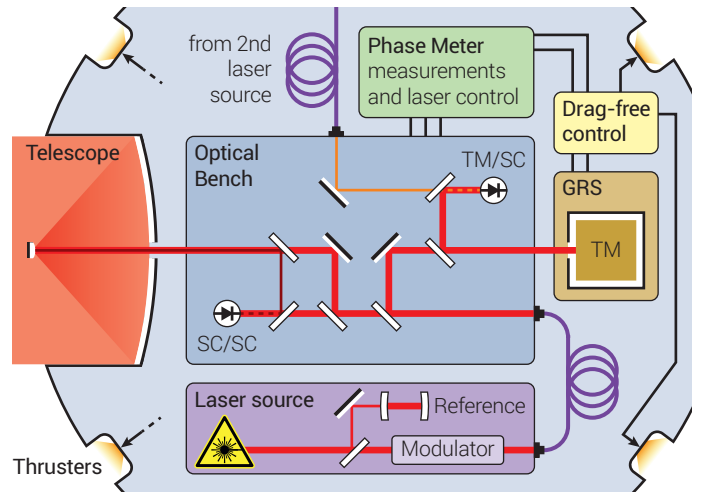


Figure 11: *eLISA* payload. Each payload unit contains a 20 cm telescope, the test mass enclosed inside the Gravitational Reference Sensor (GRS) and an optical bench hosting the interferometers. (Auxiliary reference interferometer omitted for clarity, see [15] for details.)

duce a heterodyne beat note signal between 5 and 25 MHz, which is detected by a quadrant photodiode. The phase of that beat note is measured with $\mu\text{cycle}/\sqrt{\text{Hz}}$ precision by an electronic phasemeter. Its time evolution reflects the laser light Doppler shift from the relative motion of the spacecraft, and contains both the macroscopic arm length variations on times cales of months to years, and the small fluctuations with periods between seconds and hours that represent the gravitational wave science signal. The measurement of relative spacecraft motion is then summed with a similar local interferometer measurement of the displacement between test mass and spacecraft. This yields the desired science measurement between distant free-falling test masses, removing the much larger motion of the spacecraft, which contains both thruster and solar radiation pressure noise.

Drag-free control

The spacecraft are actively controlled to remain centred on the test masses along the interferometric axes, without applying forces on the test masses along these axes. This ‘drag-free control’ around the shielded geodesic reference test masses uses the local interferometry measurement as a control signal for an array of micro-Newton spacecraft thrusters, with the residual spacecraft jitter reaching the $\text{nm}/\sqrt{\text{Hz}}$ level. These thrusters also control the spacecraft angular alignment to the distant spacecraft by detecting the laser beam wavefront with ‘differential wavefront sensing’ with $\text{nrad}/\sqrt{\text{Hz}}$ precision. Other degrees of freedom are controlled with electrostatic test mass suspensions. The only remaining degree of freedom is then the opening angle between the arms at the master spacecraft, which varies smoothly by roughly 1.5° over the year, and can be compensated for either by moving the two optical assemblies against each other or by a steering mirror on the optical bench.

The test masses are 46 mm cubes, made from a dense non-magnetic Au-Pt alloy and shielded by the GRS. The GRS core is a housing of electrodes, at several mm separation from the test mass, used for $\text{nm}/\sqrt{\text{Hz}}$ precision capacitive sensing and nN-level electrostatic force actuation on all non-interferometric degrees of freedom. The GRS also includes fibres for UV light injection for photoelectric discharge of the test mass, and a caging mechanism for protecting the test mass during launch and then releasing it in orbit. The GRS technology is direct heritage from LISA Pathfinder.

Sensitivity

The strain sensitivity (shown in Figure 12) corresponds to the noise spectrum of the instrument.

At low frequencies, it is dominated by residual acceleration noise of $3 \text{ fm s}^{-2}/\sqrt{\text{Hz}}$ per test mass. Above about 5 mHz, arm length measurement noise dominates, for which $12 \text{ pm}/\sqrt{\text{Hz}}$ are allocated, out of which $7.4 \text{ pm}/\sqrt{\text{Hz}}$ are

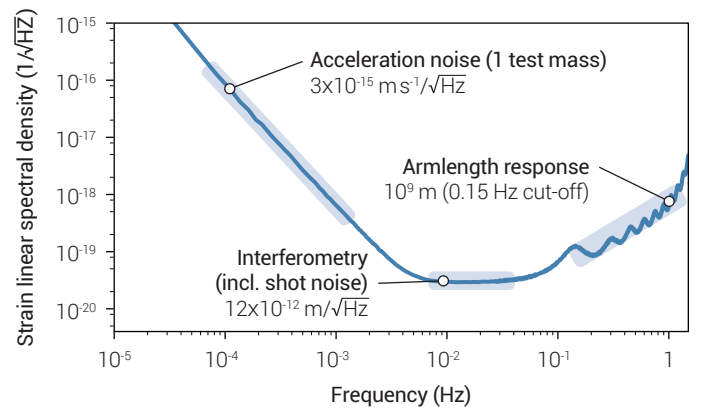


Figure 12: Time, sky and polarisation averaged eLISA sensitivity. The noise spectrum (strain sensitivity) is plotted as a linear spectral density.

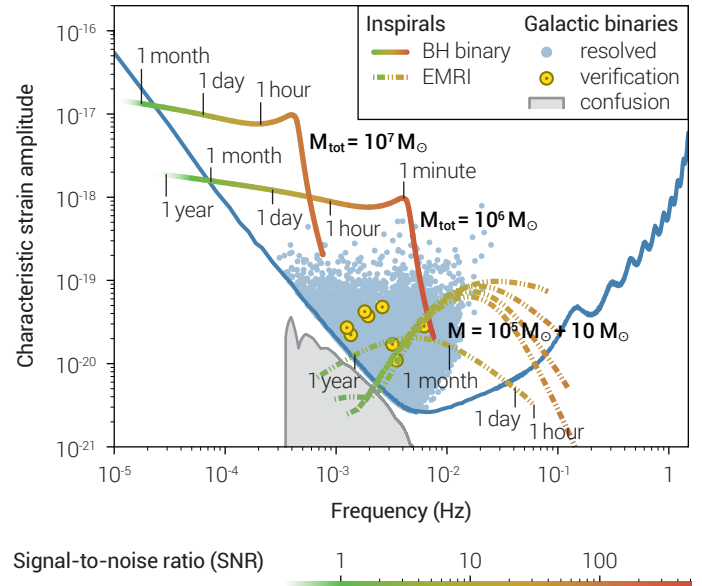


Figure 13: Examples of gravitational wave astrophysical sources in the frequency range of eLISA, compared with the sensitivity curve of eLISA. The data is plotted in terms of unitless ‘characteristic strain amplitude’. That is the strain resolvable in a single cycle for a broadband source, or the linear spectral density of Figure 12 times $\sqrt{\text{frequency}}$, or times $\sqrt{\text{number of cycles spent by the source at a given frequency during the observation}}$ for monochromatic sources. The tracks of two massive black hole binaries, located at $z = 3$ with equal masses (10^7 and $10^6 M_\odot$), are shown. The source frequency (and SNR) increases with time, and the remaining time before the plunge is indicated on the tracks. An equivalent plot is shown for an EMRI source at 200 Mpc, with 5 harmonic frequencies evolving simultaneously. Several thousand galactic binaries, with SNRs above 7, will be resolved after 1 year of observation. Some binary systems are already known, and will serve as verification signals. Millions of other binaries result in a ‘confusion noise’ that varies over the year. The average level is represented as grey shaded area.

quantum mechanical photon shot noise. At the highest frequencies, the sensitivity decreases again since multiple wavelengths of the gravitational wave fit into the arms, causing partial cancellation of the signal.

A unique feature of the eLISA interferometry is the virtual elimination of the effects of laser frequency noise. Stabilisation to a reference cavity built into the payload is not enough to suppress it completely. The remaining noise is removed by ‘Time-Delay Interferometry’ (TDI) [146], which synthesises a virtual balanced arm length interferometer in postprocessing. This requires knowledge of the absolute arm lengths to roughly 1 m accuracy, measured via an auxiliary ranging phase modulation imposed on

the laser beams. A second modulation is used to measure and remove noise caused by timing jitter of the Analogue to Digital Converter sampling clocks in the phasemeters [147].

Data Analysis

The data analysis algorithms for extracting the gravitational wave signals from the data and estimating their parameters were developed within the Mock LISA data challenge program. This program was successfully conducted from 2006 until 2011 and a new round of *eLISA* data challenges is currently being prepared. During the four rounds of the previous data challenges methods were developed for detecting gravitational wave signals from spinning massive black hole binaries, EMRIs, the population of galactic binaries, cusps formed on cosmic strings, as well as stochastic gravitational wave signals. The summary of each challenge can be found in [148–150]. The results show that we can always successfully detect and resolve gravitational wave signals, disentangling multiple sources of the same kind and of different kinds from the tens of millions of sources simultaneously present in the simulated data. In addition, the recovered parameters are always consistent with the true values to within the expected statistical uncertainty.

Technology status

All critical technologies for *eLISA* have been under intense development for more than 15 years, and today all are available in Europe, including the phasemeter. The interferometry with million km arms cannot be directly tested on ground, but it is being studied by scaled experiments and simulations. For the *eLISA* GRS, local interferometry, and the core of the drag-free and test mass control, LISA Pathfinder has allowed early identification and resolution of both technological development challenges and performance questions (see Figure 14). The GRS force noise budget has been largely verified at the level of LISA Pathfinder, and in some respects for *eLISA*, by torsion pendulum testing on the ground [151]. Additionally, tests with the LISA Pathfinder interferometer have allowed ground verification of the local displacement measurement across

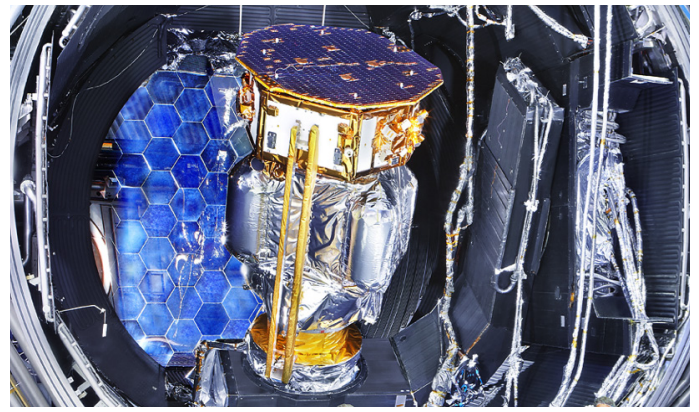


Figure 14: LISA Pathfinder. Much of the flight hardware for LISA Pathfinder is already tested and integrated on the spacecraft. In the last couple of years, the integrated system has been put through various test campaigns, from vibration and shock tests, to system level closed-loop communication and interface tests. The image shows the LISA Pathfinder satellite mounted on its propulsion module and situated in a space simulator in preparation for the Transit Orbit Thermal Test campaign. The space simulator aims to reproduce the thermal and vacuum environment in space. The bright light source at the back of the large vacuum tank simulates the sun allowing the spacecraft to be powered via its solar panel during the tests. The spacecraft has also undergone another test campaign in the same space simulator during which a Thermal Optical Qualification Model of the LTP (the science payload) was integrated. This second test campaign aimed to simulate the environment expected at L-grange point 1, with the temperature in the space simulator being cycled between the minimum and maximum predicted values. This On-Station Thermal campaign included some performance tests of the optical metrology system, the results of which showed that the performance of the system is well within the requirements.

the LISA Pathfinder band [152]. The 2015 Pathfinder flight represents a final verification and in-orbit commissioning of these systems and much of the *eLISA* metrology capability.

The ESA evaluation in the L1 process showed the NGO mission concept to be both technically feasible and compatible with the L2 cost target. It fits, with margin, into an Ariane V launch vehicle, though other launch scenarios (like the one studied for L1) are possible. There is strong interest from international partners, such as the US and China, to participate and contribute to *eLISA*, in which case a generous budget margin and/or enhanced science capability would be available. With or without international partners, Europe has the chance to take the lead in this revolutionary new science. ■

Science Objectives

Through the detection and observation of gravitational waves:

- Trace the formation, growth, and merger history of massive black holes
- Explore stellar populations and dynamics in galactic nuclei
- Test General Relativity with observations
- Probe new physics and cosmology
- Survey compact stellar-mass binaries and study the structure of the Galaxy

Event Rates and Event Numbers

Frequency band	$1 \times 10^{-4} \text{ Hz to } 1 \text{ Hz}$, ($3 \times 10^{-5} \text{ Hz to } 1 \text{ Hz}$ as a goal)
Massive black hole mergers	10 yr^{-1} to 100 yr^{-1}
Extreme mass ratio inspirals	5 yr^{-1} to 50 yr^{-1}
Galactic Binaries	~ 3000 resolvable out of a total of $\sim 30 \times 10^6$ in the <i>eLISA</i> band

V. SCIENTIFIC LANDSCAPE OF 2028

The science capabilities of the *eLISA* mission have been described in earlier sections. *eLISA* will pioneer gravitational wave observations in the rich frequency band around 1 mHz. In this section we examine this science return in the likely context of the L2 launch date of 2028. Given the predicted state of knowledge in 2028, we ask what unique contributions *eLISA* will make to our likely understanding of fundamental physics and astronomy at that time.

Naturally, science is not predictable, and the most interesting discoveries between now and 2028 will be the ones we cannot predict! But planned projects already hint at where the frontiers of science will be when *eLISA* operates. For example, massive progress can be expected in transient astronomy. Telescopes like LSST and the Square Kilometre Array (SKA) [153] are likely to identify new systems that flare up irregularly or only once, and there is a good chance that some of these will be associated with gravitational wave signals. As another example, extremely large telescopes (EELT, TMT, GMT) and large space telescopes (JWST) will be observing (proto-)galaxies at unprecedentedly high redshifts, at which *eLISA* will simultaneously observe individual merging black hole systems. As well as providing a wealth of information that will make it easier to identify the gravitational wave sources, the expected progress in all kinds of electromagnetic astronomy will sharpen the need for complementary gravitational wave observations of the unseen Universe.

Gravitational wave science by 2028

By 2028, gravitational wave astronomy will be well-established through ground-based observations operating at 10 Hz and above, and pulsar timing arrays (PTAs) at nHz frequencies. The huge frequency gap between them will be completely unexplored until *eLISA* is launched (see Figure 15).

The ground-based network of advanced interferometric detectors (three LIGO detectors, VIRGO [154], and the Kamioka Gravitational wave Detector, KAGRA [155]) will have observed inspiralling binaries up to around $100 M_{\odot}$ and measured the population statistics. Some, or all, of these detectors will have been further enhanced in sensitivity. It is possible that the third-generation Einstein Telescope (ET) will have come into operation by 2028 [156], further extending the volume of space in which these signals can be detected. At the other end of the mass spectrum, PTAs [157] will have detected a stochastic background due to many overlapping signals from supermassive black hole binaries with masses over $10^9 M_{\odot}$, and they may have identified a few individual merger events. The background will help determine the mass function of supermassive black holes at the high-mass end, but it will not constrain the mass function for the much more common $10^6 M_{\odot}$ black holes that inhabit the centres of typical galaxies and are ac-

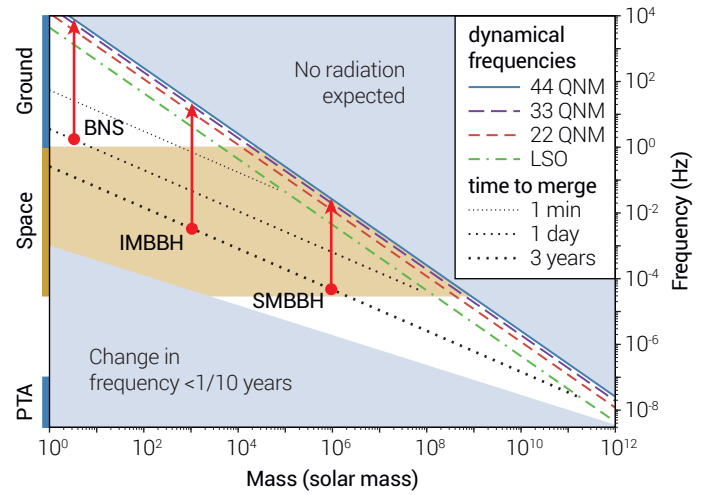


Figure 15: Measurement capabilities of ET, *eLISA*, and PTAs. The horizontal axis is the total mass of the binary system. The vertical axis is the frequency of the gravitational waves. The left side shows the frequency bands of the different instruments. The grey shaded region at the top is inaccessible because no system of a given mass can radiate at such high frequencies. The shaded region at the bottom is less interesting, because the chirp mass cannot be measured in an observation lasting less than 10 years, so the mass and distance of the source cannot be independently determined. Sloping dotted lines show the three-year, one-day, and one-minute time-to-merger lines. Sloping dashed lines are relevant dynamical frequencies: last stable orbit and the frequencies of ringdown modes of the merged black hole. Vertical lines indicate evolutionary tracks of systems of various masses as their orbits shrink and they move to higher frequencies. There is some overlap in the mass-range of sources that can be studied between ground and space detectors, but whether this translates into real science depends on the instruments' sensitivities and the source populations. Among the most interesting, would be the observation of an intermediate-mass binary black hole (IMBBH) system (if such systems exist) of around $1000 M_{\odot}$ during inspiral and coalescence with both *eLISA* and ET. This track is shown in the diagram as (IMBBH). With such a source at redshift 0.5, both instruments might have an SNR around 20 (see Figure 16), but most of the science—including the direction to the source—would come from the *eLISA* measurement. Additionally, tracks for binary neutron stars (BNS) and super massive binary black holes (SMBBHs) are shown.

cessible to *eLISA*. Ground-based gravitational wave observations are unlikely to constrain the existence and population statistics of the so-far elusive intermediate-mass black holes, although optical and X-ray observations might have done so by 2028. Besides making high-sensitivity observations of individual systems, *eLISA* will characterise the population statistics of black holes in the centres of galaxies, of intermediate mass black holes, and of the early black holes that eventually grew into the supermassive holes we see today.

By 2028, theoretical advances and predictable improvements in computer power will have made it possible to compute the complex waveforms expected from EMRIs and supermassive black hole binaries with high precision. This will allow searches in *eLISA* data to approach the optimum sensitivity of matched filtering, and it will make tests of General Relativity using these signals optimally sensitive.

eLISA and fundamental science in 2028

One of the signature goals of *eLISA* is to test gravitation theory, and it seems unlikely that any other method will achieve the sensitivity of *eLISA* to deviations of strong-field

gravity by 2028. Unlike ground-based instruments, *eLISA* will have sufficient sensitivity to be able to notice small corrections to Einstein gravity, and possibly to recognise unexpected signals that could indicate new phenomena.

By observing the long-duration waveforms from EMRI events, *eLISA* will map with exquisite accuracy the geometry of supermassive black holes, and will detect or limit extra scalar gravity-type fields. The MICROSCOPE (Micro-Satellite à traînée Compensée pour l'Observation du Principe d'Equivalence) mission [158] will, by 2028, have improved our limits on the violation of the equivalence principle or could, of course, have measured a violation, which would make *eLISA*'s strong-field observations even more urgently needed. X-ray and other electromagnetic observatories may measure the spins of a number of black holes, but *eLISA*'s ability to follow EMRI and merger signals through to the formation of the final horizon will not have been duplicated, nor will its ability to identify naked singularities or other exotic objects (such as boson stars or gravastars), if they exist.

By 2028 we will know much more about the large-scale Universe: in particular, about the nature of dark energy from the upcoming optical surveys dedicated to probing the large-scale structure. However, many questions will have remained open concerning the early Universe. From Planck and balloon flights of CMB instruments, we may know how much primordial gravitational radiation was produced at the end of inflation, which will help to pin down the actual inflationary scenario. However, there exists no means other than gravitational waves to probe the period in the evolution of the Universe ranging from reheating after inflation until big bang nucleosynthesis. Through the detection of gravitational waves, *eLISA* can gather information on the state of the Universe at much earlier epochs than those directly probed by any other cosmological observation. Gravitational waves are the next messengers to probe the very early Universe. They reach beyond the cosmic microwave background: *eLISA* has access to a fundamental frequency/energy-scale window, that of TeV. This scale is presently our boundary of knowledge in fundamental particle physics; new physics is therefore expected to emerge around that scale.

It is unclear how much progress will have been made by 2028 in understanding fundamental particle physics. The LHC will start probing new physics after its first upgrade in 2015 (reaching the scale of 14 TeV); another upgrade by a factor of 10 is expected around 2022, which will take the experiment through to 2030. It is difficult to foresee what the LHC will reveal about the nature of the Higgs, of dark matter particles, supersymmetry, extra dimensions, and so on. High-sensitivity *eLISA* observations may be crucial in providing clues here, since they explore the relevant energy scales in a completely unique way. The information contained in gravitational waves from the early Universe is complementary to, and independent of, the one accessible by particle accelerators. The presence of a first order phase

transition at the TeV scale, the presence of cosmic (super-) strings in the Universe, the properties of low-energy inflationary reheating, even the nature of the quantum vacuum state before inflation began (which could be different from the standard Quantum Field Theory nature in loop quantum gravity), are some of the fundamental issues that will still be open in 2028, and to which *eLISA* might provide some answers.

eLISA and late-time cosmology in 2028

By 2028 our understanding of the way cosmological structures formed will have been dramatically improved by high-redshift observations of QSOs and protogalaxies from missions like JWST [159], EUCLID [160] and the Wide-Field Infrared Survey Telescope (WFIRST) [161], and by the Atacama Large Millimeter/submillimeter Array (ALMA) [162] on the ground. These observations may well have constrained the supermassive black hole mass spectrum from a few times $10^{10} M_{\odot}$, or even higher, down to around $10^7 M_{\odot}$, but probably not into the main *eLISA* range of $10^4 - 10^6 M_{\odot}$, especially at $z > 2$. *eLISA* observations will fill this gap and also provide a check on selection effects and other systematics of the electromagnetic observations. By being able to measure the mass and spins of massive black holes as a function of redshift out to $z = 20$, *eLISA* will allow us to greatly improve models of how supermassive black holes grow so quickly, so as to be in place at $z \sim 7$. We will additionally learn what roles accretion and mergers play in the growth of all massive black holes. *eLISA* observations of mergers of $10^4 - 10^5 M_{\odot}$ black holes out to $z = 20$ (if they exist) can provide a strict test of the amount of growth by merger expected in these models.

eLISA and massive black holes in 2028

In the next few years, eROSITA will study tidal disruptions of stars out to redshifts of $z \sim 1$ and will look for massive black holes [163], although in the high-mass regime compared to *eLISA*. *eLISA* has extraordinary sensitivity to massive black holes in the mass-range characteristic of most galactic-core black holes (see Figure 16). Gravitational wave detectors like *eLISA* are inherently all-sky monitors: always on and having a nearly 4π steradian field of view. They naturally complement other surveys and monitoring instruments operating at the same time, like LSST [164], SKA, neutrino detectors, gamma-ray and X-ray monitors. The massive black hole mergers detected by *eLISA* out to modest redshifts ($z = 5 - 10$) could well be visible to SKA and LSST as transients in the same region of the sky. The identification of 5 to 10 counterparts during a 2 year *eLISA* mission would not be surprising. These might then be followed up by large collecting area telescopes like TMT, GMT, and EELT, providing an unprecedented view of the conditions around two merging massive black holes. Interestingly, the advent of observing with detectors like aLIGO is leading to the development of networks of optical telescopes for multimessenger astronomy. These are

designed to follow up gravitational wave triggers and find associated transient phenomena. These systems may usefully supplement LSST in picking out *eLISA* counterparts.

eLISA has very good sensitivity to as many as 4 or 5 ringdown frequencies of newly formed black holes (see Figure 16). This ‘black hole spectroscopy’ will allow *eLISA* to address important questions that, by 2028, will probably not yet have been answered. Most important will be to find evidence for naked singularities or other exotic objects: The ringdown frequencies make it possible to determine the mass and spin of the final black hole, and will be very different for any of the proposed alternatives. In addition, the ringdown modes show in detail how a dynamical black hole behaves; not even ET, if operational at that time, will have the sensitivity to make this kind of detailed study of strong gravity.

eLISA and the astrophysics of stars and the Galaxy in 2028

eLISA will perform, for the first time, a complete census of very compact binary systems throughout the Galaxy. Thousands of white-dwarf binaries are expected, along with binaries involving neutron stars and black holes in various combinations. GAIA’s catalogue will still, in 2028, be the principal optical reference for these observations, and we can expect dozens or more binaries in that catalogue to be observed by *eLISA*. *eLISA* will identify many more, the nearest of which can then be followed up with JWST if still operating, and E-ELT. These observations will lead to improved understanding of interactions, mass transfer, and double white dwarfs as supernova progenitors. By 2028, aLIGO and partners will have good statistics on the population of relativistic compact binaries out to Gpc distances, and *eLISA*’s complete census of that population in the Galaxy will allow us to compare our Galaxy with the cosmological norm, a comparison that is very difficult to make with any other stellar population, revealing much about the history of our Galaxy. If binaries of $100 M_{\odot}$ black holes exist, then *eLISA* and ET could make joint observations of a few merging systems with comparable sensitivity, improving on the angular positions which ET could measure alone.

The Event Horizon Telescope (EHT) should test general relativity and probe the horizon of Sgr A*, the massive black hole at the centre of the Milky Way [165]. EHT should also explore the shape and properties of the accretion flow onto Sgr A* and measure its spin [166–167].

Within the next few years, GRAVITY [168] may be able to observe the orbits around Sgr A* of currently unknown stars with periods of just 1–2 years, test General Relativity, and infer the hidden distribution of the dark population of objects around Sgr A*, thereby studying the different models for mass-segregation in galactic nuclei [61, 66, 169–170]. Such studies should provide more accurate estimates of the event rate for EMRIs in the *eLISA* band. ■

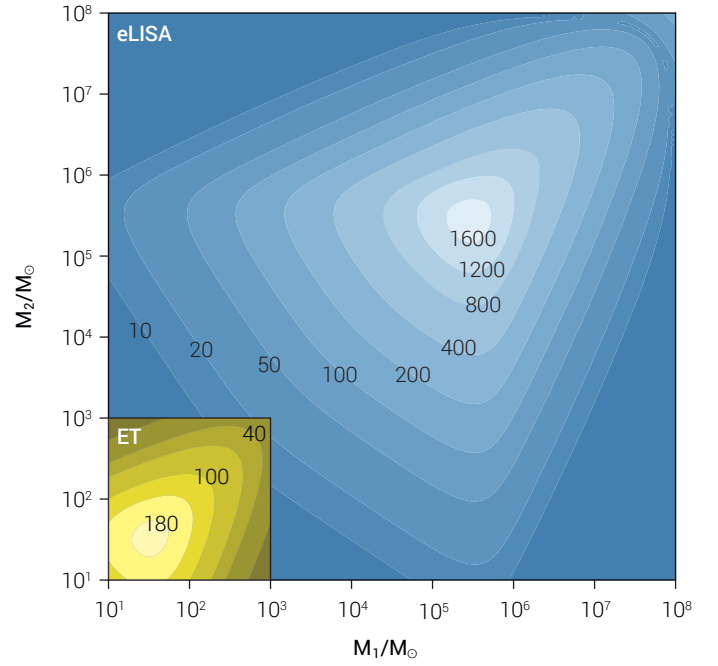


Figure 16: SNR contours of *eLISA* and ET for binary black holes at redshift 0.5. The axes are the masses (in solar masses) of the two components of the binary. The main contour plot is for *eLISA*, and the box in the lower left shows the contours for ET, on the same axes. The two observatories overlap for sources at this distance only for binaries with comparable masses, around $1000 M_{\odot}$. For smaller distances, the SNR scales as $1/r$; for larger redshifts the SNR also scales inversely with the luminosity distance and the masses redshift as well, so that the axis scales show $(1+z)M$, where M is the intrinsic (rest-frame) mass. The SNR values have been calculated using only the inspiral part of the waveform; they underestimate the true sensitivity by omitting the merger and ringdown radiation. This is particularly important at the higher mass end for each detector.

CONCLUSION

In summary, by 2028 our understanding of the Universe will have been dramatically improved by advanced observations of electromagnetic radiation. Adding a low-frequency gravitational wave observatory will add a new sense to our perception of the Universe. Gravitational waves will allow us to ‘hear’ a Universe otherwise invisible with light.

eLISA will be the first ever mission to survey the entire Universe with gravitational waves. It will allow us to investigate the formation of binary systems in the Milky Way, detect the guaranteed signals from the verification binaries, study the history of the Universe out to redshifts of order 20, test gravity in the dynamical strong-field regime, and probe the early Universe at the TeV energy scale. *eLISA* will play a unique role in the scientific landscape of 2028.

The NGO mission studied by ESA for the L1 mission selection serves as a strawman mission concept that is capable of matching the science requirements for *eLISA*. It has been evaluated by ESA as both technically feasible and compatible with the L2 cost target. ■

REFERENCES

- 1 M. Burgay et al., Nature 426, 531 (2003)
- 2 A.G. Lyne et al., Science 303, 1153 (2004)
- 3 J. Antoniadis et al., Science 340, 6131 (2013)
- 4 Planck Collaboration, ArXiv e-prints 1303.5062, (2013)
- 5 S.D.M. White and M.J. Rees, MNRAS 183, 341 (1978)
- 6 P. Madau and M.J. Rees, ApJ 551, L27 (2001)
- 7 M.C. Begelman et al., MNRAS 370, 289 (2006)
- 8 M.C. Begelman et al, Nature, 287, 307 (1980)
- 9 M. Volonteri et al., ApJ 582, 559 (2003)
- 10 S. Komossa et al. ApJ 582, L15 (2003)
- 11 X. Liu et al., ApJ 737, 101L (2011)
- 12 L. Ferrarese and D. Merritt, ApJ 539, L9 (2000)
- 13 K. Gebhardt et al. ApJ 539, L13 (2000)
- 14 T. R. Marsh, CQG 28 (9), 094019 (2011)
- 15 O. Jennrich et al., ESA/SRE(2011) 19 (2012)
- 16 P. Amaro-Seoane et al., GW Notes 6, 4 (2013)
- 17 N.A. Grogin et al., ApJS 197, 35 (2011)
- 18 R. Salvaterra et al., Nature 461, 1258 (2009)
- 19 D. Coe et al. ApJ 762,32 (2013)
- 20 D.J. Mortlock et al., Nature 474, 616 (2011)
- 21 R. Barkana and A. Loeb, PhysRep 349, 125 (2001)
- 22 R. H. Becker et al., AJ 122, 2850 (2001)
- 23 M.A. Worsely et al., MNRAS 357, 1281 (2005)
- 24 M. Tegmark et al., ApJ 474, 1 (1997)
- 25 V. Bromm et al., ApJ 527, L5 (1999)
- 26 M.Volonteri, A&ARv 18, 279 (2010)
- 27 B. Devecchi et al., MNRAS 421, 1465 (2012)
- 28 L. Mayer et al., Nature 466, 1082 (2010)
- 29 P. Madau et al., MNRAS 283, 1388 (1996)
- 30 P. F. Hopkins et al, ApJ 654, 731 (2007)
- 31 T. Di Matteo et al., Nature, 433, 604 (2005)
- 32 P. F. Hopkins et al., ApJS, 163, 1 (2006)
- 33 D. J. Croton et al., MNRAS, 365, 11 (2006)
- 34 L. Mayer et al., Science 316, 1874 (2007)
- 35 M. Dotti et al., AdAst, 2012 (2012)
- 36 M. Colpi and M.Dotti, ASL 4,181 (2011)
- 37 S. Callegari et al., ApJ 729, 85 (2011)
- 38 S. Cole et al., MNRAS 319, 168 (2000)
- 39 G. De Lucia et al., MNRAS 366, 499 (2006)
- 40 Q. Guo et al., MNRAS 413, 101 (2011)
- 41 A. Marconi et al., MNRAS 351, 169 (2004)
- 42 C. Cutler and E. Flanagan, PRD 49, 2658 (1994)
- 43 F. Ohme, CQG 29, 124002 (2012)
- 44 L. Santamaría et al., PRD, 82, 064016 (2010)
- 45 K. G. Arun et al, CQG 26, 094027 (2009)
- 46 A. Sesana et al., PRD 83, 44036 (2011)
- 47 A.Sesana et al., MNRAS 377, 1711 (2007)
- 48 T. Tanaka and Z. Haiman, ApJ 696, 1798 (2009)
- 49 T. Tanaka et al., MNRAS 425, 2974 (2012)
- 50 E. Berti and M. Volonteri, ApJ 684, 822 (2008)
- 51 N. McConnell and C. Ma, ApJ 764, 184 (2013)
- 52 A. M. Ghez et al., ApJ 689, 1014 (2008)
- 53 S. Gillessen et al., ApJ 692, 1075 (2009)
- 54 R. Genzel et al., ApJ 594, 812 (2003)
- 55 S. Konstantinidis et al., accepted at A&A, (2013)
- 56 P. Brem et al., ArXiv e-prints 1211.5601, (2012)
- 57 M. Freitag et al., ApJ 649, 91 (2006)
- 58 P. Amaro-Seoane, ArXiv e-prints 1205.5240, (2012)
- 59 L. Barack and C. Cutler, PRD 69, 082005 (2004)
- 60 J.R.Gair, CQG 26, 094034 (2009)
- 61 M. Preto and P. Amaro-Seoane, ApJ 708, L42 (2010)
- 62 P. Amaro-Seoane et al, MNRAS 429, 3155 (2013)
- 63 J. R. Gair et al., PRD 81, 104014 (2010)
- 64 S. Gezari et al., ArXiv e-prints 0903.1107 (2009)
- 65 M. Volonteri and P. Natarajan, MNRAS, 400, 1911 (2009)
- 66 P. Amaro Seoane and M. Preto, CQG 28, 094017 (2011)
- 67 P. Amaro-Seoane et al., CQG 24, R113 (2007)
- 68 N. Cornish et al., PRD 84, 062003 (2011)
- 69 C. Huwyler et al., PRD 86, 084028 (2012)
- 70 T. G. F. Li et al., PRD 85, 082003 (2011)
- 71 C. K. Mishra et al., PRD 82, 064010 (2010)
- 72 E. Berti et al., PRD 84, 101501(R) (2011)
- 73 N. Yunes et al., PRD 81, 064018 (2010)
- 74 S. Mirshekari et al., PRD 85, 024041 (2012)
- 75 S. Chandrasekhar and S. Detweiler, Proc. R. Soc. A 344, 441 (1975)
- 76 E. W. Leaver, Proc. R. Soc. A 402, 285 (1985)
- 77 E. Berti et al., CQG 26, 163001 (2009)
- 78 O. Dreyer et al., CQG 21, 787 (2004)
- 79 E. Berti et al., PRD 73, 064030 (2006)
- 80 E. Berti and V. Cardoso, Int. J. Mod. Phys. D 15, 2209 (2006)
- 81 C. B. M. H. Chirenti and L. Rezzolla, CQG 24, 4191 (2007)
- 82 E. Barausse and T. P. Sotiriou, PRL 101, 099001 (2008)
- 83 P. Pani and V. Cardoso, PRD 79, 084031 (2009)
- 84 I. Kamaretsos et al., PRD 85, 024018 (2012)
- 85 I. Kamaretsos et al., PRL 109, 141102 (2012)
- 86 E. Berti, Braz. J. Phys. online, 1 (2013)
- 87 R. Genzel et al., MNRAS 317, 348 (2000)
- 88 R. Genzel et al., Rev. Mod. Phys. 82, 3121 (2010)
- 89 F. D. Ryan, PRD 52, 5707 (1995)

- 90 N. A. Collins and S. A. Hughes, PRD 69, 124022 (2004)
- 91 K. Glampedakis and S. Babak, CQG 23, 4167 (2006)
- 92 S. A. Hughes et al., AIP Conf. Proc. 873, 13 (2006)
- 93 S. Vigeland et al., PRD 83, 104027 (2011)
- 94 S. Vigeland and S. A. Hughes, PRD 81, 024030 (2010)
- 95 S. Babak et al., CQG 28, 114001 (2011)
- 96 J. R. Gair et al., ArXiv e-prints, 1212.5575 (2013)
- 97 C. F. Sopuerta, GW Notes 4, 3 (2010)
- 98 M. Kesden et al., PRD 71, 044015 (2005)
- 99 P. Pani et al., PRD 81, 084011 (2010)
- 100 J. R. Gair et al., PRD 77, 024035 (2008)
- 101 G. Lukes-Gerakopoulos et al., PRD 81, 124005 (2010)
- 102 P. Pani et al., PRD 83, 104048 (2011)
- 103 C. F. Sopuerta and N. Yunes, PRD 80, 064006 (2009)
- 104 P. Canizares et al., PRD 86, 044010 (2012)
- 105 E. Berti et al., PRD 85, 122005 (2012)
- 106 S. T. McWilliams, PRL 104, 141601 (2010)
- 107 C. G. Böhmert et al., PRD 71, 084026 (2005)
- 108 K. Yagi and T. Tanaka, PRD 81, 064008 (2010)
- 109 V. Cardoso et al., PRL 107, 241101 (2011)
- 110 J. Gair and N. Yunes, PRD 84, 064016 (2011)
- 111 D. E. Holz and S. A. Hughes, ApJ 629, 15 (2005)
- 112 B. F. Schutz, Nature 323, 310 (1986)
- 113 J. D. Schnittman, CQG 28, 094021 (2011)
- 114 A. Petiteau et al., ApJ 732, 82 (2011)
- 115 C. L. MacLeod and C. J. Hogan, PRD 77, 043512 (2008)
- 116 J. R. Gair and E. K. Porter, ASP Conf. Ser. 467, 173 (2012)
- 117 E. Witten, PRD 30, 272 (1984)
- 118 C. J. Hogan, MNRAS 218, 629 (1986)
- 119 A. D. Dolgov et al., PRD 66, 103505 (2002)
- 120 A. Kosowsky et al., PRD 66, 024030 (2002)
- 121 C. J. Hogan, PRL 85, 2044 (2000)
- 122 L. Randall and G. Servant, JHEP 0705, 54 (2007)
- 123 S. Khlebnikov and I. Tkachev, PRD 56, 653 (1997)
- 124 R. Easther and E. A. Lim, JCAP 04, 010 (2006)
- 125 G. N. Felder and L. Kofman, PRD 75, 043518 (2007)
- 126 R. Brustein et al., Phys. Lett. B. 361, 45 (1995)
- 127 A. Buonanno et al., PRD 55, 3330 (1997)
- 128 A. Buonanno, ArXiv e-prints, gr-qc/0303085 (2003)
- 129 E. J. Copeland et al., JHEP 6, 13 (2004)
- 130 P. Binetruy et al., JCAP 1206, 027 (2012)
- 131 G. Nelemans et al., A&A 368, 939 (2001)
- 132 N. Ivanova et al., A&ARv 21, 59 (2013)
- 133 R. F. Webbink, ApJ 277, 355 (1984)
- 134 H. B. Perets et al., Nature, 465, 322 (2010)
- 135 A. L. Piro and S. R. Kulkarni, ApJ 762, L17 (2013)
- 136 M. R. Adams and N. J. Cornish, PRD 86, 124032
- 137 A. Farmer and E. S. Phinney, MNRAS 346, 1197 (2003)
- 138 J. A. Edlund et al., PRD 71, 122003 (2005)
- 139 A. J. Ruiter et al., ApJ 693, 383 (2009)
- 140 S. Nissanke et al., ApJ 758, 131 (2012)
- 141 K. Belczynski et al., ApJ 725, 816 (2010)
- 142 T. B. Littenberg et al., MNRAS, 429, 2361 (2013)
- 143 M. Dan et al., MNRAS 422 (3), 2417 (2012)
- 144 Antonucci et al., CQG 28, 094002 (2011)
- 145 Antonucci et al., CQG 29, 124014 (2012)
- 146 S. J. Mitryk et al., PRD 86, 122006 (2012)
- 147 G. Heinzl et al., CQG 28, 094008 (2011)
- 148 K. A. Arnaud et al., CQG 24, S529 (2007)
- 149 S. Babak et al., CQG 25, 114037 (2008)
- 150 S. Babak et al., CQG 27, 084009 (2010)
- 151 A. Cavalleri et al., CQG 26, 094017 (2009)
- 152 H. Audley et al., CQG 28, 094003 (2011)
- 153 F. Aharonian et al., ArXiv e-prints 1301.4124, (2013)
- 154 J. Aasi et al., ArXiv e-prints 1304.0670, (2013)
- 155 K. Somiya, CQG 29, 124007 (2012)
- 156 M. Punturo et al., CQG 27, 194002 (2010)
- 157 G. Hobbs et al., CQG 27, 084013 (2010)
- 158 P. Touboul et al., CQG 29, 184010 (2012)
- 159 M. Stiavelli et al., Baltimore: STScI, (2008)
- 160 R. Laureijs et al., ArXiv e-prints 1110.3193, (2011)
- 161 J. Green et al., ArXiv e-prints 1108.1374, (2011)
- 162 A. Wootten and A. R. Thompson, Proc. IEEE 97, 1463 (2009)
- 163 Z. Ivezic et al., ArXiv e-prints 0805.2366, (2008)
- 164 A. Merloni et al., ArXiv e-prints 1209.3114, (2012)
- 165 H. Falcke et al., AIP Conf. Ser. 522, 317 (2000)
- 166 S. S. Doeleman et al., Nature 455, 78 (2008)
- 167 S. S. Doeleman et al., ArXiv e-prints 0906.3899, (2009)
- 168 F. Eisenhauer et al., Proc. SPIE 7013, 1 (2008)
- 169 J. Miralda-Escudé and A. Gould, ApJ 545, 847 (2000)
- 170 M. Freitag et al., ApJ 649, 91 (2006)

Links to the original documents can be found at
<http://elisascience.org/references>

On the cover page: Gravitational waves generated by the white dwarf pair RX J0806.3+1527 (artist's impression) in front of the Milky Way's Galactic Centre (wide spectrum composite image) with black hole Sagittarius A*. Credit: NASA, ESA, SSC, CXC, STScI, and AEI.

SOLARIS: SOLAR sail Investigation of the Sun

Spokesperson: APPOURCHAUX, Thierry
Institut d'Astrophysique Spatiale, Orsay, France
+(33) 1 69 85 86 29
Thierry.Appourchaux@ias.u-psud.fr

Supporters

AUCHERE, Frederic	Institut d'Astrophysique Spatiale, Orsay, France
ANTONUCCI, Ester	Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Italy
GIZON, Laurent	Max Planck Institute for Solar System Research, Katlenburg Lindau, Germany
MACDONALD Malcolm	Advanced Space Concepts Laboratory - Glasgow G1 1XJ - Scotland
HARA, Hirohisa	National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan
SEKII, Takashi	National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan
MOSES, Dan	Naval Research Laboratory, Washington DC, USA
VOURLIDAS, Angelos	Naval Research Laboratory, Washington DC, USA

I.1. SCIENTIFIC OBJECTIVES

Understanding the origins of solar magnetic activity has been at the forefront of solar and stellar physics since the discovery of the 11-year sunspot cycle nearly two centuries ago. Unravelling this mystery has broad implications not only for promoting a deeper knowledge of the Sun itself but also for understanding the Sun's influences on the heliosphere, the geospace environment, and potentially the Earth's climate system. Such influences regulate space weather, with increasing economic impacts on our technological society as our reliance on telecommunications systems, power grids, and airline travel continues to grow. As a readily observable example of an astrophysical magneto-hydrodynamic (MHD) dynamo, the Sun also provides unique insights into the generation of magnetic fields by turbulent plasma flows throughout the universe, from planetary and stellar interiors to stellar and galactic accretion disks to interstellar clouds.

The global magnetic polarity of the Sun reverses during each 11-year sunspot cycle so that the overall period of the solar magnetic activity cycle is 22 years. It is a formidable challenge to understand how such remarkable regularity arises from the highly turbulent conditions of the solar convection zone and how magnetic flux emerges from the solar interior to energize the solar atmosphere and power solar variability. Large-scale flows (differential rotation and meridional circulations) established by turbulent convection, plasma instabilities, and nonlinear feedbacks all play an important role, spanning many orders of magnitude in spatial and temporal scales.

Modern solar observations coupled with sophisticated theoretical and numerical models have yielded important insights into many aspects of solar magnetism but the basic physical mechanisms responsible for generating these fields are still not understood. To make great

scientific progress in our understanding the Sun and the fundamental problems of cosmic magnetism, there is no doubt that we need to continue both theoretical and observational efforts. On the observational side, measuring solar internal flows is of the greatest importance. For this task helioseismology has proven to be a powerful tool.

Helioseismic measurements are based on surface wavefield data, normally and preferably temporal series of photospheric Dopplergrams, which are then analyzed to probe the solar interior structure and flows. With the so-called global methods, the wavefield data are used to measure (mainly acoustic) eigenfrequencies of the Sun. The eigenfrequencies are then analyzed, often by way of inverse methods, to probe the solar interior for thermal and dynamical structure of high degrees of symmetry, such as the spherically symmetric distribution of sound speed, or differential rotation as the axisymmetric component of flows. With new local methods, wavefield data are used to measure local resonant properties or wave propagation time for a given pair of points, by cross-correlating local wavefields. These travel-time data are then analyzed to probe the interior for local and/or asymmetric structures, such as meridional flow, convection and flows around active regions (Gizon et al. 2010).

Differential rotation and meridional flows have already been measured by such helioseismology techniques, up to about 60° latitude with a typical uncertainty of the order of a m s^{-1} (Thompson et al 1996). It is essential to extend this measurement to the polar region, partly because without such measurement we will never be confident of our understanding of dynamics of the Sun as a whole, and partly because the polar region is where the magnetic flux reverses and the meridional flow, which plays an important role in carrying the magnetic flux, should turn in towards the solar interior, and where polarity reversals take place in the surface layers.

Another related mystery is the total solar irradiance variation over the solar cycle. The total irradiance of the other solar-like stars that exhibits activity at a level that is similar to the Sun, on average varies around 0.3 per cent over their activity cycles. On the other hand, the solar irradiance varies only by 0.1 per cent. There is a well-founded suspicion that the solar irradiance depends on latitude, thereby creating a great interest in measuring the solar irradiance from high heliographic latitudes.

Because of the orbit characteristics of the vast majority of spacecraft, the solar flux has predominantly been measured at Earth or at least in the plane of the ecliptic. Therefore, the existing data do not directly demonstrate the fact that the latitudinal distribution of the extreme-ultraviolet (EUV) solar flux is largely anisotropic. Indeed, in the EUV the nonuniform distribution of very contrasted bright features (i.e., active regions) and dark features (i.e., coronal holes) at the surface of the Sun produces both the obvious rotational (or longitudinal) modulation of the flux and also a strong latitudinal anisotropy. This has been demonstrated by Auchère et al. (2005), but no direct measurements exist. Although largely ignored up to now, the latitudinal anisotropy has several important implications. A first consequence is that the total EUV output of the Sun is currently overestimated, which affects comparisons with other stars. The EUV latitudinal anisotropies also affect the photo-ionization rates of helium that are used to derive the helium abundance in the LISM from measurements made in-situ in the heliosphere (e.g. by GAS/Ulysses). The helium abundance in the LISM is a major tracer of galactic evolution. Measuring the solar flux in at least two EUV wavebands will allow to test and to complete the models of the irradiance anisotropies derived from in-Ecliptic observations.

Many facts point to a weak coupling between the magnetic fields in the northern and southern hemispheres of the Sun. The reversals of the magnetic field polarity at the poles do not occur simultaneously, but one can be delayed of about two years relative to the other. Emergence in

the photosphere of the magnetic field was organized in a persistent large-scale pattern with different rotation periods in the northern and southern hemispheres, in cycles 20 and 21, pointing to a weak interdependence of the field systems originating in the two hemispheres (Antonucci et al, 1990). Galactic cosmic rays modulated by solar activity are affected by a north-south gradient that reverses at the reversal of the solar magnetic dipole, indicating a different level of activity, and thus magnetic flux, in the two hemispheres. A significant north-south asymmetry in the EUV solar flux was discovered in solar cycle 23. Measurements of solar irradiance at the poles have not yet been possible and therefore we cannot assess whether this quantity is the same at the north and the south Pole. Thus observations suggest that the magnetic field systems generated in the two hemispheres can indeed evolve independently, although they tend to exchange characteristics at the polarity reversal, following the magnetic cycle.

Recent results show that the corona itself is site of a vast variety of fluctuations. The coronal images, obtained with STEREO, show the existence of quasi-periodic non-stationary density variations characterized by a wide range of temporal and spatial scales and strongly confined by the magnetic topology. In closed field line structures, at mid and low latitudes, the density variations might be interpreted as due to slow standing magneto-acoustic waves excited by the convective super-granular motions (Telloni et al. 2013). In the outer corona, observed in the ultraviolet line HI Ly alpha with UVCS-SOHO, density fluctuations are coherent and persistent in the slow coronal wind, at mid and low latitudes, whilst in the fast wind, at high latitudes, they are stochastic and non-correlated (Telloni et al. 2009). The analysis and interpretation of this kind of coronal phenomena is still in its infancy and a complete scenario which links the coronal fluctuations back to their magnetic roots in the photosphere and which helps to assess their role in the energetics of the solar corona and the solar wind, is still to be developed.

Finally, several aspects of the large-scale magnetic configuration of the corona and the interplay between explosive activity and background magnetic field remain either elusive or ambiguous. Questions such as: *what is the 3D structure of the global corona? What is the longitudinal extent of coronal mass ejections (CMEs)? What is the magnetic connectivity between the outgoing CME and low corona? Are post-CME rays the plasma envelopes of the post-CME current sheet predicted by theories of eruptions?* remain open issues despite several years of multi-viewpoint observations from STEREO. For example, both STEREO spacecraft lie on the ecliptic and hence can provide information for the density distribution of the background corona and CMEs only normal to the ecliptic. For this reason, inversion techniques cannot be used reliably unless out-the ecliptic information on the longitudinal spread of coronal structures is available. Such observations provided by a wide angle EUV imager will also be able to provide directly the longitudinal evolution of eruptions from the very low corona ($< 70,000$ km) to the outer corona (~ 15 Rs) and verify whether there exists a super-radial expansion phase during the early part of the event, as postulated by Patsourakos & Vourlidas (2012), whether the changes in the CME angular width are due simply to projection effects (i.e. rotation) or due to interaction with the ambient coronal structures, whether CMEs can be deflected longitudinally by nearby coronal holes (Makela et al 2013), and whether CMEs can have angular momentum coupling with the Sun as proposed by Liu et al (2010). All these effects can have important consequences for understanding the development of CMEs and eventually predicting their geoeffectiveness. Wide angle EUV observations of post-CME structures, from high latitudes, will finally uncover the natures of these features. They will show both the extent and connectivity of these structures between the outgoing CME and post-eruption flaring arcade. Although the upcoming Solar Orbiter mission will make the first inroads towards addressing these problems, its modest maximum

latitude of 34° does not result in a significantly different view of coronal structures in regards to ecliptic observations. Besides, those extreme latitudes will last for only a few days making it difficult to study the intermittent solar eruptive activity from that platform. The best science return can be achieved from long-term synoptic observations from high latitudes as envisioned in this white paper.

The outstanding issues confronting our current understanding of the solar dynamo and its manifestation in the corona may be summarized through several key scientific questions:

- How is the global, cyclic, solar magnetic field generated?
- What is the nature of flows in the polar regions of the Sun and how do they interact with magnetism?
- Which is the degree of coupling between the magnetic field generated in the northern and southern hemispheres of the Sun?
- How does the radiative energy output of the Sun depend on latitude?
- How does the solar dynamo work and drive connections between the Sun and the heliosphere?
- What is the 3D structure of the global corona?

Progress on these scientific questions requires detailed observations of the solar polar regions, where data is currently scarce and where much of the subtle interplay between plasma flows and magnetic fields that gives rise to cyclic polarity reversals is thought to occur. The out-of-ecliptic observations of the Sun, for the first time, will provide an opportunity for detailed investigations of the magnetic structure and dynamics of the polar region. High-latitude photospheric observations will also provide an unprecedented vantage point for helioseismic imaging that can be used to probe flows and fields in the deep convection zone and tachocline where solar activity is ultimately thought to originate.

In addition to measurements at the photospheric level, the structures of the outer solar atmosphere in polar region and the heliospheric structures merit observations from outside the ecliptic. The poles of sun undergo dramatic change during the 11-year solar cycle, driven by the dynamo action in the solar convection zone. The polar vantage point gives unique opportunities for understanding the origin of the fast solar wind spectroscopically and for stereo viewing of surface vector magnetic fields, coronal structures, and Earth-directed CMEs in coordination with observatories near the Earth. The unique inclination for the SOLARIS observatory will also permit unprecedented measurements of the total solar irradiance. This may help resolve the discrepancy of the cycle variation of the solar irradiance of $\sim 0.1\%$ while solar analogues vary, on average, by 0.3% .

With this in mind, we propose the following prime measurement targets for the SOLARIS mission:

- T1) Photospheric magnetic flux distribution and evolution in the polar region
- T2) Dynamical coupling between magnetic fields and flows
- T3) High-precision measurement of total solar irradiance
- T4) Dynamics of the background and transient solar winds
- T5) Latitudinal variations of the EUV irradiance and coupling with the magnetic activity

I.2. INSTRUMENTATION AND MISSION REQUIREMENTS

SOLARIS from its highly inclined orbit around the Sun, aims to combine helioseismic and magnetic observations, solar irradiance measurements and EUV images at various latitudes. A total mass of 35 to 50 kg is envisaged for the following three instruments.

The highest science objective of SOLARIS is measuring subsurface flow at high-latitude regions, by local helioseismology techniques, based on Dopplergrams acquired by an HMI-MDI type instrument. From this measurement we will derive differential rotation, meridional circulations and convective flows in the upper convection zone. These flow-field measurements will then be cross-correlated with surface magnetic field measurement, to reveal how magnetic flux is transported to the polar region, and how the polarity reversals take place as interplay between plasma flows and magnetic fields. The predominantly vertical kG-field patches that Hinode has found in the polar regions are large enough to be observed. A serious attempt will also be made for stereoscopic helioseismology, for investigating deeper layers, including the best ever shot of the solar tachocline region. Space heritage on this type of instrument is extremely high if we include also the lighter development done for the Polarimetric and Helioseismic Imager of Solar Orbiter.

Solar irradiance measurements at various latitudes will, for the first time, enable us to measure the anisotropy of the total solar irradiance. The unexplained low photometric variability of the sun may be explained by higher variability at higher latitudes, likely caused by faculae. If it is not the case, then we must conclude that the Sun is a rather atypical star, which will lead to more fundamental questions in astrophysics. Space heritage is also very high given the various irradiance monitor flown on VIRGO, PICARD and other missions.

The wide-field EUV telescope must be able to produce images of the corona in at least two EUV wavelengths (e.g. 30.4 nm and 17.4 nm) and cover a field of view (FOV) of 6 -10 degrees ($\sim 10-15 R_s$ from 0.4 AU) centered on the Sun. Over the past 10 years, the quality of the EUV optics has improved to such a degree that we can seriously contemplate using EUV imagers, which are lighter, more compact and easier to build, to replace low corona coronagraphs whose design remains challenging due to the intrinsically high contrast between the solar disk and the corona at visible wavelengths. Indeed the most recent developments in EUV technologies make it possible to design and build a wide field of view (5 degrees or more), compact (shoebox size) and lightweight (5kg) multi-band EUV telescope. The instrument will be based on the compact design of the Full Sun Imager (FSI) on board Solar Orbiter (Auchère et al. 2005; Halain et al. 2010). FSI includes a movable occulting disk to limit the level of instrumental stray-light for observations above $3R_s$. The 30.4 nm and 17.4 nm bands are dominated by spectral lines formed by resonant scattering above 2 to 3 R_s , which produces a signal proportional to the electron density, as in white light coronagraphs. Several technical options, such as using a smaller pitch detector, will be studied to make the design even more compact than on Solar Orbiter

The main mission requirements are as follows:

- Data collection for the primary science goals shall begin within five years of launch
- Data collection for the primary science goals shall begin when the payload can observe above ± 60 degrees of solar latitude
- In an orbit passing directly over the solar poles, the pole shall be continuously visible at an observation-zenith angle of less than 30 degrees (TBC) for one full rotation period
- The orbit shall have a very low eccentricity < 0.2 (close to a circular orbit).

- EUV observations will be performed at all latitudes with primary data collection at latitudes ± 45 degrees of solar latitude.
- At most two spacecraft for having a 3D view of the corona.

These requirements are much more severe than those put on the Solar Orbiter mission which is not suited for long helioseismic observation and solar irradiance measurements mainly due to the lower inclination to the highly eccentric orbit (>0.8).

I.3. GOSSAMER WORKING GROUP

In the framework of a study led by DLR and ESA, the Gossamer working group (GWG) was set up in order to study how missions based on the use of solar sail could be feasible. The GWG is led by M. Macdonald (Chair and engineer PI) and by T. Appourchaux (Science PI) assisted of several European engineers and of several scientists comprising L. Gizon (Max Planck Institute for Solar System Research) and T. Sekii (NAOJ). The activities started in November 2011 and will continue until a workshop on Solar Sail that will take place in Glasgow in June 2013.

	Architecture option			
	A ²	B ²	C ²	D ²
Solar pole maximum OZA in one-sidereal rotation period (in deg)	50	40	30	30
Target solar radius (in AU)	0.393	0.447	0.550	0.550
Target orbit period (in years)	0.25	0.30	0.41	0.41
Required sail characteristic acceleration (in mm s ⁻²)	0.2843	0.3103	0.3655	0.5300
Time to target solar radius, i.e. phase 1 duration (in years)	2.44	2.16	1.68	1.16
Phase 2 duration, i.e. time to reach 60 deg. solar latitude (in years)	2.56	2.84	3.32	2.29
Time to 60 deg. solar latitude (in years)	5.00	5.00	5.00	3.45
Time to 90 deg. solar latitude (in years)	6.75	6.94	7.25	5.00

Table 1: Summary and comparison of solar sail mission architecture options and requirements. Superscript number indicates number of trajectory phases.

The preliminary mission analysis assumes that the orbit is reached using a two-phase approach: Phase 1 during which a solar sail is deployed and the orbit radius is reduced; Phase 2 during which the orbit is cranked to 90 degrees. Table 1 provides the results for different options. Table 2 provides the spacecraft and payload characteristics only for option A². With a solar sail of 100 m, a payload mass of 35 kg is enabled within a total launch mass of about 320 kg. Increasing the sail size to 125 m, a payload mass of 60 kg is enabled within a total launch mass of about 500 kg.

Architecture option A ²						
Sail Side Length (m)	S/C Mass Range with 7.5 μ m film (kg)	S/C Mass Range with 2.5 μ m film (kg)	Platform Mass range with 7.5 μ m film (kg)	Platform Mass range with 2.5 μ m film (kg)	Payload Mass with 7.5 μ m film (kg)	Payload Mass with 2.5 μ m film (kg)
50	120 – 107	95 – 83	<i>negative</i>	<i>negative</i>	<i>negative</i>	<i>negative</i>
75	202 – 172	146 – 116	< 13	39 – 69	< 5	10 – 15
100	306 – 255	207 – 156	23 – 73	122 – 172	5 – 15	25 – 35
125	435 – 360	281 – 205	78 – 154	232 – 308	15 – 30	45 – 60

Table 2: Summary of mission architecture option A² mass allowance estimates, including ROM payload allowance estimate.

I.4. INTERNATIONAL COMMUNITY

The French community interested is located in the Paris area (IAS, LESIA, CEA/IRFU). These laboratories are involved in the Solar Orbiter mission and were involved in the study of the Plan A of Solar-C. Additional laboratory that are interested are the Observatoire de la Côte d'Azur and IRAP in Toulouse.

The international community (besides the supporters) interested is the same as the one that proposed several out-of-ecliptic concepts such as Solar Polar Investigation (SPI), POLARIS or the Plan-A of Solar-C. These are namely the Jet Propulsion Laboratory (USA), Stanford University (USA), the Max-Planck institute for Solar System Research, the World Radiation Center (Switzerland) and the NAOJ (Japan). China is also potentially interested since they develop the SPORT mission that is also an out-of-ecliptic mission.

I.5. PROGRAMMATIC ASPECTS

The POLARIS mission was submitted in 2007 as L-class mission to the Cosmic Vision programme of ESA (Appourchaux et al, 2009). It was rejected based on the low technology readiness of the solar sails. The Gossamer road map established by the Working Groups of DLR/ESA aims at the successful deployment of a solar sail system of 50 m square by 2018 (Gossamer-3). Therefore we can reasonably envisage a demonstrated solar-sail technology by 2020 for a potential mission in the framework 2025-2030. Such a mission would require a large international collaboration involving several national agencies (ESA, Japan, US, China). A mission of the L-class seems to be realistic. At the time of writing, there is no programmatic window either in Europe or in the US.

The prospect of having a potential dual-spacecraft solar mission out of ecliptic using solar sail is an exciting opportunity for great scientific return with large discovery-like potential.

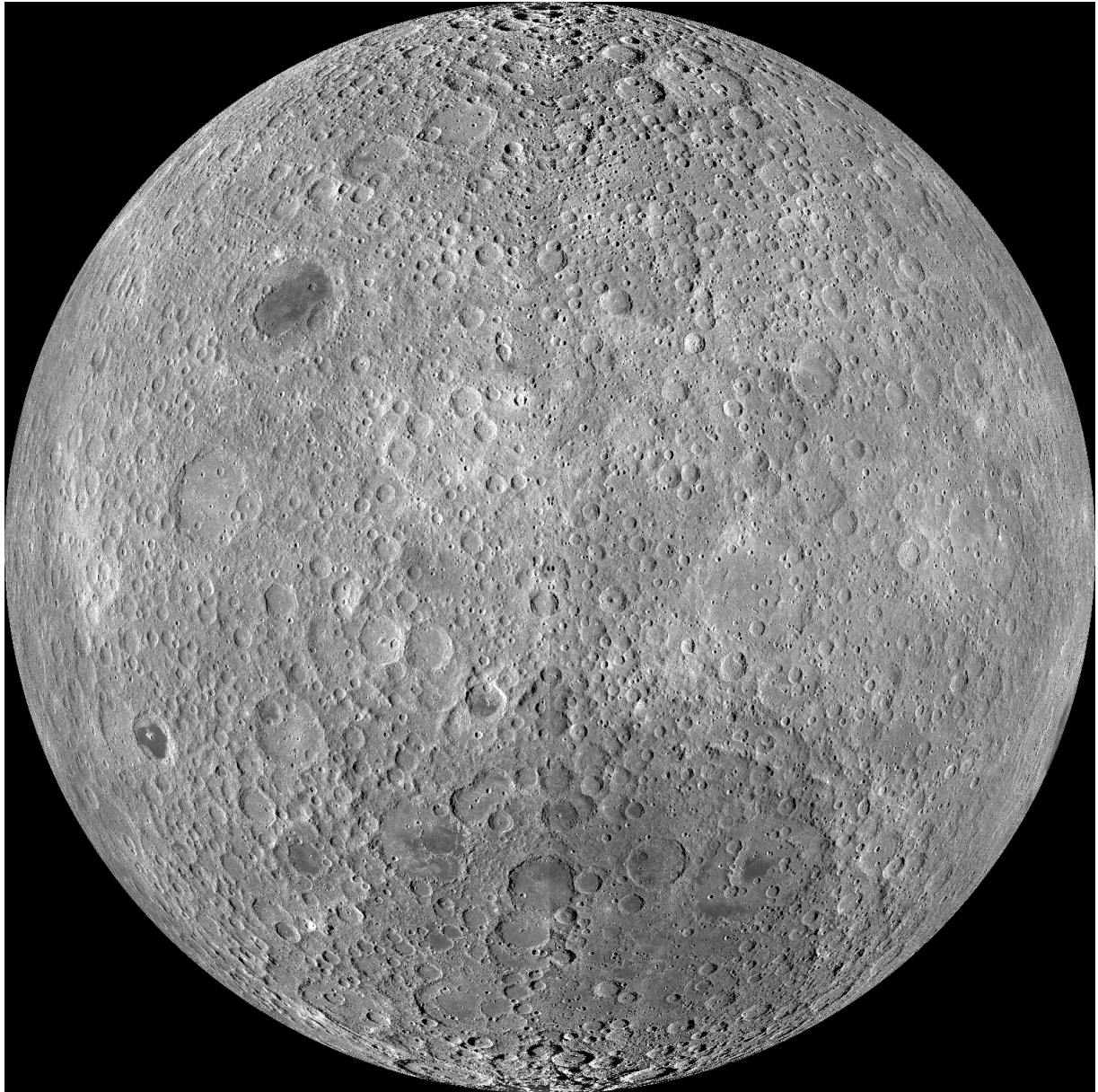
I.6. REFERENCE

- Antonucci, E., Hoeksema, T. and Scherrer, P., 1990, *ApJ*, **360**, 296
- Appourchaux, T. et al, 2009, *Experimental Astronomy*, **23**, 1079
- Auchère, F. et al. 2005, *ApJ*, **625**, 1036
- Auchère, F. et al. 2005, *SPIE*, **5901**, 298
- Gizon, L., Birch, A.C., and Spruit, H.C., 2010, "Local Helioseismology: Three-Dimensional Imaging of the Solar Interior", *ARAA*, **48**, 289
- Halain, J.-P., et al. 2012, *SPIE*, **8443**, 844307

- Liu, Y, et al 2010, *ApJ*, **722**, 1762
- Makela, P. N. et al, 2013, *Sol. Phys.* doi:10.1007/s11207-012-0211-6
- Patsourakos, S. & Vourlidas, A., 2012, *Sol. Phys.*, **281**, 187
- Thompson, M. J. et al., 1996, *Science*, **272**, 1300
- Telloni et al. 2009, *ApJ*, **693**, 1022
- Telloni et al., 2013, *ApJ*, **767**, 138
- Vourlidas, A. et al. 2010, *ApJ*, **722**, 1522

Science from the Farside of the Moon

A response to ESA's call for proposals for Cosmic Vision L2/3 science themes



The farside hemisphere of the Moon, as imaged by NASA's Lunar Reconnaissance Orbiter.

Mark Wieczorek
Institut de Physique du Globe de Paris, France

PROPOSERS

Mark Wieczorek (primary contact); wieczor@ipgp.fr, +33 1 57 27 53 08

Institut de Physique du Globe de Paris, Université Paris Diderot, Case 7071, 75205 Paris Cedex 13, France

Bruce Banerdt	<i>Jet Propulsion Laboratory, Pasadena, California, United States</i>
David Baratoux	<i>Institut de Recherche en Astrophysique et Planétologie, Toulouse, France</i>
Sylvain Bouley	<i>Université Paris-Sud, Orsay, France</i>
Jean-Louis Bougeret	<i>LESIA, Observatoire de Paris-CNRS, Meudon, France</i>
James Carpenter	<i>ESA-ESTEC, The Netherlands</i>
Baptiste Cecconi	<i>LESIA, Observatoire de Paris-CNRS, Meudon, France</i>
Sylvain Chaty	<i>Université Paris Diderot, Institut Universitaire de France, Paris, France</i>
Garreth Collins	<i>Imperial College London, London, United Kingdom</i>
Anthony Cook	<i>Aberystwyth University, Aberystwyth, United Kingdom</i>
Ian Crawford	<i>Birkbeck College London, London, United Kingdom</i>
Sébastien de Raucourt	<i>Institut de Physique du Globe de Paris, Paris, France</i>
Heino Falcke	<i>Radboud Universiteit Nijmegen, Nijmegen, The Netherlands</i>
Vera Assis Fernandes	<i>Museum für Naturkunde, Berlin, Germany</i>
Joachim Flohrer	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Jörg Fritz	<i>Museum für Naturkunde, Berlin, Germany</i>
Raphael Garcia	<i>Institut de Recherche en Astrophysique et Planétologie, Toulouse, France</i>
Robert Grimm	<i>Southwest Research Institute, Boulder, Colorado, United States</i>
Matthias Grott	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Leonid Gurvits	<i>Joint Institute for VLBI in Europe, Delft University of Technology, The Netherlands</i>
Junichi Haruyama	<i>ISAS/JAXA, Sagamihara, Kanagawa, Japan</i>
Harry Hiesinger	<i>Institut für Planetologie, Westfälische Wilhelms-Universität, Münster, Germany</i>
Kazumasa Imai	<i>Kochi National College of Technology, Kochi, Japan</i>
Takahiro Iwata	<i>ISAS/JAXA, Sagamihara, Kanagawa, Japan</i>
Boris Ivanov	<i>Institute for Dynamics of Geospheres, Russian Academy of Sciences, Moscow, Russia</i>
Ralf Jaumann	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Catherine Johnson	<i>University of British Columbia, Vancouver, British Columbia, Canada</i>
Katherine Joy	<i>The University of Manchester, Manchester, United Kingdom</i>
Gunter Kletetschka	<i>Charles University, Prague, Czech Republic</i>
Martin Knapmeyer	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Naoki Kobayashi	<i>ISAS/JAXA, Sagamihara, Kanagawa, Japan</i>
Alexander Konovalenko	<i>Institute of Radio Astronomy, Kharkov, Ukraine</i>
David Lawrence	<i>Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland, United States</i>
Philippe Lognonné	<i>Institut de Physique du Globe de Paris, Paris, France</i>
Daniel Mège	<i>Polish Academy of Sciences, Institute of Geological Sciences, Wrocław, Poland</i>
David Mimoun	<i>Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France</i>
Clive Neal	<i>University of Notre Dame, Indiana, United States</i>
Jürgen Oberst	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Nils Olsen	<i>Technical University of Denmark, Niels Bohr Institute, Copenhagen, Denmark</i>
Gordon Osinski	<i>University of Western Ontario, London, Ontario, Canada</i>
Thomas Pike	<i>Imperial College London, London, United Kingdom</i>
Huub Röttgering	<i>University of Leiden, Leiden, The Netherlands</i>
Hiroaki Shiraishi	<i>ISAS/JAXA, Sagamihara, Kanagawa, Japan</i>
Tilman Spohn	<i>German Aerospace Center, Institute of Planetary Research, Berlin, Germany</i>
Satoshi Tanaka	<i>ISAS/JAXA, Sagamihara, Kanagawa, Japan</i>
Wim van Westrenen	<i>VU University Amsterdam, Amsterdam, The Netherlands</i>
Susanne Vennerstrom	<i>Technical University of Denmark, National Space Institute, Copenhagen, Denmark</i>
Stephanie Werner	<i>University of Oslo, Oslo, Norway</i>
Graham Woan	<i>Glasgow University, Glasgow, United Kingdom</i>
Kai Wuennemann	<i>Museum fuer Naturkunde Berlin, Germany</i>
Philippe Zarka	<i>LESIA, Observatoire de Paris-CNRS, Meudon, France</i>

EXECUTIVE SUMMARY

The farside hemisphere of the Moon is a unique place in the Solar System for a large range of scientific investigations. Being shielded from terrestrial radio-frequency interference, the farside of the Moon is the most radio-quiet environment in near-Earth space. The farside hemisphere faithfully records the primary differentiation of the Moon and hosts the largest recognized impact basin in the Solar System. From the Earth-Moon L2 Lagrange point, the farside hemisphere of the Moon is ideal for the continuous monitoring of meteoroid impacts with the lunar surface. A scientific mission to the Moon in the framework of a Cosmic Vision L-class mission could place **three or four robotic landers** on the farside hemisphere of the Moon and put **an instrumented relay satellite** into a halo orbit about the Earth-Moon L2 Lagrange point. During the course of a 4-year nominal mission, three broad scientific investigations would be conducted.

First, from the vantage point of the lunar surface, **the first extensive radio astronomy measurements in the most radio-quiet region of near-Earth space would be made.** The first sky mapping at low frequencies would be performed, as would be pathfinder measurements of the red-shifted neutral hydrogen line that originates from before the formation of the first stars. Low-frequency radio bursts from our Sun would be quantified, as would auroral emissions from the giant planets in the outer Solar System, pulsars, and the interaction of ultra-high energy cosmic rays with the lunar surface. This radio astronomy experiment would be a pathfinder demonstration for a future large-scale radio array on the farside of the Moon.

Second, this mission would **make precise geophysical measurements of the Moon's interior and measure the composition of its surface.** From seismological, heat flow, and electromagnetic sounding measurements, these data would determine the bulk composition of the Moon, the thickness of its crust, the size and composition of its core, the size of a solid inner core, and the temperature profile of its interior. The surface geochemical data would provide critical ground truth measurements for the interpretation of orbital remote-sensing data sets, and would help decipher the origin of two of the Moon's most prominent geologic provinces: the giant South Pole-Aitken basin and the ancient farside highlands. Geochemical measurements made at the poles would place constraints on the abundances of volatiles, such as water ice, that would be invaluable for future human exploration of the lunar surface.

Third, from the vantage point of the relay satellite, this mission would **quantify near-Earth impact hazards by continuously monitoring the farside of the Moon for meteoroid impacts.** Unspoiled by Earthshine and an intervening atmosphere, by the detection of impact flashes, this experiment would measure the Earth-Moon impact flux, the size-frequency distribution of impactors in near-Earth space, and spatial and temporal variations in the lunar impact rate during the lunar night. The measured impact times and locations would be used as known seismic sources for the seismology experiment, allowing for highly detailed interior modeling.

A mission to the farside of the Moon would develop soft-landing capabilities on airless bodies, and would benefit from existing state-of-the-art geophysical and astronomical instrumentation in Europe. The scientific objectives of such a mission are supported jointly by the radio astronomy and lunar science communities, address directly all four of the top-level themes of ESA's Cosmic Vision program (Bignami et al. 2005), and are identified as top priorities in the United States planetary science decadal survey (Committee on the planetary science decadal survey, 2010). An ESA-led mission to the farside of the Moon would demonstrate European leadership in the renewed international interest in exploring Earth's nearest celestial neighbor.

1. INTRODUCTION

The past decade has seen a dramatic surge in the interest of exploring Earth's only natural satellite. Orbital missions to the Moon have included spacecraft from ESA (SMART-1), Japan (Kaguya), India (Chandrayaan-1), China (Chang'e-1 and 2) and the United States (Lunar Reconnaissance Orbiter, and Gravity Recovery and Interior Laboratory, GRAIL). Two missions included impact experiments (Chandrayaan-1, and NASA's Lunar Crater Observation and Sensing Satellite, LCROSS), and several missions are planned for launch in the immediate future (Chang'e-3, Chandrayaan-2 and NASA's Lunar Atmosphere and Dust Environment Explorer, LADEE).

Given the huge amount of data that has been collected from lunar orbit, it is surprising that there are still so many fundamental questions about the Moon that remain unanswered. The origin of the Earth-Moon system is currently being debated (Ćuk and Stewart 2012; Canup 2012), the lunar bulk composition of important refractory elements are unknown to within a factor of two (Taylor et al. 2006), the origin of magnetic anomalies is highly contested (Hood and Artemieva 2008, Wieczorek et al. 2012), and the size, composition, and state of the core are largely unconstrained (Neal 2012). The cratering record of the Moon has suggested a dramatic spike in the impact cratering rate about 4 billion years ago, perhaps associated with a reorganization of the orbits of the outer giant planets (Tsiganis et al. 2005), but the timing and magnitude of this so-called "late heavy bombardment" is fraught with large uncertainties (Norman 2009). Analyses of lunar samples led to the idea that many planets possessed "magma oceans" during their early stages of accretion and differentiation, but the details of the crystallization stage and the applicability of this paradigm to other rocky planets are still largely unexplored (Grove and Krawczynski 2009).

Though the exploration of the Moon from orbit is now largely completed, the exploration and utilization of the Moon from its surface has not progressed since the end of NASA's Apollo program in 1972 and the end of the Russian Luna program in 1976. Furthermore, it is now recognized that all of the Apollo landings were made near an anomalous geological province, and that the inferences drawn from these sites are probably not representative of the Moon as a whole. Further progress on unlocking the Moon's secrets will be made only by returning to its surface with scientific instruments.

From the surface of the Moon, a number of pioneering and innovative measurements could be made. Pioneering radio astronomy measurements would be made at frequencies inaccessible to terrestrial observatories. From a station on the farside, in the most radio-quiet region of near-Earth space, the first sky mapping at low frequencies would be performed, and pathfinder measurements of signals originating from before the formation of the first stars in our universe would be quantified. Next-generation geophysical measurements would illuminate the interior structure of the Moon, and geochemical measurements of its surface would unlock the secrets of terranes not visited by the Apollo and Luna missions. From a lander at one of the poles, the abundance and composition of near surface volatiles, including water ice, would be characterized, resources that are certain to play an important role in the return of humans to the Moon. Finally, from the monitoring of the lunar surface from orbit, flashes generated by meteoroid impacts would be detected, quantifying not only impact hazards in near-Earth space, but also providing times and locations of seismic sources for use with investigations of the Moon's interior structure.

The science objectives of this proposal for an ESA L-class mission are conceived to exploit the unique environment offered by the farside hemisphere of the Moon, and address directly all four of the top-level themes of ESA's Cosmic Vision program (Table 1). Three primary investigations are dictated by the properties of this platform (Mimoun et al. 2010).

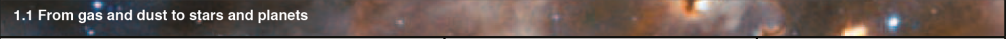
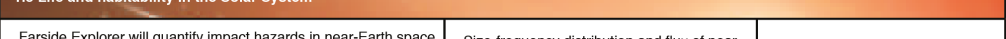





Theme	Farside Explorer Relevance to CV theme	Measurement Objective	Instrument
Theme 1 <i>What are the conditions for planet formation and the emergence of life?</i>	1.1 From gas and dust to stars and planets 		
	Farside Explorer will investigate the late stages of planetary formation, including the event that formed the Earth-Moon system and the impact that formed the South Pole-Aitken basin.	Bulk composition of the Moon, crustal thickness, core size, internal structure, surface geology	Seismometer, heat flow probe, electromagnetic sounder, orbiting magnetometer, surface camera, gamma-ray spectrometer
	1.3 Life and habitability in the Solar System 		
Theme 2 <i>How does the Solar System work?</i>	Farside Explorer will quantify impact hazards in near-Earth space, elucidate the consequences of giant impact events that could have frustrated the development of life, and detect volatile elements in the regolith.	Size-frequency distribution and flux of near-Earth objects, crustal thickness within the South Pole-Aitken basin, composition and abundance of volatiles	Impact flash camera, seismometer, geochemistry experiment
	2.1 From the Sun to the edge of the Solar System 		
	Farside explorer will measure low-frequency radio emissions from the Sun, uncontaminated by terrestrial radio-frequency interference.	Low-frequency radio monitoring of the Sun	Radio astronomy receiver
	Farside Explorer will quantify how the solar wind interacts with atmosphereless bodies.	Electromagnetic measurements from the surface and orbit	Electromagnetic sounder, orbiting magnetometer
	Farside Explorer will investigate how planets differentiate into a crust, mantle and core, and how tectonic processes work on single-plate planets.	Internal structure of the Moon, surface heat flow	Seismometer, heat flow probe, electromagnetic sounder, orbiting magnetometer
	2.2 The giant planets and their environments 		
	Farside Explorer will measure the magnetospheric emissions of the giant planets, their time variations, and the coupling with their satellites.	Low-frequency radio monitoring of the outer planets, uncontaminated by solar and terrestrial emissions	Radio astronomy receiver
Theme 3 <i>What are the fundamental physical laws of the Universe?</i>	2.3 Asteroids and other small bodies 		
	Farside Explorer will constrain the flux, size-frequency distribution, and physical properties of small near-Earth objects.	Monitoring of impact flashes in the visual and infrared domains	Impact flash camera, seismometer
Theme 4 <i>How did the Universe originate and what is it made of?</i>	3.3 Matter under extreme conditions 		
	Farside Explorer will detect interactions between ultra high energy cosmic rays and the lunar surface.	Low-frequency radio measurements	Radio astronomy receiver
Theme 4 <i>How did the Universe originate and what is it made of?</i>	4.1 The early Universe 		
	Farside Explorer will investigate the cosmological dark ages through the red-shifted neutral hydrogen 21-cm line.	Low-frequency radio measurements uncontaminated by solar and terrestrial emissions	Radio astronomy receiver

Table 1. Scientific relevance of an L-class *Farside Explorer* mission to the Cosmic Vision science objectives.

1. The farside of the Moon is a unique environment for low-frequency radio measurements. On Earth, several sources of radio-frequency interference exist, such as radio broadcasts and lightning. Furthermore, extra-terrestrial signals with frequencies below 10-20 MHz cannot be studied since they are reflected off of, or are severely distorted by, the Earth's ionosphere. The farside of the Moon does not suffer from either of these problems given that the Earth and its orbiting communications satellites are not in direct view from this platform. It is the ideal place to investigate one of the last unexplored regions of the electromagnetic spectrum (Jester and Falcke 2009).

From the vantage point of the lunar farside, even a single dual-polarized antenna would make extraordinary measurements: Low-frequency radio bursts from the outer giant planets of our Solar System would be monitored, the red-shifted neutral hydrogen (HI) line that originates from before the formation of the first stars would be investigated, and the interaction of the Moon's surface with ultra-high energy cosmic rays that exceed the energies of modern particle accelerators would be detected. Given the broad interest in the creation of a low-frequency aperture synthesis radio array on the Moon's surface, such pioneering low-frequency interferometric radio measurements would demonstrate the feasibility of such a later larger-scale project.

2. The farside of the Moon is a unique laboratory for investigating planetary formation and evolution. The Earth-Moon system is believed to have formed by a giant impact between the nascent Earth and a Mars-sized object. This event led to the formation of globe-encircling magma oceans on both bodies and contributed to their primary differentiation and core formation. In contrast to the Earth, the interior structure and composition of the Moon have evolved little since this time. Both the Earth and Moon suffered subsequently the consequences of large impact events, but this important period of time has been almost totally erased from the Earth's surface, whereas it is ideally preserved

on the Moon. The Moon is an end-member of terrestrial planetary evolution and is the nearest celestial object to have preserved the record of early planetary processes.

Though the geophysical investigations of the Apollo era provided tantalizing clues to lunar formation, differentiation, and evolution, in retrospect, the Apollo nearside landing sites were not ideal for this purpose. As shown in Figure 1, it is now recognized that these measurements are biased by their proximity to an atypical geological province called the Procellarum KREEP Terrane (Jolliff et al. 2000). The most prominent terrane that records the formation of the primordial crust is located largely on the farside hemisphere, as is the largest and oldest recognized impact basin in our Solar System, the South Pole-Aitken basin. The farside hemisphere of the Moon is the best place to investigate the relics of planetary differentiation that are recorded beneath its surface.

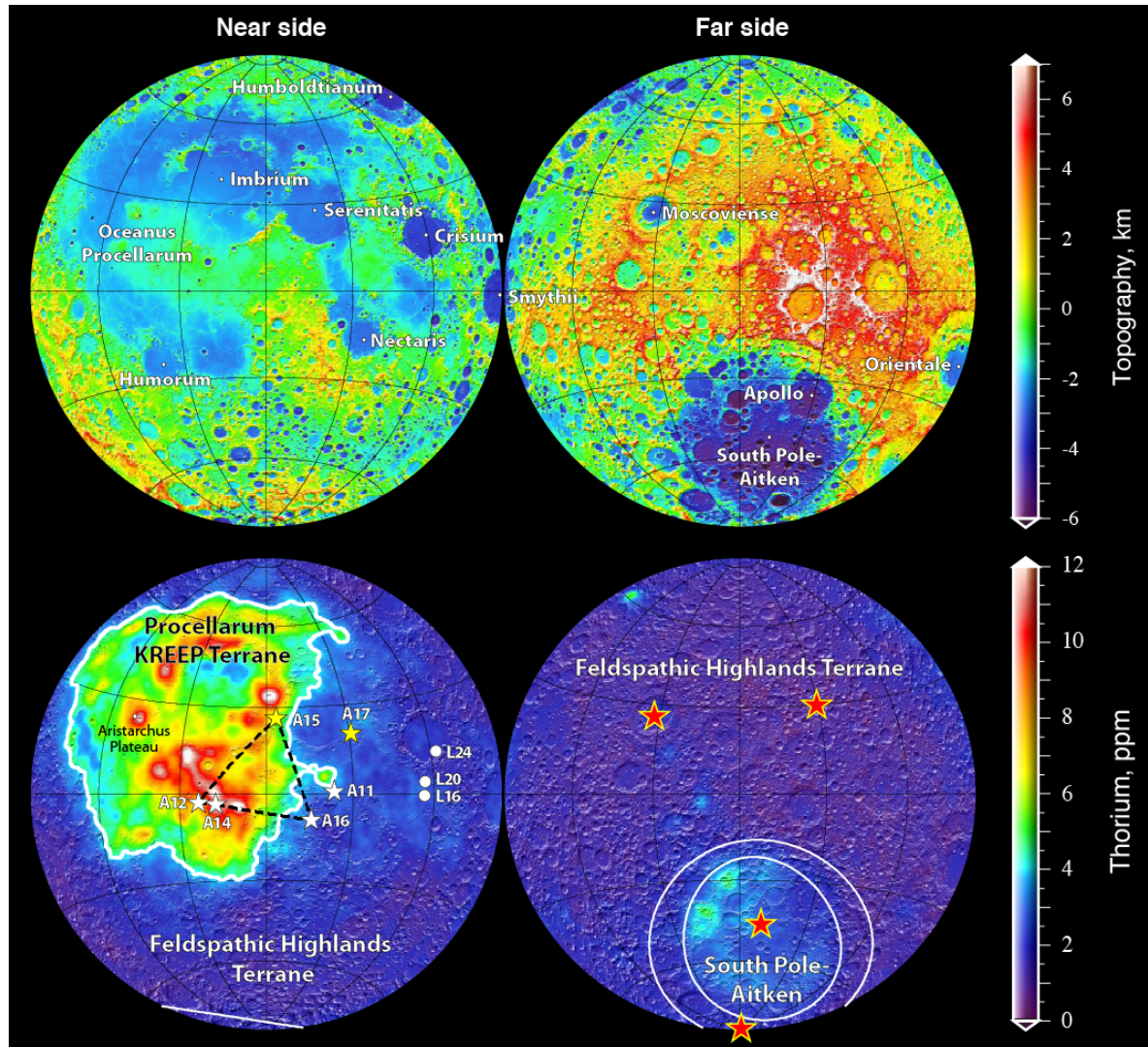


Figure 1. (top) Topography of the Moon from the Lunar Reconnaissance Orbiter altimeter (Smith et al. 2010) and (bottom) surface abundance of thorium from the Lunar Prospector gamma-ray spectrometer (Lawrence et al. 2004). The thick white line delineates the confines of the nearside Procellarum KREEP Terrane, and the thin white ellipses on the farside outline the floor and structural rim of the South Pole–Aitken basin. Stars and circles on the nearside represent the Apollo and Luna sample return stations, respectively. Yellow stars correspond to the locations where two heat-flow measurements were made, and the black dashed lines connect the four Apollo stations containing seismometers that operated in a network manner. The ALSEP geophysical network covered only a small portion of the nearside hemisphere that is highly atypical. The four red stars on the farside hemisphere mark high science priority landing sites for a Cosmic Vision L-class mission, which include the center of the South Pole–Aitken basin, the South Pole, the interior of the Moscoviense basin where the crust is nearly absent, and the ancient farside highlands.

3. Impacts between near-Earth objects and the Moon are ideally monitored from the Earth-Moon L2 Lagrange point. From the vantage of the L2 Lagrange point, about 60,000 km above the lunar surface, a camera could make measurements of the visible and thermal flashes generated by meteoroid impacts during the lunar night that would be superior to those made on the nearside hemisphere from Earth (Ortiz et al. 2006, Suggs et al. 2008). These observations would not be affected by local weather conditions, and given the lack of an intervening atmosphere, critical near-infrared observations would be possible. Since there is no Earthshine light pollution on the farside of the Moon, and since L2 is significantly closer to the Moon than is the Earth, impact flashes would be detectable for objects considerably smaller than could be seen from a comparable optical system on the Earth's surface.

In combination with simultaneous seismic measurements on the surface, these impact flash measurements will make it possible to constrain both the present day impact flux and size-frequency distribution of centimeter to meter sized near-Earth objects. Through long-term monitoring, both temporal and spatial variations in the impact flux on the Moon would be quantified (Le Feuvre and Wicczorek 2011), allowing for a better understanding of the origin of these objects.

2. SCIENCE OBJECTIVES

The next major step in lunar exploration will be made from its surface. Here, we describe three major science investigations that could be accomplished by two or more stations operating in a network manner: (1) radio astronomy, (2) lunar science, and (3) the quantification of Earth-Moon impact hazards. Though first-rate science could be obtained from landers located on the nearside hemisphere of the Moon, the most rewarding scientific returns would come from placing landers on the Moon's farside hemisphere.

A number of high priority science targets exist for the landing sites on the farside of the Moon. These include: the center of the South Pole-Aitken basin, which is the largest and oldest impact basin on the Moon; the South Pole, close to one of the regions that are permanently shadowed; the center of the Moscoviense basin, where the crust is almost entirely absent; and the feldspathic highlands that formed from the crystallization of the lunar magma ocean.

Radio Astronomy

The Moon has an extremely thin ionosphere that allows radiofrequency measurements down to 500 kHz during the day, and even lower during the night (Vyshlov et al. 1976). Its farside hemisphere is shielded from terrestrial radio interference, satellite interference, lightning, and auroral emissions, and during the lunar night the farside is also shielded from strong solar radio emissions (Figure 2). For these reasons, the farside of the Moon has been considered to be an ideal site for low-frequency radio astronomy since the 1960s (Bely et al. 1997). The two Radio Astronomy Explorer satellites launched in 1968 and 1973 are the only spacecraft to have made low frequency radio measurements in the frequency range of 0.02 to 13.1 MHz (Alexander et al. 1975). From the collected data (total flux only), these spacecraft could study only solar, Jovian and terrestrial radio emissions (Kaiser et al. 1977).

Astrophysics at frequencies below 10-20 MHz is today almost entirely unexplored. Though solar and planetary studies can be performed from spacecraft above the Earth's ionosphere, these observations are made in permanent view of strong sources from the Sun, Jupiter, and Earth. No dedicated measurements exist from a place as quiet as the lunar farside. Being an unexplored portion of the electromagnetic spectrum (Figure 3), low-frequency radio measurements will elucidate many processes. A few investigations enabled by a radioastronomy receiver on the farside of the Moon include the following.

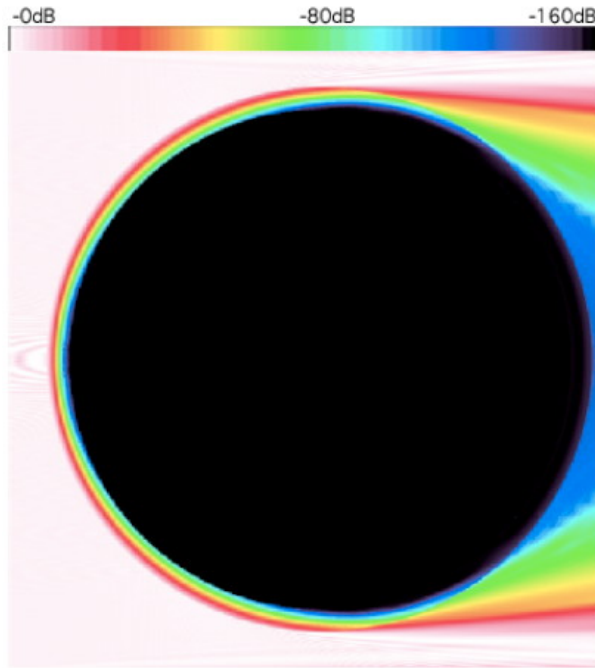


Figure 2. Attenuation of a 60 kHz continuous plane radio wave (incident from the left), as it propagates through and around the Moon. On the farside, at latitudes equatorward of 45°, strong radio signals originating from the Earth, Sun, or Jupiter, would be attenuated in strength by more than a factor of 10^{10} (Takahashi 2003).

Cosmology. The frequency range less than 40 MHz corresponds to the red-shifted neutral hydrogen (HI) 21-cm line at redshifts $z \geq 35$, and thus to the so-called “dark ages” that preceded the epoch of reionization in the Universe (Pritchard and Loeb 2010).

Sky mapping. Combining the time-variable occultation of one-half of the sky by the Moon, goniopolarimetric measurements by each lander, and inversion of series of interferometric measurements, radio sky maps could be built at several frequencies together with an accurate spectrum of the sky background intensity and polarization (Dulk et al. 2010).

Solar physics. The frequency range of 0.1-40 MHz corresponds to plasma frequencies in the solar corona between about two and several tens of solar radii from the center of the Sun. As solar type II and III radio bursts are emitted at 1-2 times the plasma frequency, occurrence and evolution of these structures would be monitored throughout the solar corona (Mann et al. 1999).

(Exo-)planetary magnetospheres. The detection of magnetospheric radio emissions from all giant planets would be possible on a regular basis from the surface of the Moon. Long-term observations would permit fundamental studies such as the accurate determination of planetary rotation periods and their variations. These measurements would set the context for future exoplanet low-frequency radio searches (Zarka et al. 2007, 2012).

Pulsars and radio propagation. Due to their periodic nature, radio emissions from a few known intense pulsars should be detectable. These challenging observations would provide the first measurements of a few pulsar spectra down to very low frequencies, and allow us to quantify propagation effects down to increasingly lower frequencies and constrain emission models (Bruck et al. 1976, Popov et al. 2006).

Transient events. Radio transient events produced by the interaction of ultra high-energy cosmic rays (UHECRs) with the Moon’s surface would be investigated by a radio astronomy experiment on the surface. High-energy neutrinos would behave like UHECRs, but with a much deeper penetration depth and detector volume (Jester and Falcke 2009).

The Extragalactic Universe. Star-forming galaxies, clusters of galaxies and active galactic nuclei are major ingredients of the visible universe that can also be extremely luminous at ultra-low frequencies. Observations from the radio quiet platform offered by the farside of the Moon would constrain the physical accretion and ejection mechanisms that lead to the relativistic gas streams and high magnetic fields associated with active galaxies, but also in accreting black holes and neutron stars within our own Galaxy.

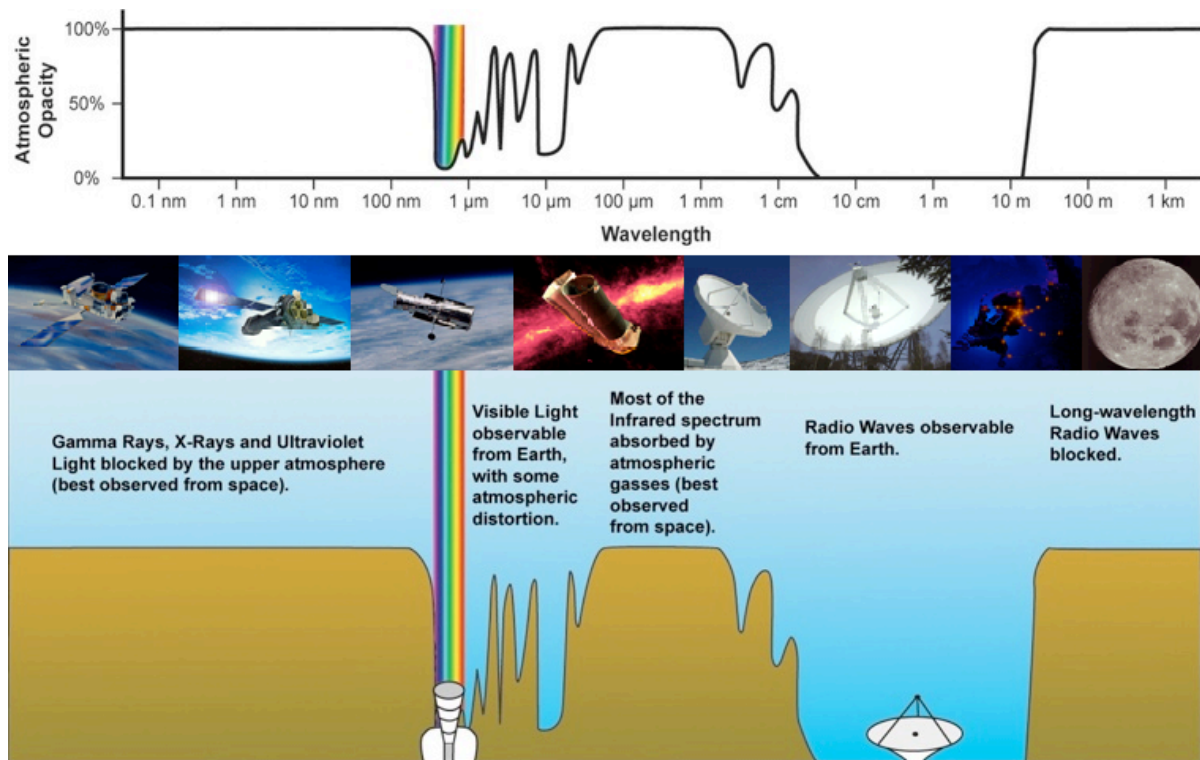


Figure. 3. Atmospheric and ionospheric effects allow only a small portion of the electromagnetic spectrum to be investigated from Earth.

Local environment studies. A surface radio astronomy experiment would monitor the peak plasma frequency above the lander as a function of lunar local time, solar activity, and traversal of the Earth's magnetotail. Charging of the lunar regolith due to ultra-violet illumination, charged particles, or micrometeoritic impacts could result in detectable electrostatic discharges.

Pathfinder measurements for a lunar array. A longer-term goal in radio astronomy is the construction of a large low-frequency aperture synthesis radio array on the Moon's farside hemisphere. Radioastronomy measurements from the farside would provide invaluable pioneering measurements that will influence the design of such an array.

Lunar Science

The Moon is the only terrestrial object for which we have samples from known locations, geophysical data from dedicated stations on the surface, and observations from field geologists. From these data, the origin of the Moon from a giant impact with the early Earth, the existence of a globe-encircling magma ocean that crystallized into an ancient primary crust, the existence of distinct geologic terranes, and a three-billion year record of volcanic activity have been elucidated (Jolliff et al. 2006, and references therein).

Geophysical measurements of the lunar interior were made from the Apollo Lunar Surface Experiment Package (ALSEP) that included a passive seismic experiment, heat flow probes, magnetometers, and laser retroreflectors. However, because of the limited extent of the ALSEP network, limitations of late 1960s digital technology, and the unfortunate placement of these stations near the boundary of the two most prominent geologic terranes (Figure 1), the interior structure and early geologic evolution of the Moon remains elusive (Figure 4). Geophysical experiments on the lunar surface would enable the following science investigations.

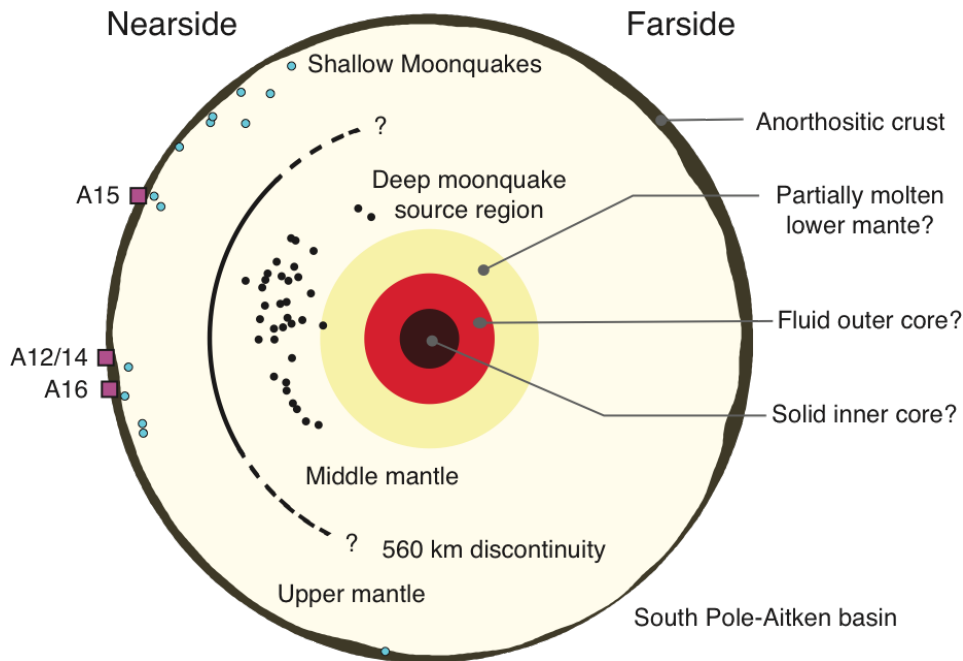


Figure 4. Schematic diagram (to scale) of the Moon's interior structure. Shown are the Apollo seismic stations (squares), all shallow moonquakes (blue circles), the deep moonquake source regions that are periodically activated by Earth-raised tides (black circles), inferred variations in crustal thickness, and a possible discontinuity ~500 km below the surface. The structure below the deep moonquake source region is only weakly constrained, and little is known about the farside hemisphere. Image from Wieczorek et al. (2006).

Interior structure. This thickness of the crust is a key parameter for understanding lunar differentiation, but seismic estimates beneath the Apollo network range from 30 to 60 km. An apparent seismic discontinuity was reported about 500 km below the surface, and this had been interpreted to mark the base of the lunar magma ocean. However, some recent models without discontinuities are compatible with the data, and the limited footprint of the Apollo seismic network leaves open the question as to whether such a discontinuity is a widespread feature of the mantle (Lognonné and Johnson 2007). Many studies have attempted to determine if the Moon has a metallic core (Weber et al. 2011, Garcia et al. 2011), but its size, state (molten or solid) and composition are still poorly constrained, and the existence of a solid inner core remains controversial.

Bulk Composition. Previously reported bulk silicate compositions have suggested that the Moon is either similar in composition to that of the Earth, or enriched in refractory elements by a factor of about 2 (e.g., Taylor et al. 2006). Better determination of the geophysical parameters of the lunar core, mantle, and crust would provide improved bulk compositional constraints that could be used to test hypotheses regarding lunar formation (e.g., Wieczorek et al. 2013).

Thermal evolution. To decipher the Moon's magmatic and thermal evolution, it is necessary to know the quantity of heat that is currently escaping its surface. The surface heat flow provides important constraints on the Moon's internal temperature and bulk composition, and is a critical measurement that any thermal evolution model must satisfy. Two heat flow measurements were made during Apollo, but these were later found to be made in an atypical region on the boundary of two major terranes: one enriched in heat producing elements, and the other depleted (Wieczorek and Phillips, 2000). Representative heat flow measurements from each of the major lunar terranes are required.

Magnetic field. Portions of the lunar crust are associated with strong magnetic anomalies and some of the lunar samples possess remanent magnetization. After decades of study, the origin and timing of the fields that magnetized the crust, and the origin of the magnetic carriers, are still being

debated (Suavet et al. 2013, Wieczorek et al. 2012). A magnetic field observatory on the surface of the Moon would not only help in deciphering the magnetic properties of the crust, but would also allow electromagnetic sounding of the lunar interior, providing electrical conductivity and temperature profiles of the deep interior (Hood and Sonnett 1982).

Surface composition. The samples returned by the Apollo and Luna missions are well documented and have provided a wealth of data, but only from a few locations on the surface. Global remote sensing data have shown that there are rock types in both the highlands and South Pole-Aitken basin that are not found in the Apollo and Luna sample collections. In situ “ground truth” compositional measurements, such as from a combined gamma-ray and neutron spectrometer, would provide essential clues to deciphering the origin of these rock types.

Lunar outgassing and volatiles. The Moon presents unique challenges in resource detection, recovery, and utilization. An important challenge is the detection of volatiles and the study of how they are transported in the tenuous lunar exosphere (Lucey 2009). The production and transport of the short-lived radioactive element radon is not only a valuable tracer of lunar regolith-exosphere exchange processes, but can also be used to monitor the current degassing activity of the Moon and to localize active geodynamic regions.

Earth-Moon Impact Hazards

Asteroids and comets derived from beyond the orbit of Mars occasionally encounter the Earth. The vast majority of these meteoroids are small and burn up in the atmosphere, but a smaller number of larger objects pass through the atmosphere to make an impact crater on the surface. Such events can be catastrophic to life, either locally or globally. Two primary objectives of investigating impact hazards are to determine the size-frequency distribution of these objects, and to determine the probability that they will collide with the Earth.

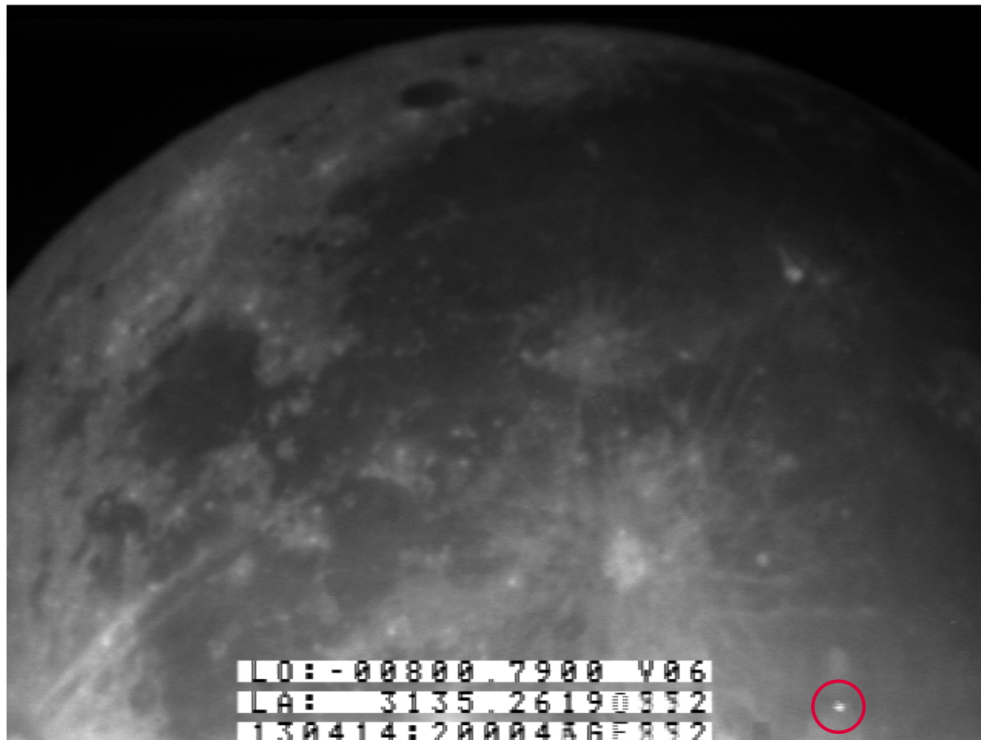


Figure 5. A 0.1 second impact flash (circled in red) detected from one of the telescopes dedicated to their monitoring in Morocco (Baratoux et al., 2012). Today, several hundred impact flashes have been observed on the Moon. In this image taken during the lunar night, the nearside surface of the Moon is illuminated only by Earthshine.

The Moon is subjected to the same flux of impactors as the Earth, but unlike the Earth, where small bodies are destroyed in the atmosphere before hitting the ground and where impact structures are rapidly eroded, its surface is an excellent recorder of collisions with objects at all sizes. The present flux of impactors gives rise to new impact craters that can be imaged from orbit, and the impact itself produces a visible and thermal emission (an impact flash) that is detectable from Earth or lunar orbit, as shown in Figure 5 (Bouley et al. 2012). Observation of the Moon's surface thus provides a complementary approach to understanding the flux and size distribution of objects in near-Earth space, as well as providing the time and location of seismic sources that could be used in investigations of the Moon's interior structure.

The monitoring of impact flashes on the nearside hemisphere of the Moon (during the lunar night) can be accomplished from the Earth (Ortiz et al. 2006, Suggs et al. 2008), but as these observations are affected by local weather conditions, a network of telescopes across the globe is required. To achieve global coverage, measurements of the farside would be required from a dedicated space-based camera. With a lack of obscuring weather, continuous monitoring of the lunar farside meteoroid flux could be made from the vantage point of the Earth-Moon L2 Lagrange point. These observations would be ideal because there is a complete lack of Earthshine illuminating the unlit portions of the Moon's farside hemisphere, and the lack of an intervening atmosphere would allow for continuous infrared measurements. Combined with the proximity of L2 to the lunar surface, these observations would allow for the detection of much smaller impacts than could be accomplished from a comparable terrestrial observatory. From these measurements, a more precise picture of impact processes in the Earth-Moon neighborhood would be obtained.

Considering the expected detections of several hundreds of events per year, the main goals of an impact monitoring program would be to: (1) determine seismic source locations and times for a seismology experiment on the surface, (2) obtain the average impact rate on the Moon and quantify spatial and temporal variations, and (3) determine the optical-magnitude impact-frequency relationship. Combining the measured optical magnitudes with simultaneous data from a seismology experiment would offer a unique opportunity to explore the partitioning of impactor kinetic energy into seismic and thermal energy at scales that are impossible to reach in laboratory. If a camera were able to image the newly formed crater, this would provide invaluable information regarding the relationship between the projectile's kinetic energy and final crater diameter.

3. MODEL PAYLOAD

The *Farside Explorer* mission concept consists of two essential components: an instrumented relay satellite to be inserted into a halo orbit about the Earth-Moon L2 Lagrange point, and several identical spacecraft that make soft landings on the lunar surface. The landers would contain a suite of state-of-the-art instrumentation: a radio astronomy receiver, long- and short-period seismometers, a heat flow probe, an electromagnetic sounder, surface cameras, and a geochemical/volatiles experiment. Though several possibilities exist for the geochemistry experiment, a combined gamma-ray and neutron spectrometer would benefit from the long duration spent on the surface. If excess mass is available, other experiments, such as a ground penetrating radar, could be considered. A robotic arm is included on each lander for deployment of the seismometer and heat flow experiment, but lighter alternatives with fewer degrees of freedom are also available.

The relay satellite would contain an impact flash camera, and a magnetometer. Radio astronomy interferometric correlations could be performed on the orbiter in order to reduce the amount of data transmitted to Earth. As shown in Figure 6, the surface and orbital experiments are synergistic. The times and locations of impact events on the lunar surface obtained from the orbiting camera would be used as known sources for the seismic experiment. By monitoring the magnetic fields on the surface of

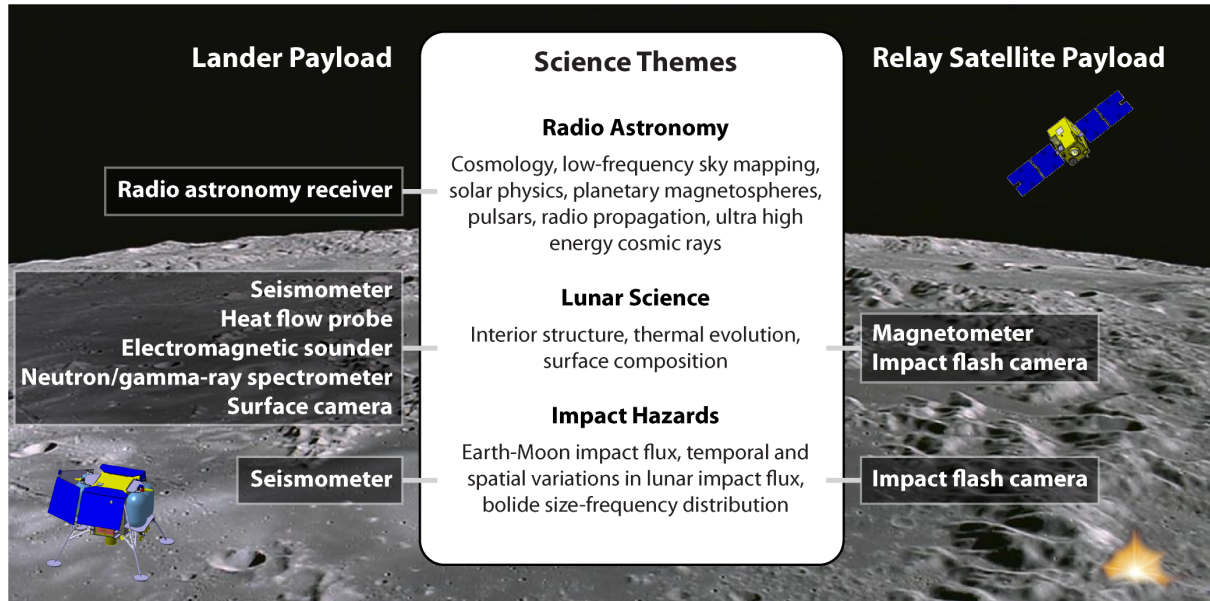


Figure 6. Relationship of Farside Explorer science themes to the lander and orbiter payload. Lunar science and impact hazard objectives are addressed by synergistic measurements from the surface and orbit.

the Moon and from orbit, it would be possible to determine an electrical conductivity profile of the Moon from the crust to core.

Much of the proposed payload has already a high-degree of maturity or flight heritage. As examples, a seismometer is currently being built at IPGP (France) and Imperial College/Oxford (UK), and a heat flow probe is being built at DLR (Germany), all for launch on NASA's next Discovery class mission to Mars (InSight). Components of the radio astronomy experiment have numerous predecessors (WIND, Ulysses, Cassini, STEREO, and BepiColumbo), as do the magnetometers (Osted, Astrid-2, CHAMP, SAC-C, and PROBA-2) and cameras.

Current best estimates for the mass, power, and data rates of the payload elements are given in Table 2. The mass of the full science payload on the lander is approximately 33 kg (including 20% margin). If spacecraft mass margins require, portions of this payload could be descoped with only marginal impact on the science. For example, one heat flow probe could be flown instead of two on each lander, 1 axis of the very broad band seismometer could be removed from each lander, and the electrical sounding aspects of the electromagnetic sounding experiment could be removed. The mass of a reduced payload that meets all the science objectives is approximately 25 kg. To conserve energy during the night (if solar panels and batteries are used), many of the instruments can be turned off, or put into a low power hibernation mode.

Lander Payload	Remark	Mass (kg)		Avg Power (W) with 20% margin		Data Rate Gbits/Lun
		Current best estimate	with 20% margin	Day	Night	
Robotic Arm		5.8	7.0	-	-	-
Radio astronomy receiver	3 dipoles	1.5	1.8	1.2	1.2	3.3
Electromagnetic sounding	Fluxgate, search coil, 5 electrodes	7.2	8.6	4.0	0.1	1.6
Seismometer	3 long period and 3 short period	7.0	8.4	2.6	1.6	6.3
Heat flow probe	2 moles	2.8	3.4	3.4	0.3	0.0
Surface camera		1.0	1.2	1.5	0	1.0
Neutron/Gamma-ray spectrometer		2.5	3.0	4.0	0	0.1
Total		27.8	33.4	16.7	3.2	12.3
Satellite Payload						
Impact flash camera		10.0	12.0	33.6		1.6
Magnetometer	50 Hz sampling	1.4	1.7	3.0		16.0
Total		11.4	13.7	36.6		17.6

Table 2. Potential lander and satellite payload, and current best estimates of the mass, power and data rates.

4. MISSION PROFILE

The Farside Explorer space segment includes three or four spacecraft to land on the farside of the Moon, an instrumented relay satellite, and the launcher (an Ariane 5). The proposed mission concept is innovative by using a halo orbit about the Earth-Moon L2 Lagrange point (LL2) to provide a communications relay to the farside landers while simultaneously enabling the impact flash monitoring program.

As previous lunar lander studies have proven to be marginal in mass (such as the Astrium Satellites “MoonTwins” study), we instead propose a longer but more mass-efficient trajectory to the Moon, similar to that used by NASA’s GRAIL mission. The spacecraft are first injected into a geosynchronous transfer orbit (GTO) and then fly along a weak stability boundary (WSB) trajectory that goes first to the Earth-Sun L1 Lagrange point (EL1) and then to the Earth-Moon L2 (see Figure 7). This approach offers several advantages over a direct trajectory: it allows several launch windows per month with a similar ΔV , it requires less post-launch ΔV , and it provides a longer cruise time for calibrations, readiness tests, and flight system performance.

Following insertion into GTO, two staging options are considered:

1. Each spacecraft includes the propellant for the rest of the mission (transfer to EL1, LL2 and landing for the probes; transfer to EL1 and LL2 for the relay satellite).
2. The upper lander and relay satellite are bound together during the Earth-Moon cruise as a composite. After LL2 halo orbit injection, the satellite and lander are separated.

The second strategy allows for a very simple bus for the relay satellite as it does not require a propulsion stage.

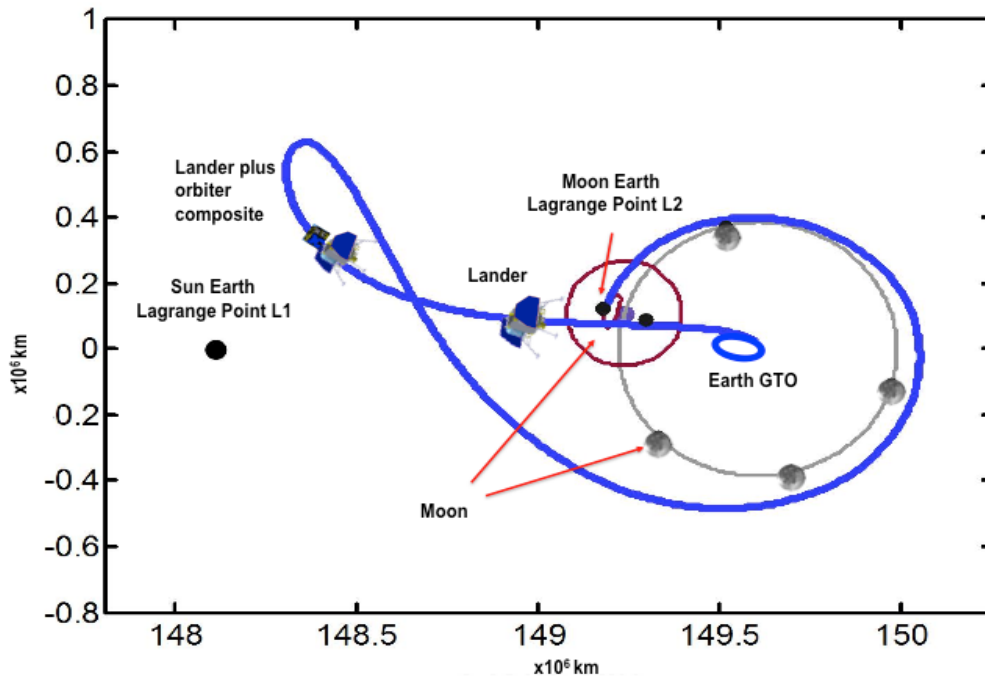


Figure 7. Low ΔV trajectory to the Moon. After insertion in a geosynchronous transfer orbit, the two spacecraft are injected into a 150-day weak stability boundary trajectory along the Earth-Sun L1 gravity manifold. This trajectory returns to the vicinity of the Moon, allowing insertion into the Earth-Moon L2 halo orbit. Modified after Parker (2006).

From GTO, the lunar probes are injected one after another into a 150-day ballistic trajectory along the EL1 gravity manifold. The proposed trajectory uses the instability of the manifold next to EL1 to return to the vicinity of the Moon. At lunar arrival, the intersection of the manifolds of the Earth-Sun and Earth-Moon system allows for the insertion, at a very low ΔV , of the mission elements into an Earth-Moon LL2 halo orbit. The LL2 halo orbit will be the quasi-final orbit for the relay satellite, but only a waiting orbit for the landers. The spacecraft LL2 waiting orbit allows the access to any location on the lunar surface with a single large maneuver (the braking burn).

The chosen operational halo orbit for the relay satellite results from a compromise between several mission requirements: use for a relay satellite with almost constant visibility of the farside landing sites from Earth, implementation of the lunar impact flash and magnetic field payloads, minimization of maintenance requirements, and low cost injection from the transfer orbit. Several LL2 halo orbits are possible, and for operational reasons, a 2:1 resonant orbit (14 days period) is preferred with altitudes that vary between about 30,000 and 60,000 km. For this pseudo-orbit, the relay satellite would be in constant visibility of the equatorial landers and would be in visibility of the polar landers 90% of the time. A slightly higher altitude orbit is possible that would enable a constant visibility of the two landers.

The landing strategy is to follow the unstable WSB down to the surface and to set to zero the relative speed with respect to the landing site. Ground control would monitor constantly the spacecraft landing through the relay satellite. As a result of the waiting orbit strategy, the control of only one probe at a time is required for the critical landing phase. In order to simplify the probe guidance, navigation and control (GNC), no precision landing capabilities are foreseen. The American and Soviet space programs have demonstrated that precision landing and hazard avoidance capabilities are not required to safely land a robotic spacecraft on the Moon. Whereas all previous successful robotic landings were performed with little detailed information about the lunar surface, site selection for this mission would mitigate against potential surface hazards by using existing high-resolution images of the surface, slope maps from high precision laser altimetry and stereo images, and rock abundances from orbital radar and radiometer data.

5. SPACECRAFT DESIGN

Lander. Given the stringent mass constraints on the lander, a design relying on the heritage of previous lunar lander studies (Astrium Satellites “MoonTwins”) is chosen, as shown in Figure 8. The power system relies on solar power provided by fixed solar panels and enough batteries to enable a minimum set of instruments to operate during the night at reduced levels. An automated wake-up system is implemented in the power system to restore avionics to nominal operations at sunrise. The avionics architecture is based on a dual structure: the command and data management unit (CDMU) contains the processor module and memory, and the electrical interface unit (EIU) manages the interface and power lines with the attitude and orbit control system (AOCS), the propulsion units, and the heater lines. The architecture is based on an RS422 interface bus, and has the ability to be switched on and off by a timer to allow the minimization of energy spent during the night.

The power subsystem is based on a classical design (PCDU, unregulated 28 V). The available power provided by the solar arrays is about 200 W during the day, and the available power during the night is about 4 W. Night operations for the payload will be mostly in a low power mode and with a duty cycle that is compatible with the battery resources. A low energy consumption mode is used for payloads that remain on during the night, with all possible bus equipment switched off. Small periodic wake-ups (with a negligible impact on energy budget) are used to store data in the CDMU and perform basic communications, such as time synchronization of the landers with the orbiter. Data will be stored at instrument level during the night and uploaded during the day.

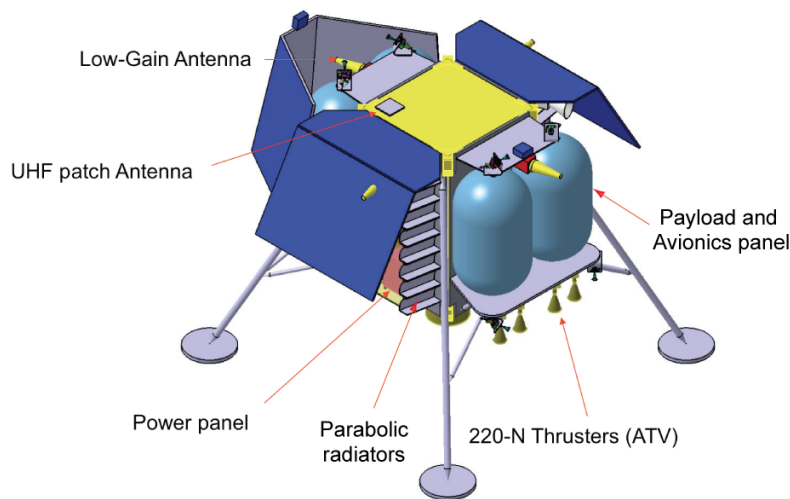


Figure 8. Lander preliminary accommodation. After landing, an arm (not shown) will deploy the seismometer and heat flow probe on the ground (Image courtesy of Astrium Satellites).

Surviving the lunar night, while keeping enough resources for minimum payload operation and keeping the avionics warm to operate during 14 terrestrial days, is one of the biggest technical challenges of this mission. Two concepts have been considered that have a considerable impact on the lander spacecraft sizing. One concept makes use of radioisotope heating units (RHUs) to keep the critical parts of the spacecraft (batteries and sensor electronics) warm and uses a second surface mirror (SSM) shadowed from the Sun to dissipate the excess heat during the day. An alternate design that does not make use of RHUs has more important consequences on spacecraft design. Keeping the critical avionics warm during the night requires a very good insulation of the warm electronics box, and thermal switches and parabolic shaped radiators to allow sufficient heat rejection during the day.

The nominal operational life of the lander is four years. Time synchronization of the landers, which is required by the science payload, will be performed through the lander-satellite link. The initial localization of the landers on the surface will rely on differential Doppler measurements and lander star-tracker attitude and inertial measurements. Comparing images from the surface with existing high-resolution Lunar Reconnaissance Orbiter images and topography will enable a more precise localization. The high precision required by the radio astronomy interferometric measurements will be eventually reached using known radio-sources.

Relay Satellite. The requirements of the relay satellite are to accommodate a small scientific payload (an impact flash camera and magnetometer) while providing a telecommunications relay for the two landers. A tailored Myriad-class (CNES mini-satellite) or mini-geo bus design is considered to accommodate the scientific payload, as shown in Figure 9. Even if this satellite class is designed for a shorter life duration (1 year), the Demeter and Parasol missions have already proven reliability compatible with our requirements. The control of the satellite attitude allows satellite orientation on 3 axes. In the nominal mission, the relay satellite is oriented along a North-South line, with the relay antenna on its “East” side. The satellite remains pointed towards the night side of the Moon for impact monitoring.

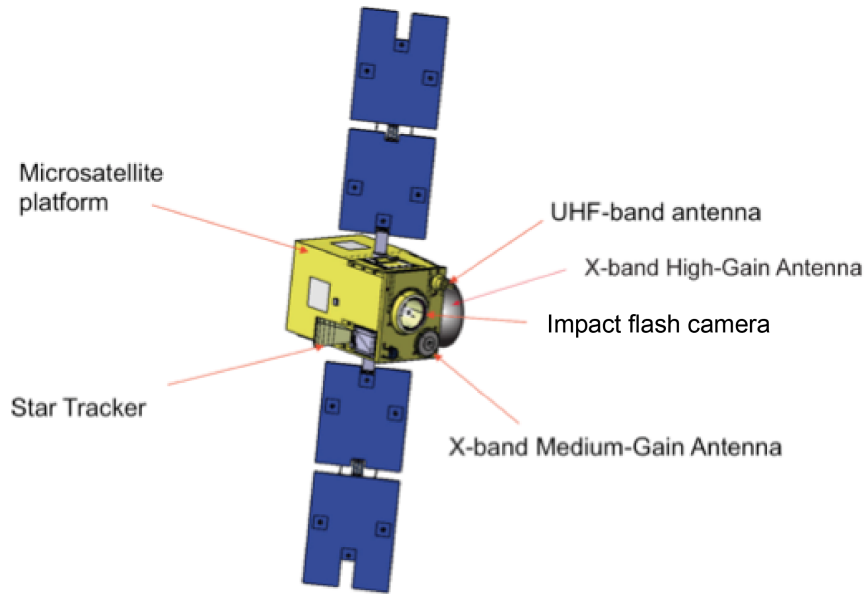


Figure 9. The relay satellite is based on a micro-satellite platform. The satellite includes the impact monitoring camera on the front side, and a deployable high gain antenna on the side. The magnetometer (not shown) is to be deployed on a boom.

To allow minimal changes to on-board software with respect to existing designs, the satellite follows a tabulated attitude as a function of the orbit (pseudo-)period, and the solar panels rotate with the Earth-Moon synodic period. To enable the telecom relay capability, two options are considered: either an X-band beam forming network (BFN) would keep a synthetic tracking of Earth, or Earth pointing would occur when impact flashes could not be observed (such as when the farside disk is fully illuminated by the Sun). Satellite pointing precision is about 5×10^{-3} arc degrees, which is sufficient for the impact flash detection payload. In the nominal mode, the satellite uses a stellar sensor and four reaction wheels, with Sun sensors being provided for safe mode. Momentum wheel desaturation is performed with cold gas thrusters. The electrical power subsystem (EPS) derives from the Myriad power generator, but uses 4-hinged panels (folded during launch) instead of 2. The overall surface area of the solar panels is approximately 1.6 m^2 , which allows 360 W of power. A UHF local link is used for communication between the probes and relay satellite (following the Proximity-1 Space Link Protocol) with an uplink capability of about 128 to 256 kbps. An X/S band link is used for direct communication of the relay satellite to Earth.

Launch Vehicle Accommodation. The accommodation of two landers, with the relay satellite attached to the upper lander, is shown in Figure 10. There is enough space within the Ariane 5 fairing to accommodate two additional landers.

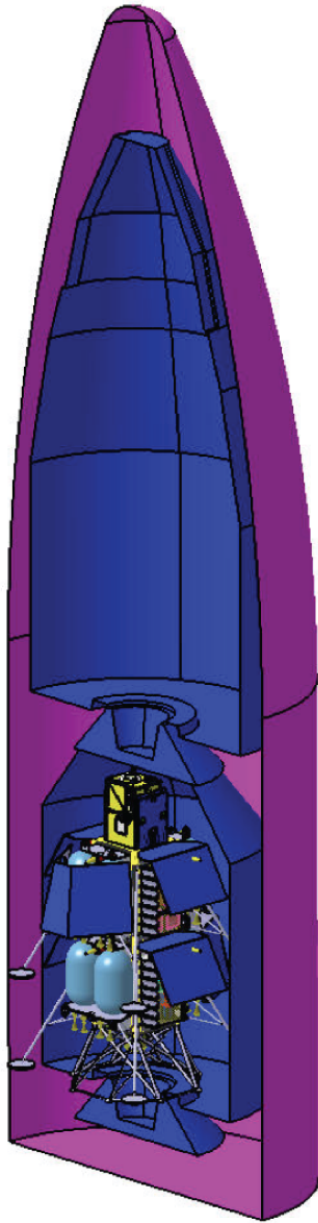


Figure 10. Ariane 5 spacecraft accommodation showing two landers and the relay satellite that is attached to the upper lander. At least two additional landers (not shown) can be accommodated within the upper portion of the fairing. Image courtesy of Astrium Satellites.

ACKNOWLEDGEMENTS

The authors would like to thank Astrium Satellites and Thales Alenia Space who contributed to the preparation of this paper. George Smoot, Sean Solomon, Claude Jaupart and Stuart Ross Taylor are thanked for their input and support of this mission proposal.

REFERENCES

- Alexander, J.K., M. L. Kaiser, J. C. Novaco, F. R. Grena, and R. R. Weber (1975), Scientific instrumentation of the Radio-Astronomy-Explorer-2 satellite, *Astron. Astrophys.*, **40**, 365-371.
- Baratoux, D., H. Chennaoui-Aoudjehane, F. Colas, Z. Benkhaldoun, A. Jambon, A. Leroy, P. Lognonné, O. Azagrouze, M. Ait Moulay Larbi, S. Bouley, A. Bounhir, M. Calvet, S. Chaabout, S. Chevrel, A. Daassou, R. Garcia, A. Habib, A. Jabiri, N. Larouci, P. Pinet, M. Sabil, and J. Vaubaillon (2012), A French-Morrocon project for the studie of impact processes on the Earth and Moon, *Lunar Planet. Sci. Conf.*, abstract 1667.
- Bely P.Y., R. J. Laurance, S. Volonte, R. R. Ambrosini, A. Ardenne, C. H. Barrow, J.-L. Bougeret, J. M. Marcaide, and G. Woan (1997), Very Low Frequency Array on the Lunar Far Side. *ESA report SCI(97)2*, European Space Agency.
- Bignami, G., P. Cargill, B. Schutz, and C. Turon (2005), Cosmic Vision, *BR-247*, ESA publications division, Noordwijk.
- Bouley, S., D. Baratoux, J. Vaubaillon, A. Mocquet, M. Le Feuvre, F. Colas, Z. Benkhaldoun, A. Dassou, M. Sabil, and P. Lognonné (2012), Power and duration of impact flashes on the Moon: Implication for the cause of radiation, *Icarus*, **218**, 115-124.
- Canup, R. (2012), Forming a Moon with an Earth-like composition via a giant impact, *Science*, **338**, 1052-1055.
- Cuk, M., and S. T. Stewart (2012), Making the Moon from a fast-spinning Earth: A giant impact followed by resonant despinning, *Science*, **338**, 1047-1052.
- Bruck, Yu., M., Ustimenko, and B. Yu (1976), Decametric radio emission from four pulsars, *Nature*, **260**, 766-767.
- Committee on the planetary science decadal survey (2010), Vision and Voyages for Planetary Science in the Decade 2013-2022, National Research Council, National Academies Press, Washington, DC.
- Dulk, G.A., W. C. Erickson, R. Manning, and J.-L. Bougeret (2010), Calibration of low-frequency radio telescopes using the galactic background radiation, *Astron. Astrophys.*, **365**, 294-300.
- Garcia, R. F., J. Gagnepain-Beyneix, S. Chevrot, and P. Lognonné (2011), Very preliminary reference Moon model, *Phys. Earth Planet. Inter.*, **188**, 96-113.
- Grove, T. L., and M. J. Krawczynski (2009), Lunar mare volcanism: Where did the magmas come from? *The Elements*, **5**, 29-43.
- Hood, L. L., F. Herbert, and C. P. Sonett (1982), The deep lunar electrical conductivity profile: Structural and thermal inferences, *J. Geophys. Res.*, **87**, 5311-5326.
- Hood, L. L., and N. A. Artemieva (2008), Antipodal effects of lunar basin-forming impacts: Initial 3D simulations and comparisons with observations, *Icarus*, **193**, 485-502.
- Jester, S., and H. Falcke (2009), Science with a lunar low-frequency array: From the dark ages of the Universe to nearby exoplanets, *New Astron. Rev.*, **53**, 1-26.
- Jolliff, B. L., J. J. Gillis, L. A. Haskin, R. L. Korotev, and M. W. Wieczorek (2000), Major lunar crustal terranes: Surface expressions and crust-mantle origins, *J. Geophys. Res.*, **105**, 4197-4216.
- Jolliff, B., M. Wieczorek, C. Shearer, and C. Neal (eds.), *New views of the Moon. Rev. Mineral. Geochem.*, **60**, 721pp. (2006)
- Kaiser, M. L. (1977), A low-frequency radio survey of the planets with RAE 2, *J. Geophys. Res.*, **82**, 1256-1260.
- Lawrence, D. J., S. Maurice, and W. C. Feldman (2004), Gamma-ray measurements from Lunar Prospector: Time series data reduction from the gamma-ray spectrometer, *J. Geophys. Res.*, **109**, E07S05, doi:10.1029/2003JE002206.
- Le Feuvre, M., and M. W. Wieczorek (2011), Nonuniform cratering of the Moon and a revised crater chronology of the inner solar system, *Icarus*, **214**, 1-20.
- Lucey, P. (2009), The poles of the Moon, *The Elements*, **5**, 41-46.
- Lognonné P., and C. L. Johnson (2007), Planetary seismology, *Treatise on Geophysics*, **10**, 69-122.
- Mann, G., F. Jansen, R. J. MacDowall, M. L. Kaiser, and R. G. Stone (1999), A heliospheric density model and type III radio bursts, *Astron. Astrophys.*, **348**, 614-620.
- Mimoun, D., M.A. Wieczorek, L. Alkalai, W.B. Banerdt, D. Baratoux, J.-L. Bougeret, S. Bouley, B. Cecconi, H. Falcke, J. Flohrer, R.F. Garcia, R. Grimm, M. Grott, L. Gurvits, R. Jaumann, C.L. Johnson, M. Knapmeyer, N. Kobayashi, A. Konovalenko, D. Lawrence, M. Le Feuvre, P. Lognonné, C. Neal, J. Oberst, N. Olsen, H.

- Röttgering, T. Spohn, S. Vennerstrom, G. Woan, and P. Zarka (2012), Farside Explorer: Unique science from a mission to the farside of the Moon, *Exp. Astron.*, **33**, 529-585
- Neal, C. (2012), The Moon 35 years after Apollo: What's left to learn? *Chemie der Erde, Geochem.*, **69**, 3-43.
- Norman, M. D. (2009), The lunar cataclysm: reality of "mythconception"?, *The Elements*, **5**, 23-28.
- Ortiz J. L., F. J. Aceituno, J. A. Quesada, J. Aceituno, M. Fernandez, P. Santos-Sanz, J. M. Trigo-Rodriguez, J. Llorca, F. J. Martin-Torres, P. Montanes-Rodriguez, and E. Palle (2006), Detection of sporadic impact flashes on the Moon: Implications for the luminous efficiency of hypervelocity impacts and derived terrestrial impact rates, *Icarus*, **184**, 319-326.
- Parker, J. (2006), Families of low-energy lunar halo transfers, *AAS/AIAA SpaceFlight Mechanics Conference*, AAS06-132, Tampa, Florida, January 22-26.
- Popov, M. V., A. Kuz'min, O. Ul'yanov, A. Deshpande, A. Ershov, V. Zakharenko, V. I. Kondrat'ev, S. V. Kostyuk, B. Losovskii, and V. Soglasnov (2006), Instant radio spectra of giant pulses from the Crab pulsar from decimeter to decameter wavelengths, *Astron. Reports*, **50**, 562-568.
- Pritchard, J. R., and A. Loeb (2010), Constraining the unexplored period between the dark ages and reionization with observations of the global 21 cm signal, *Phys. Rev. D*, **82**, 023006.
- Smith, D. E., M. T. Zuber, G. A. Neumann, F. G. Lemoine, E. Mazarico, M. H. Torrence, J. F. McGarry, D. D. Rowlands, J. W. Head III, T. H. Duxbury, O. Aharonson, P. G. Lucey, M. S. Robinson, O. S. Barnouin, J. F. Cavanaugh, X. Sun, P. Liiva, D.-d Mao, J. C. Smith, and A. E. Bartels (2010), Initial observations from the Lunar Orbiter Laser Altimeter (LOLA), *Geophys. Res. Lett.*, **37**, L18204, doi:10.1029/2010GL043751.
- Suavet, C., B. P. Weiss, W. S. Cassata, D. L. Schuster, J. Gattacceca, L. Chan, I. Garrick-Bethell, J. W. Head, T. L. Grove, and M. D. Fuller (2013), Persistence and origin of the lunar core dynamo, *Proc. Nat. Acad. Sci.*, doi:10.1073/pnas.1300341110.
- Suggs, R. M., W. J. Cooke, R. J. Suggs, W. R. Swift, and N. Hollon (2008), The NASA lunar impact monitoring program, *Earth Moon Planets*, **102**, 293-298.
- Takahashi, Y. D. (2003), New astronomy from the Moon: A lunar based very low frequency array, M.Sc. thesis, University of Glasgow, 90pp.
- Taylor, S. R., G. J. Taylor, and L. A. Taylor (2006), The Moon: A Taylor perspective, *Geochim. Cosmochim. Acta*, **70**, 5904-5918.
- Tsiganis, K., R. Gomes, A. Morbidelli, and H. F. Levison (2005), Origin of the orbital architecture of the giant planets of the Solar System, *Nature*, **435**, 459-461.
- Vyshlov, A. S., N. A. Savich, M. B. Vasilyev, L. N. Samoznaev, A. I. Sidorenko, and D. Y. Shtern (1976), Some results of cislunar plasma research, *NASA Special Publication*, **397**, 81-85.
- Weber, R. C., P.-Y. Lin, E. J. Garnero, Q. Williams, and P. Lognonné (2011), Seismic detection of the lunar core, *Science*, **331**, 309-312.
- Wieczorek, M. A., B. L. Jolliff, A. Khan, M. E. Pritchard, B. P. Weiss, J. G. Williams, L. L. Hood, K. Richter, C. R. Neal, C. K. Shearer, I. S. McCallum, S. Tompkins, B. R. Hawke, C. Peterson, J. J. Gillis, and B. Bussey (2006), The constitution and structure of the lunar interior, *Rev. Mineral. Geochem.*, **60**, 221-364.
- Wieczorek, M. A., G. A. Neumann, F. Nimmo, W. S. Kiefer, G. J. Taylor, H. J. Melosh, R. J. Phillips, S. C. Solomon, J. C. Andrews-Hanna, S. W. Asmar, A. S. Konopliv, F. G. Lemoine, D. E. Smith, M. M. Watkins, J. G. Williams, and M. T. Zuber (2013), The crust of the Moon as seen by GRAIL, *Science*, **6120**, 671-675.
- Wieczorek, M.A., B. P. Weiss, and S. T. Stewart (2012), An impactor origin for lunar magnetic anomalies, *Science*, **335**, 1212-1215.
- Wieczorek, M.A., and R. J. Phillips (2000), The "Procellarum KREEP Terrane": Implications for mare volcanism and lunar evolution, *J. Geophys. Res.*, **105**, 20417-20430.
- Zarka, P., L. Lamy, B. Cecconi, R. Prangé, and H. O. Rucker (2007), Modulation of Saturn's radio clock by solar wind speed, *Nature*, **450**, 265-267.
- Zarka, P., J.-L. Bougeret, C. Briand, B. Cecconi, H. Falcke, J. Girard, J.-M. Griessmeier, S. Hess, M. Klein-Wolt, A. Konovalenko, L. Lamy, D. Mimoun, and A. Aminaei (2012), Planetary and Exoplanetary Low Frequency Radio Observations from the Moon, *Planet. Space Sci.*, **74**, 156-166.

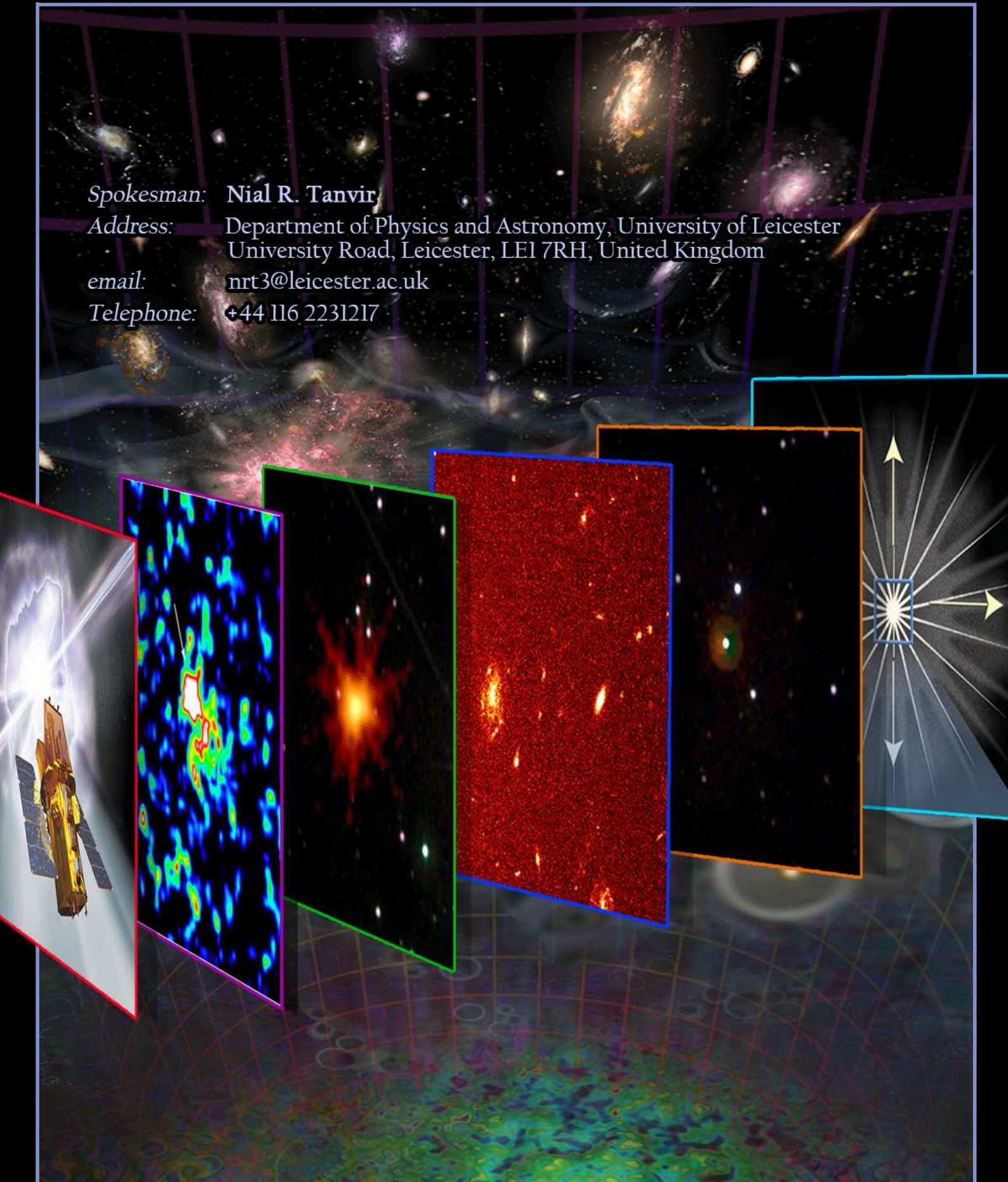
Light from the Cosmic Frontier: Gamma-Ray Bursts

Spokesman: **Nial R. Tanvir**

Address: Department of Physics and Astronomy, University of Leicester
University Road, Leicester, LE1 7RH, United Kingdom

email: nrt3@leicester.ac.uk

Telephone: +44 116 2231217



Authors:

L. Amati (INAF/IASF Bologna), J.-L. Atteia (IRAP/CNRS/UPS Toulouse),
L. Balazs (Konkoly Obs.), S. Basa (LAM Marseille), J. Becker Tjus (Univ. Bochum),
D.F. Bersier (Liverpool Univ.), M. Boër (CNRS-ARTEMIS Nice), S. Campana (INAF/OAB Brera),
B. Ciardi (MPA Garching), S. Covino (INAF/OAB Brera), F. Daigne (IAP Paris),
M. Feroci (INAF/IAPS Rome), A. Ferrara (SNS Pisa), F. Frontera (INFN/Univ. Ferrara),
J.P.U. Fynbo (DARK Copenhagen), G. Ghirlanda (INAF/OAB Brera),
G. Ghisellini (INAF/OAB Brera), S. Glover (Uni. Heidelberg), J. Greiner (MPE Garching),
D. Götz (CEA Saclay), L. Hanlon (UCD Dublin), J. Hjorth (DARK Copenhagen),
R. Hudec (AI AS CR Ondřejov & CTU Prague), U. Katz (Univ. Erlangen),
S. Khochfar (Univ. Edinburgh), R. Klessen (Univ. Heidelberg), M. Kowalski (Univ. Bonn),
A.J. Levan (Univ. Warwick), S. McBreen (UCD Dublin), A. Mesinger (SNS Pisa),
R. Mochkovitch (IAP Paris), P. O'Brien (Univ. Leicester), J.P. Osborne (Univ. Leicester),
P. Petitjean (IAP Paris), O. Reimer (Univ. Innsbruck), E. Resconi (TU München),
S. Rosswog (Univ. Stockholm), F. Ryde (KTH Stockholm), R. Salvaterra (INAF/IASF Milano),
S. Savaglio (MPE Garching), R. Schneider (INAF/Oss. Rome), G. Tagliaferri (INAF/OAB Brera),
A. van der Horst (Univ. Amsterdam)

Supporters:

M. Ackermann (DESY Zeuthen), Z. Bagoly (Eötvös Univ.), E. Bernardini (DESY Zeuthen),
J.H. Black (Chalmers Univ. of Techn.), P. Clark (Uni. Heidelberg), B. Cordier (CEA Saclay),
J.-G. Cuby (LAM Marseille), F. Ferrini (Univ. Pisa), C. Finley (Stockholm Univ.),
S. Klose (Tautenburg Obs.), A. Klotz (IRAP/CNRS/UPS Toulouse),
T. Krühler (DARK Copenhagen), N. Langer (Univ. Bonn), K. Mannheim (Würzburg Univ.),
E. Nakar (Wise Obs.), C.-N. Man (ARTEMIS, CNRS/OCA Nice),
M. Pohl (DESY Zeuthen, Potsdam Univ.), P. Schady (MPE Garching), S. Schanne (CEA Saclay),
V. Springel (ITS Heidelberg), P. Sutton (Cardiff Univ.), N. van Eijndhoven (Brussel Univ.),
J.-Y. Vinet (ARTEMIS, CNRS/OCA Nice), A. Vlasov (CPA Leuven),
D. Watson (DARK Copenhagen), K. Wiersema (Univ. Leicester)

Supporters from non-ESA member states:

V. Bromm (Univ. Texas Austin), N. Gehrels (GSFC), N. Kawai (Tokyo Inst. Techn.)

1 Executive Summary

Gamma-Ray Bursts (GRBs) are the most powerful cosmic explosions since the Big Bang, and thus act as signposts throughout the distant Universe. Over the last 2 decades, these ultra-luminous cosmological explosions have been transformed from a mere curiosity to essential tools for the study of high-redshift stars and galaxies, early structure formation and the evolution of chemical elements. In the future, GRBs will likely provide a powerful probe of the epoch of reionisation of the Universe, constrain the properties of the first generation of stars (Fig. 1), and play an important role in the revolution of multi-messenger astronomy by associating neutrinos or gravitational wave (GW) signals with GRBs.

Here, we describe the next steps needed to advance the GRB field, as well as the potential of GRBs for studying the Early Universe and their role in the upcoming multi-messenger revolution. We identify the following fundamental questions as the prime science drivers for the next two decades:

- When did the first stars form, what are their properties, and how do Pop III stars differ from later star formation in the presence of metals?
- When and how fast was the Universe enriched with metals?
- How were the first structures formed which then developed into the first galaxies?
- How did reionisation proceed as a function of environment, and was radiation from massive stars its primary driver?
- What is the relation between GRB rate and star formation rate, and what is its evolution with time? What is the true redshift distribution and corresponding luminosity function of long GRBs?
- How are γ -ray and neutrino flux in GRBs related, and how do neutrinos from long GRBs constrain the progenitor and core-collapse models?
- Can short-duration GRBs be unambiguously linked to gravitational wave signals, and what do they tell us about the neutron star merger scenario?
- What are the electromagnetic counterparts to gravitational waves and neutrino bursts?

These questions relate directly to the Cosmic Vision theme #4, “*How did the Universe originate and what is it made of?*”, in particular to 3 out of the 8 goals: (1) *Find the first gravitationally-bound structures that were assembled in the Universe – precursors to today’s galaxies and cluster of galaxies – and trace their evolution to the present.* Since GRBs can be detected from extreme distances ($z \sim 30\text{--}60$ [2]), accurate localisations provide the best-possible pointers to the first stars, and to the proto-galaxies where they form. This will find the first black holes (BH), likely to be the seeds of the supermassive BHs which dominate the X-ray luminosity of the current Universe.

GRBs in Cosmological Context

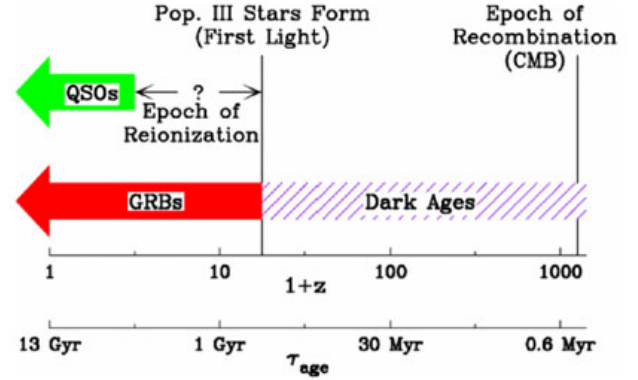


Figure 1: GRBs in the cosmological context. (From [1])

(2) ...Trace the life cycle of chemical elements through cosmic history. Since long GRBs are produced by massive stars [3], which have short life-times, the GRB rate traces the star formation (SF) rate. The extraordinary brightness and minimised absorption bias make GRBs particularly useful SF indicators. Optical/NIR afterglow spectroscopy allows to measure line-of-sight metallicity at exquisite detail, and map the cosmic chemical evolution with high- z GRBs. Short GRBs are likely linked to the formation of the heaviest elements in the Universe, such as platinum.

(3) ...Examine the accretion processes of matter falling into black holes..., and look for clues to the processes at work in gamma-ray bursts. The prompt γ /X-ray emission of GRBs, combined with polarisation measurements provides direct clues on the accretion and jet ejection processes. Optical/NIR absorption line diagnostics will link this to progenitor properties, thus allowing us to understand the basic picture of GRB-production.

To fully utilise GRBs as probes of the early Universe and/or as multi-messengers requires (i) a detection system for 5000 GRBs, among them 50 GRBs at $z > 10$, (ii) a means to localise them quickly and accurately, (iii) instrumentation to determine their redshift quickly, preferentially on the fly, and (iv) low-latency communication to the ground.

Instead of proposing a single strawman mission concept, we describe the instrumentation needed to answer each of the above fundamental questions, and describe options how to realise the measurements depending on the priorities among those questions.

The authors of this WP agree in the choice of emphasising the role GRBs can and will play in the study of the high-redshift Universe, but also recognise the huge potential of GRBs for the coincident detection of GW or neutrinos and electromagnetic signals [4]. It may confirm the basic model of the short GRBs, finally clarify the origin of the heaviest elements, and allow for a precise measurement of the expansion rate of the Universe.

2 The astrophysics landscape until 2030

The landscape of astronomy 15-20 years from now will be dominated by a wealth of new facilities across the entire electromagnetic (EM) spectrum and beyond, as non-photonic sources open a previously inaccessible window into many of the most extreme regions of the Universe. It is against this landscape that new space missions and ground-based facilities must be measured and judged. While any attempt to prescribe precisely the likely scientific frontiers at this time is fraught with uncertainty, a variety of possibilities bear consideration. We outline a range of facilities that may be operational in this time scale, along with their contribution to the science questions. A GRB-focussed mission would provide a huge step in our understanding of the early Universe, impossible by any of the facilities in the planning, and at the same time would enable some of these planned facilities to perform science that would be otherwise impossible. GRBs are the most luminous sources on the sky, releasing in less than a minute the energy output of the Sun over its entire life. Several GRBs occur each day, and thus GRBs act as frequently available signposts throughout the Universe. Two sub-groups of GRBs are distinguished according to their duration (Fig. 2): (i) Long-duration GRBs (>2 s) are firmly linked to the collapse of massive stars, thus probing sites of star formation with little delay, as the star's lifetimes are measured in megayears and not gigayears. GRBs have been seen up to the highest measured redshifts. (ii) Short-duration GRBs likely originate from the merging of compact stars and are expected to produce strong gravitational waves. Both types of GRBs are powerful neutrino sources. As stellar sized objects at cosmological scales, they connect different branches of research and thus have a broad impact on present-day astrophysics.

2.1 High-Energy satellites

Over the next few years, progress on GRBs is likely to remain driven by the *Swift* mission. Its launch heralded an unprecedented period of progress towards GRB progenitors, as well as highlighting the varied and diverse high-energy sky in ways that were unanticipated prior to its launch. *Swift* achieved this due to a combination of a broad complement of instruments dedicated to GRB detection and follow-up, and the implementation of a novel autonomous rapidly-slewing spacecraft. It has found the first GRBs at $z > 6, 7, 8$ and 9, pinpointed the locations of short-GRB afterglows, identified nearby GRBs with and (importantly) without supernovae. Despite its 8 yr in orbit, it continues to discover new populations of high-energy transients in previously unexplored parameter space. *Swift* was joined in 2008 by the *Fermi* Gamma-ray Telescope – a powerful satellite with an unparalleled spectral range, opening new insights into

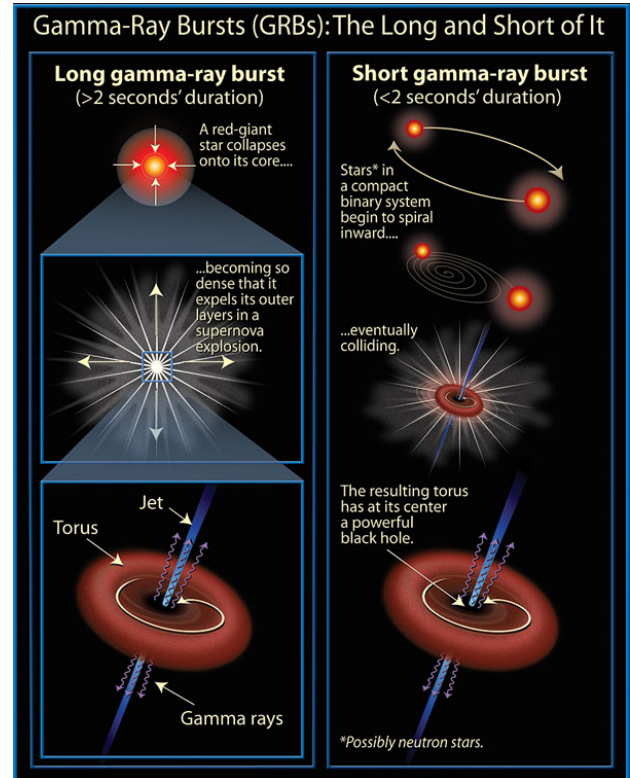


Figure 2: Long-duration GRBs (left) are thought to originate in the collapse of massive stars, while short-duration GRBs are likely produced in the merger of two compact objects. Both scenarios result in a relativistic jet which is responsible for producing the γ -ray emission. (From [5])

the nature of the γ -ray emission from GRBs, and enabling sensitive tests of differing models for quantum gravity. Real-time GRB detections are also provided by *INTEGRAL*, *AGILE*, *Suzaku*, *MAXI/ISS* and the interplanetary network (IPN) satellites.

These missions are all working well at present, but have finite lifetimes, governed both by orbital decay, instrument lifetime and, perhaps more importantly, financial constraints. It is unlikely that any of them will still be in operation well into the 2020s. Our window onto the transient high-energy sky thus revolves around new initiatives. Those likely in the interim period until 2028 are specialised instruments, often with lower sensitivity than *Swift* which will focus on individual science questions.

Four larger scale missions are the approved Indian *Astrosat*, the Japanese *Astro-H*, and the German/Russian *SRG*, as well as the planned French/Chinese *SVOM*. *Astrosat* is a multi-wavelength observatory covering the UV to hard X-ray bands scheduled for launch within a year, and may be expected to detect of order a dozen GRBs per year with its Scanning Sky Monitor. *Astro-H* is scheduled for launch in 2015, and might be able to obtain high-resolution spectra of GRB afterglows if target-of-opportunity observations can be rapidly scheduled. *SRG* will perform a sensitive all-

sky survey in the 0.3–12 keV band with the *eROSITA* and ART-XC telescopes, starting in 2015. In particular, *eROSITA* [6] with its good sensitivity is expected to detect 4–8 GRB afterglows per year [7], over a mission lifetime of at least 4 years. *SVOM* in many ways is modelled on the remarkable success of *Swift*, carrying γ -ray, X-ray and optical telescopes. The softer response (triggering at lower energy) and larger, red optimised optical telescope may enhance the recovery fraction for high- z GRBs.

Other future high-energy missions remain at an earlier stage of development, although there are plans in progress to launch small to moderate size detectors either as stand-alone missions or via the ISS. *LOFT* is currently under consideration for ESA's M3 launch slot as primarily a timing experiment, providing spectral sensitivity and timing resolution much better than *RXTE*. Its wide-field monitors make it a capable GRB detector. However, it has no automated slewing capability. In a 4 year mission it will see only < 2 GRBs at $z > 8$, even under optimistic assumptions [8]. Japan is planning an upgrade of *MAXI* on the ISS within the next few years, and NASA has recently accepted *NICER* [9], an X-ray timing and spectroscopy experiment for the ISS, with a launch date in 2017 which potentially could be used for GRB afterglow observations. The Russian space agency is planning for a small GRB mission within the next year with *UFFO*-pathfinder, a rapid detection system for prompt optical emission which later might evolve into a larger *UFFO* mission [10]. Planned for launch in the next years on the Chinese Tiangong-2 is a hard X-ray polarimeter for the study of GRBs, though this relies on the localisation and spectral measurements by a different satellite.

2.2 Multi-wavelength & multi-messenger domain

Outside of the high energy arena, the next years should see the long awaited start of routine multi-messenger astronomy, for which neutrinos from SN 1987A offered the first hints. The power of this non-photonic messenger, and gravitational waves as well, is to probe into highly enshrouded environments, invisible to electromagnetic observers. Both messengers are currently the subject of major investments, still have to reach a positive detection of signals from GRBs, but are expected to remedy this situation in the next decade.

Firstly, the upgrades to the *LIGO* and *VIRGO* interferometers will reach the point at which routine astrophysical detections of gravitational waves become reality [11]. This point should be reached towards the end of this decade [12]. Further away is a next generation of GW interferometer known as the Einstein Telescope (ET) [13] with a target operational date in the mid-2020's. Since these detectors measure gravitational wave strain, the observational horizon scales linearly with the sensitivity (unlike the inverse square

law for electromagnetic detectors). ET will be capable of seeing compact binary mergers to $z \sim 3$ (compared to the $z \sim 0.1$ for the next generation detectors). It will provide detection rates of 10^4 (10^5) for binary BH (NS) mergers [14], enabling detailed population and evolution studies. Mergers also provide a precise gravitational wave luminosity distance, giving a powerful probe of cosmology that can independently measure H_0 , Ω_M , Ω_Λ , w and \dot{w} . However, positional accuracy will be poor, even for ET operating in conjunction with further upgraded *ALIGO/AVIRGO* detectors. An EM trigger-system will therefore be needed to pinpoint source locations to an accuracy that allows measuring their redshifts, since it is the comparison of the redshift to the GW-determined distance that enables cosmological studies.

Secondly, IceCube, the largest neutrino telescope built so far, has been in operation since two years. First studies now reach beyond the level of predicted neutrino fluxes, but yielded no detection of GRB neutrinos so far [15]. The recently announced first hint of an astrophysical signal seen by IceCube provides great prospects for the identification of cosmic ray sources [16]. Possible reasons could be the choice of parameters for the standard neutrino flux calculations, in particular the Lorentz factor of the source, and the relation between accelerated electrons and protons. A more detailed treatment of the microphysics leads to a reduction of the general flux at fixed parameters, but does not take into account the general assumption that GRBs are the sources of ultra-high energy cosmic rays (UHECRs) [17]. Further studies will have to show if there is any significant spatial or temporal clustering that could be connected to GRBs and will help to study GRBs as possible cosmic ray sources. The future European neutrino telescope KM3NeT, to be deployed within this decade in the Mediterranean Sea, should provide another sensitivity boost, so that expectations remain high.

The end of this decade (or the start of the next) will see the advent of the James Webb Space Telescope (JWST) and large ground-based optical telescopes (ELT's). These are observatories rather than dedicated missions, with a science remit from exo-planets to cosmology. Central to the science case for each is the study of the early Universe. These large telescopes will be used to pin-point some of the most distant galaxies yet observed as well as providing spectroscopic capability beyond the limit of HST photometry. Nonetheless, even with these next generation facilities, spectroscopic studies remain challenging. If the faint end slope of the galaxy luminosity function is genuinely very steep [18, 19] then even these facilities will not probe it far down. If the first stars form in relatively faint and low mass haloes, it is quite likely that they will not be found directly by either facility, even in their deepest fields [20]. As we will describe later, GRBs offer a route around this prob-

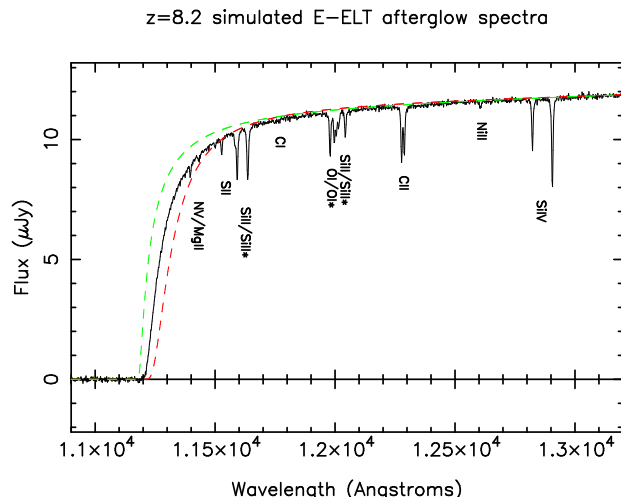


Figure 3: Simulated spectrum (solid line) around the Ly- α break showing the quality of data which would be obtained with a 30–40 m telescope such as the proposed E-ELT for an afterglow with magnitude approximately the same as that obtained for GRB 090423 (as observed by the VLT). The host galaxy was chosen to have an HI column density of 10^{21} cm $^{-2}$ and a metallicity of $1/10 Z_{\odot}$, and the IGM to be 100% neutral. The green dashed line shows a model with just a neutral IGM (with redshift fixed at that given by the host metal lines). High S/N data can be used to decompose IGM and host galaxy contributions (red dashed line), thus determining each with good precision. This simulation also shows the excellent measurements of metal abundances that could be achieved with such observations.

lem. The 30–40 m telescopes will be able to provide unique information on the chemical enrichment and re-ionization history *if* they can be fed with accurate locations of high- z GRB afterglows (see Fig. 3 for a simulated E-ELT/HARMONI spectrum).

This period will also mark the launch of ESA’s GAIA and Euclid satellites. GAIA is primarily an astrometry mission, but due to the temporal sequence of its sky scans it will detect a large number of transients, among those up to 40 GRB afterglows [21, 22]. Euclid is predominantly aimed at providing precision measurements of cosmological parameters via weak lensing and baryon acoustic oscillations. The deep survey should reach optical/NIR magnitudes of $VYJH \sim 24$ over half of the sky by 2027, and may turn up a reasonable fraction ($\sim 30\%$) of low-redshift GRB hosts.

The Large Synoptic Survey Telescope (LSST) is due to start operation around 2022. It will locate of order 10^6 transients per night. High-energy coverage of a good fraction of these transients would be of significant interest to the community providing distinction between orphan GRB afterglows, tidal disruption flares, extreme supernovae, radioactively powered transients from GW sources and other yet unimagined transients. While LSST is expected to discover 4 GRB afterglows per night [21], it will be limited to $z < 7$ due to its filter set.

There are also several significant ground-based investments across the electromagnetic spectrum, from the high frequency of the Cherenkov Telescope Array (CTA) to the low frequency radio arrays associated with the Square Kilometer Array (SKA). CTA will be sensitive to the highest energy γ -rays (> 10 TeV in some cases), and will probe high-energy emission from GRBs and their shocks in the first minutes after the bursts. The properties of the bursts at such high energy remain poorly understood at present, although the recent GRB 130427A was detected up to 120 GeV (rest-frame; [23]). Assuming the spectral-temporal extrapolations from presently detected GRBs by Fermi-LAT, CTA might detect just a few GRB/yr [24], but its orders-of-magnitude better sensitivity on short timescales compared to Fermi-LAT in the overlapping energy regime will provide a vast amount of photon data allowing to sensitively probe spectral-temporal evolution of GRBs at the upper end of the accessible electromagnetic spectrum. This will provide a handle on the prompt emission properties, and will be a powerful complement to our proposed mission – importantly, CTA will have a narrow field of view, and so will require triggers in order to re-point at GRBs. CTA should be operational towards the end of the decade. We may gain a somewhat earlier insight of the high-energy properties of GRBs via HAWC (the High Altitude Water Cherenkov Gamma-Ray observatory), which is already partially operational, and has the ability to trigger on > 1 TeV γ -ray photons across 15% of the sky, (though it lacks the sensitivity of CTA and its effective area decreases rapidly away from zenith).

Moving to longer wavelengths, the Atacama Large Millimeter Array (ALMA) will continue to be a workhorse instrument for astronomy, and its unique sensitivity to warm and cool dust in distance galaxies provides a means of probing the nature of the earliest galaxies, in particular of the highest redshift GRB hosts detected. In addition, ALMA will allow us to map systematically the GRB afterglow emission near its spectral peak, thus providing beaming-independent energy estimates. Finally, SKA, which should be operational with a partial array around 2020, will provide new insights into the formation of the first structures in the Universe and the re-ionisation through mapping the 21 cm line emission at different epochs. In the GRB field, SKA as well as its predecessors LOFAR, MWA and PAPER will be powerful facilities to study the radio afterglow emission, and will provide unique insights into the physics and environment properties of these sources. We expect that SKA will be sensitive enough to detect all afterglows of GRBs of a next generation γ -ray detector, and for 50% of those, will allow the estimate of their true (collimation corrected) energetics through late time (>100 days) radio calorimetry [25].

3 Open Questions

3.1 GRBs and the Early Universe

3.1.1 GRB observability

The identification of very high redshift ($z > 7$) sources is challenging due to their great luminosity distances, and the difficulties of observing in the near-IR (NIR) from the ground. This is exacerbated by the effects of hierarchical structure growth, which means that galaxies start out increasingly small and faint intrinsically, and that bright quasars are exceptionally rare at $z > 7$. The scientific importance of studying this era has motivated very large investments (or planned investments) in new NIR facilities (e.g. aboard JWST) that are expected to detect sources at redshifts up to $z \sim 13$ (H -band dropouts), but even they will struggle to find, much less confirm, galaxies beyond this.

To study the origin of the first stars and luminous structures in the universe, observational advances to redshifts exceeding $z \sim 13$ are essential. The exploration of this challenging high- z realm may be enabled by sources that are very bright, and have emission predominantly in the high energy regime, namely GRBs. Their specific advantages are: (i) they likely exist out to the highest redshifts due to their creation in the deaths of massive stars, (ii) the brightest bursts can easily be detected at the highest redshifts due to their huge intrinsic luminosities and energy spectra peaking at $\sim 100\text{--}300$ keV; (iii) their pan-chromatic afterglows can also be extraordinarily bright, providing backlights for detailed spectroscopy which is otherwise unprecedented at such distances; (iv) they probe the epoch we seek to understand as their progenitor stars are likely representative of those responsible for the reionisation of the Universe: their current distance record is $z \approx 9.4$ [26]; and (v) a favourable relativistic k -correction implies that they do not get fainter beyond $z \sim 3$. Yet, present and near-future ground- and space-based capabilities are limiting the measurement of redshifts at $z \sim 13$ (as H -band drop-outs), and their afterglows above $2.5 \mu\text{m}$ are too faint by many magnitudes for 8–10 m telescopes.

To fully utilise GRBs as probes of the early Universe one must localise large samples quickly and accurately, and be able to identify which of these are worth the valuable 30–40 m telescope time for detailed study, implying the determination of their (at least photometric) redshifts onboard.

3.1.2 Structure formation scenarios

From studying the cosmic microwave background we know that the Universe started out very simple. It was by and large homogeneous and isotropic, with small fluctuations that can be described by linear perturbation analysis. The present Universe, on the contrary, is highly structured and complicated. Cosmic evolution is thus a progression from simplicity to complexity,

with the formation of the first stars and protogalaxies marking a primary milestone in this transition. Compiling and characterising a sample of very high redshift GRBs will help us directly probe this key phase of cosmic structure formation, as follows.

The first stars that give birth to high- z GRBs must form out of gas that collected inside dense dark matter (DM) potential wells. Structure formation in a cold dark matter (CDM)-dominated Universe is "bottom up," with low-mass halos collapsing first. In the current concordance cosmology, with densities in CDM and dark energy of $(\Omega_M, \Omega_\Lambda) \approx (0.3, 0.7)$ as emerged from WMAP and Planck, DM halos with masses of $10^5\text{--}10^6 M_\odot$ [27, 28] form from $\sim 3 \sigma$ peaks of the initial primordial density field as early as $z \sim 25$. It is natural to identify these condensations as the sites where the first astrophysical objects, including the first massive stars, were born. Thus, one expects to find GRBs out to this limiting redshift but not beyond. While the standard CDM model has been remarkably successful in explaining the large-scale structures in the Universe and the cosmic microwave background, some discrepancies remain at small scales, $\lesssim 1$ Mpc. Proposed alternatives are either baryonic feedback or Warm DM (WDM; \sim keV particles) models [29]. In the latter case, the resulting effective pressure and free-streaming would decrease structures on small scales [30]. If indeed DM was 'warm', the high-redshift Universe would be rather empty, such that even a single GRB at $z > 10$ would already provide strong constraints on the WDM models [31]. Present constraints rule out WDM particles with masses smaller than 1.6–1.8 keV at 95% confidence level, but depend on assumptions on the slope of the luminosity function and the GRB to SFR rate ratio. Any improvements on these constraints requires a substantially larger number of GRBs with measured redshifts at $z \gtrsim 5$ [32].

On a similar note, GRBs might be used to get independent constraints on the amount of primordial non-Gaussianity in the density field [33]. Deviations from the Gaussian case can only be found at high z .

Measurements of a statistically significant sample of GRBs (minimum ~ 50) at $z > 10$ will therefore help to answer the question:

How were the first structures formed which then developed into the first galaxies?

3.1.3 When and how did the first stars form?

The nature of the first stars in the Universe, and understanding how their radiative, chemical and mechanical feedback drove subsequent galaxy evolution, provide one of the grand challenges of modern cosmology [34]. The earliest generations of stars ended the so-called cosmic dark ages and played a key role in the metal enrichment and reionisation of the Universe, thereby shaping the galaxies we see today

[34, 35, 36, 37]. These so-called Population III (or Pop III) stars build up from truly metal-free primordial gas at extremely high redshift. They have long been thought to live short, solitary lives, with only one extremely massive star with about 100 solar masses or more forming in each DM halo [41, 42, 43, 44]. However, the most recent calculations [45, 46, 47] suggest that Pop III stars formed as members of multiple stellar systems with separations as small as the distance between the Earth and the Sun [48, 49]. Although these recent fragmentation calculations suggest an initial mass function (IMF) that reaches down to sub-solar values, most of the material is probably converted into intermediate mass stars with several tens of solar masses [50, 51]. This agrees with the analysis of abundance patterns of extremely metal-poor stars in the Galactic halo [52], which requires a minimum level of enrichment to form low-mass and long-lived stars [53] and is consistent with enrichment from core collapse supernovae of stars in the intermediate mass range $20 - 40 M_{\odot}$ rather than from pair-instability supernovae of very massive progenitors with $\sim 200 M_{\odot}$ [54, 55, 56, 57, 58].

Second generation stars, sometimes termed Pop II.5 stars, have formed from material that has been enriched from the debris of the first stars. Unlike the very first stars, for which we have no direct detections yet, low-mass members of the second generation may have already been found in surveys looking for extremely metal-poor stars in our Milky Way and neighbouring satellite galaxies. The relative fraction of high-mass stars amongst Pop II.5 stars is still unknown. It is a key question in early galaxy formation to understand the transition from truly primordial star formation to the mode of stellar birth we observe today [59, 60]. When and where did this transition occur? Was it smooth and gradual or rather sudden and rapid? It is therefore important to learn more about the IMF of the first and second generation of stars and to find observational constraints on the star formation process at different redshifts. This would culminate in the more general question:

When did the first stars form, what are their properties, and how do Pop III stars differ from later star formation in the presence of metals?

3.1.4 Detecting high- z objects

Direct detections of Pop III or Pop II.5 stars in the early Universe appear highly unlikely even with upcoming observatories such as the James Webb Space Telescope (JWST) or the proposed 30–40 m ground-based telescopes (such as the E-ELT). Individual stars are much too faint, and only rich clusters of very massive stars might be bright enough to lie above the detection limits in long exposures (e.g. [61]). High-redshift observations seem only able to provide indirect constraints on the physical properties (mass, lu-

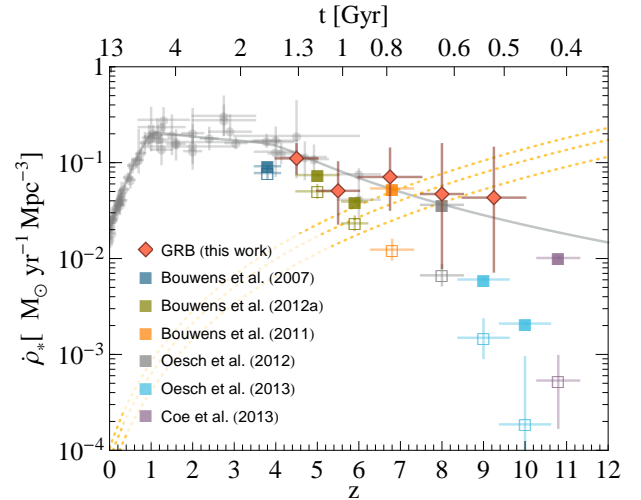


Figure 4: Star formation rate density (SFRD) Low- z data (circles) are from [38]. The diamonds are obtained using Swift GRBs. The open squares show the result of integrating the LBG UV luminosity functions down to the lowest measured value, M_{vis} , while the solid squares use $M_{\text{cut}} = -10$. All assume a Salpeter IMF. For comparison, the critical $\dot{\rho}_*$ for $\mathcal{C}/f_{\text{esc}} = 40, 30, 20$ ([39], dotted lines), top to bottom) are shown. (From [40])

minosity, frequency, etc.) of the first and second generations of stars, for instance, by looking at their influence on reionisation or on the cosmic metal enrichment history [37].

The polarisation data of WMAP, the Wilkinson Microwave Anisotropy Probe (and likely soon the Planck mission) indicate a high electron scattering optical depth, hinting that the first stars formed at high redshift [2, 62, 63]. Massive, low-metallicity Pop III stars may produce very powerful long GRBs [64, 65, 66]. Thus, GRBs offer a powerful alternative route (Fig. 4) to identifying high- z objects, as demonstrated by GRBs 080913 at $z = 6.7$ [67], 090423 at $z = 8.2$ [68, 69] and 090429B at $z = 9.4$ [26]. Indeed, studying GRBs is the only realistic pathway towards the direct detection of Pop III and high-mass Pop II.5 stars and thus towards constraining their mass spectrum as well as their multiplicity. From the predicted mass range of Pop III stars and their high binary frequency it was concluded that a $<0.6\%$ – 2% fraction of Pop III stars ended their lives in GRBs. While at the *Swift* sensitivity level only $\sim 10\%$ of GRBs detected at $z > 6$ could be powered by Pop III stars, this fraction increases to 40% at $z > 10$ [63]. In addition, both main production channels of GRBs, core collapse supernovae of massive stars (long GRBs) as well as binary mergers involving Roche-lobe overflow and common-envelope evolution (short GRBs) [70], are likely to be present. This makes high- z GRB observations the ideal probe of studying early star formation (Fig. 5). The rate of GRBs is expected to track the global cosmic star-formation rate [71, 72, 73] (Figs. 4, 5), though possibly with different efficiencies at high- z

and low- z [74, 75]. Deduced from a principal component analysis on *Swift* GRB data, the level of star formation activity at $z=9.4$ could have been already as high as the present-day one [76], a factor 3–5 times higher than deduced from high- z galaxy searches through drop-out techniques. If true, this might alleviate the longstanding problem of a photon-starving reionisation; it might also indicate that galaxies accounting for most of the star formation activity at high redshift go undetected by even the deepest searches. Clearly, observing more GRBs would be crucial to shrink the currently large error bars at the highest redshifts, thus answering the question:

What is the relation between GRB rate and star formation rate, and what is its evolution with time?

Already with current technology we can characterise GRBs up to redshifts of $z \sim 10$ [26, 68, 69], but reaching larger redshifts requires a new approach and a dedicated mission. The present *Swift* samples of GRBs, both large biased samples as well as smaller but nearly complete samples, indicate a fraction of $5.5 \pm 2.8\%$ GRBs at $z > 5$ [77, 78]. Using standard cosmology and star formation history description (Fig. 4), this translates into a fraction of 1% of all GRBs located at $z > 10$, or 0.1% of all GRBs at $z > 20$ [73]. With 1000 GRBs per year, and a nominal lifetime of 5 yr (goal 10 yr) we would expect 50 (goal 100) GRBs at $z > 10$, and 5 GRBs (goal 10) at $z > 20$. Thus, the measured GRB redshift count will be large enough to observationally constrain the cosmic star formation rate at very high redshifts and it will allow us to determine the earliest cosmic time when star formation became possible - thus answering the question:

What is the true redshift distribution and corresponding luminosity function of long-duration GRBs?

3.1.5 Chemical evolution in the Early Universe

Beside their direct detection, clues about the first stars can be obtained by studying the gas polluted by first supernova explosions [79]. Recent models for the formation of Pop III stars suggest that their typical masses are similar to those of present-day O stars, implying that they will die as standard core-collapse supernovae (CCSNe). However, it is also possible that some Pop III stars may have much larger masses, of the order of a few hundred solar masses. These stars would die as pair-instability supernovae (PISNe), leaving no remnants and producing large quantities of metals and dust [80]. The metal abundance ratios produced by CCSNe and PISNe are quite distinct, and hence by measuring their relative contributions to the metal enrichment of high-redshift gas, we can constrain the form of the Pop III IMF.

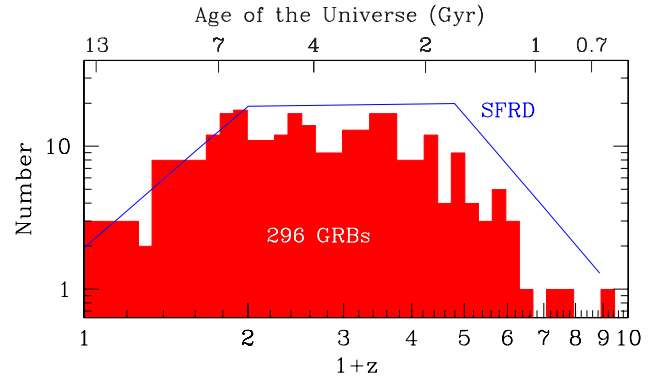


Figure 5: Histogram of the observed number of GRBs (spectroscopic redshift only) per redshift bin by May 2013 (in units of $\log(1+z)$). The number of GRBs increases from $z = 0$ to ~ 1 , is steady up to $z \sim 2.8$, then it decreases down to zero at $z \sim 9$. At low or high z , redshifts are mainly measured from the host galaxy or the DLA detected in the optical afterglow, respectively. Dust (mainly at $z = 1 - 3.5$) and γ -ray flux detection limits (for $z > 3.5$) affect our high- z detections. This is consistent with the comparison with the SFRD (co-moving volume change included) derived from field galaxies [38], scaled to match the observed $z < 1$ GRB histogram. This suggests that a substantial fraction of GRBs at high redshift is presently missed.

GRBs offer a particularly rewarding opportunity to study the physical conditions of the surrounding medium, in various ways. i) The UV radiation of the GRB and its early afterglow ionise the neutral gas and destroy most molecules and dust grains up to tens of parsecs away. Interestingly, rotational levels of molecules and metastable states of existing species (O I, Si II, Fe II) are populated by UV pumping followed by radiative cascades. As the GRB afterglow fades rapidly, recombination prevails and the populations of these levels changes on timescales of minutes to hours, imprinting variable absorption lines in the otherwise flat (synchrotron) afterglow spectrum. This allows us to measure with unprecedented accuracy the density, composition and ionisation state of the surrounding ISM [81]. ii) Other tracers of ionization are molecules forming by the impact of photons (or cosmic rays) on neutral hydrogen, via the formation of H_2^+ , which rapidly leads to the production of H_3^+ and heavier molecules [82]. GRBs, provide a good environment to induce molecule building processes via ionisation. iii) The detection of metals through optical absorption lines in the highest redshift GRBs (e.g., $z=6.3$, Fig. 10) will allow us to determine whether CCSNe or PISNe are primarily responsible for enriching the gas in these high redshift systems. This has important implications for models of the initial stages of reionization [83, 84, 85, 86] and the metal enrichment of the IGM [87, 88], thus answering the question:

When and how fast was the Universe enriched with metals?

3.1.6 The first galaxies

Identifying objects beyond $z \sim 7$ has proven extremely difficult. None of the previously claimed UDF galaxy candidates at $8.5 < z < 10$ could be confirmed by the deeper multi- λ UDF12 campaign [91] (although new candidates were identified). Even if found, such galaxies only represent the tip of the iceberg, in star-formation terms: increasing evidence suggests the bulk of early star formation happened in small, low-mass, and very faint galaxies, inaccessible to optical/NIR surveys. This is illustrated by the finding that, at $z > 5$, six GRB host fields have been observed with deep HST/VLT imaging [19, 92] with null detections in all cases. If no dust correction is applied (dust is not expected to be abundant in the Universe at an age of less than 1 Gyr, especially in small, low metallicity galaxies), the UV luminosity limit can be translated into $\text{SFR} < 2.5 \text{ M}_\odot \text{ yr}^{-1}$ [93]. Particularly remarkable is the deep $m_{AB} > 30.3$ mag NIR limit with *HST* of the host galaxy of GRB 090423 at $z = 8.23$ [19], which gives an incredibly low $\text{SFR} < 0.06 \text{ M}_\odot \text{ yr}^{-1}$.

This finding is in agreement with recent semi-analytic numerical simulations (Fig. 6) that predict that about 70% of GRB hosts at $z > 6$ will be small, with stellar mass in the range $M_\star = 10^6 - 10^8 \text{ M}_\odot$, while star formation and metallicity are in the intervals $\text{SFR} = 0.03 - 0.3 \text{ M}_\odot \text{ yr}^{-1}$ and $\log Z/Z_\odot = 0.01 - 0.1$, respectively [89]. For comparison, the deepest rest-frame luminosities achieved by the HUDF can only reveal down to $\text{SFR} \sim 0.2 \text{ M}_\odot \text{ yr}^{-1}$ at $z \sim 8$ [94].

Thus, GRBs provide a unique, and above $z \geq 13$ perhaps the only, way of pin-pointing the vast bulk of star-forming galaxies as well as their individual building blocks. Furthermore, the faintness of even the brightest galaxies at $z > 8$ makes spectroscopic confirmation very demanding. GRBs provide the opportunity of probing individual stars at these times, and their afterglows may provide not only redshifts, but detailed information about abundances, gas columns etc. via absorption line spectroscopy. Indeed, *JWST* would be able to obtain $R \sim 3000$ spectroscopy at $S/N \sim 10$ even 7 days after the GRB explosion, while the 30–40 m ground-based telescopes will be able to provide unique information on the chemical enrichment and reionisation history if they can be fed with accurate locations of high- z GRB afterglows (see Fig. 3 for a simulated E-ELT/HARMONI spectrum).

GRB lines-of-sight typically contain more gas than most QSO Damped Ly- α systems (DLAs) as they generally probe dense SF regions within their host galaxies, and in that sense are more representative of high- z star forming environments. It required observation of more than 12,000 DLA absorbers towards $\sim 10^5$ quasars to identify 5 systems with $\log N_{\text{HI}} [\text{cm}^{-2}] \geq 22$ (0.04%, [95]). In contrast, of the 31 DLAs with $\log N_{\text{HI}} [\text{cm}^{-2}] \geq 21.4$ de-

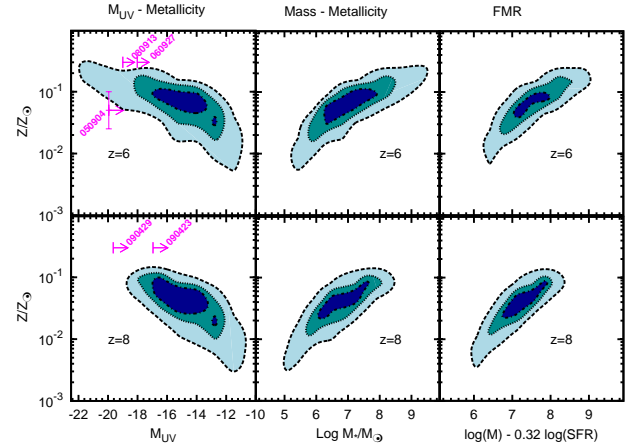


Figure 6: Luminosity-metallicity (left column), mass-metallicity (central column) and Fundamental Metallicity Relation (right column) for the LGRB host galaxy simulations at $z = 6$ and $z = 8$ [89]. Contour plots report the 30%, 60%, and 90% probability of hosting a GRB. Arrows refer to [19] and, in the absence of a measured metallicity, have been positioned arbitrarily at $Z = 0.3Z_\odot$, while the metallicity of GRB 050904 has been obtained by [90].

tected in the GRB afterglow population, 35% have $\log N_{\text{HI}} [\text{cm}^{-2}] \geq 22$ (e.g. [96, 97]).

GRBs are already allowing us to see into the heart of star-forming galaxies from $z \approx 0$ to $z > 8$ [68, 69, 98]. With afterglow spectroscopy (throughout the electromagnetic spectrum from X-rays to the sub-mm) we can characterise the properties of star-forming galaxies over cosmic history in terms of mass function, metallicity, molecular and dust content, ISM temperature, etc. Deep follow-up searches for their hosts can then place strong constraints on the galaxy luminosity function, either through weak detections (unlike LBG searches this does not require multi-band photometry for SED fitting), or non-detections which indicate the amount of star formation in undetectable galaxies.

3.1.7 Initial stages of re-ionisation

The reionization of the IGM is the subject of intensive investigation currently, and this is likely to continue for the foreseeable future. The fundamental unanswered question in the field is whether radiation from early stars was sufficient to have brought about this phase change? If not, then we will be compelled to find alternative sources of ionising radiation which, given that emission by quasars seems to fall well short of providing the necessary ionising flux at $z > 3$, may entail new physics such as decaying particle fields. Conversely, if early stars are the explanation, then reionization will teach us about their nature and the time-line of their creation. At the present time it is hard to reconcile the measured star formation with the required ionising background without invoking, e.g., a high Lyman continuum escape fraction, and/or a dom-

inant contribution from a large population of dwarf galaxies (as is, in fact, indicated by studies of high- z GRB hosts [19]), but different from [40].

Various observational windows on the process itself have begun to produce important results. Estimates of the optical depth to electron scattering of the cosmic microwave background (CMB) by WMAP and Planck indicate a reionization redshift of $z \sim 10.4 \pm 1.2$ for an instantaneous reionisation. From an analysis of 17 $z > 5$ quasar spectra it was concluded [99], that the HI fraction x_{HI} evolves smoothly from $10^{-4.4}$ at $z = 5.3$ to $10^{-4.2}$ at $z = 5.6$, with a robust upper limit $x_{\text{HI}} < 0.36$ at $z = 6.3$. However, most limits are model dependent; in fact it was shown that reionization extending to $z < 6$ is not ruled out by current data [84, 100, 101, 102].

In the near future, redshifted 21 cm mapping with LOFAR, MWA and PAPER are likely to better establish the timescale of reionisation, and ultimately much more precisely with SKA. However, the fine-scale topology of the process, and the key question of the nature of the sources responsible for the ionising radiation will remain uncertain. GRBs can provide a unique census of early massive star formation, and a route to understanding the populations of galaxies in which they formed. Crucially, in addition, high-S/N infrared spectroscopy of GRB afterglows can provide simultaneous estimates of the neutral hydrogen column density both in the host [97] and in the IGM surrounding it [103, 104] via the shape of the red damping wing of the Ly α line. While most of the flux on the blue side of Ly α is simply absorbed for a wide range of neutral fractions, the shape of the red wing depends on the neutral hydrogen fraction of the IGM in which the source is embedded, the host neutral column and the extent of any Strömgren region around the host [105]. Although this is complicated by the requirement to disentangle the HI absorption in the host from that in the IGM, in principle, it can be done as exemplified by GRB 050904 at $z = 6.3$ despite a low-S/N spectrum and high host N_{HI} (Fig. 10) [106]. A large sample of high- z GRBs will likely provide a fraction of absorbers with low column density, allowing us to cleanly isolate the IGM damping wing. The scatter in the IGM absorption from an inhomogeneous reionisation is itself a robust reionisation signature, which can be statistically isolated in a reasonably large GRB sample [107]. The exciting prospects for such studies in the era of 30–40 m ground-based telescopes is illustrated by the simulation in Fig. 3.

Thus, a sample of a few dozen GRBs at $z > 8$ would constrain not only the progress of reionisation, but its variance along different lines of sight (which may be correlated with identified galaxy populations at the same redshift), and also the typical escape fractions of radiation from early massive stars. The latter is a crucial, but extremely hard to quantify, piece of the puzzle, since only if the ionising radiation can escape

unimpeded from a significant fraction of massive stars (say, $> 20\%$), will they be successful in driving reionization. Measuring directly the neutral columns to many GRBs will establish how many lines of sight provide such an unabsorbed view. With a fiducial GRB-finder with 1000 GRBs/yr and immediate redshift estimates, ground-based spectroscopy can be secured for many dozen GRBs in the $6 < z < 13$ range. An unique and independent way of probing the high- z UV radiation field with GRBs is through its effect on high-energy photons. The expected UV field at these redshifts can cause appreciable attenuation above a few GeV, that can be observable with e.g. CTA [108]. In conclusion, a powerful GRB detection and localisation mission, in tandem with future facilities expected to be available on a 15–20 year time frame, will answer the question:

How did reionisation proceed as a function of environment, and was radiation from massive stars its primary driver?

3.1.8 Warm-hot IGM studies

The redshift distribution of X-ray absorbing column densities, N_{HI} , as detected in GRB afterglows by *Swift*/XRT shows a significant excess of high N_{HI} values at redshifts $z \geq 2$ with respect to the low redshift GRBs [109, 110, 111]. This excess absorption has been tentatively interpreted as due to the presence of absorbing matter along the line of sight not related to the GRB host galaxy. This can be either diffuse (i.e. located in diffuse structures like the filaments of the Warm-Hot Intergalactic Medium - WHIM [112, 113]) or concentrated into intervening systems (i.e. galaxies or clouds along the line of sight [109, 110]). The study of X-ray absorptions for a larger sample of GRBs at redshifts $z \geq 2$ could provide new insight on the nature of the intergalactic medium and in particular allowing to constrain its metal content.

Quasars are the alternative target for this kind of studies, in fact WHIM signatures have been detected when observing the bright blazar Mkn 421 (e.g. [114]). GRBs provide a much larger flux, if observed promptly, allowing us to extend these studies to larger distances. It is not easy to disentangle filaments from intervening systems, whereas a sufficient spectral resolution will allow us to detect distinct absorption features (originating at a given redshift) versus a truly diffuse medium (across a redshift range). As a by-product a direct measurement of the GRB redshift can be obtained from the X-ray data alone [115].

3.2 The GRB origin

3.2.1 GRBs and neutrinos

Neutrinos are electrically neutral, weakly interacting elementary particles which are produced as the result of radioactive decay, nuclear reactions or proton-proton collisions. Examples are the fusion reaction in

the Sun, electron capture during the collapse of stars into a supernova, and particle acceleration (jets) in e.g., active galactic nuclei, microquasars, supernova remnants or GRBs [116]. Due to the very small interaction cross section, neutrinos are difficult to detect, but the detection sensitivity increases dramatically with neutrino energy. This disadvantage is however an advantage for the search for neutrinos at the same time: they are neither absorbed nor deflected on their way to Earth, so that the production region can be studied. This makes them unique in the search for the origin of ultra-high energy cosmic rays.

The vast majority of the neutrinos from a GRB is emitted at moderate energies (~ 20 MeV) from the central engine's accretion disk. Their moderate energies together with the steep energy dependence of the interaction cross sections make them hard to detect. Chances are much better for those neutrinos produced by the ultra-relativistic outflow that is responsible for the GRB prompt emission. The GRB fireball phenomenology predicts spatially and temporally correlated neutrino emission to occur from proton-proton or proton-photon interaction. For a neutrino flux distributed as a power law $\propto E^{-2}$, this implies that energies in the range TeV to PeV are most promising for neutrino detection from distant sources [117].

A number of possible neutrino production sites from long GRBs have been identified: within the exploding star, within the relativistic outflow, and within the reverse shock that is formed as the afterglow is developing. Neutrinos can be formed in proton-proton and proton-photon interactions in the jet cavity that is formed as the jet penetrates the collapsing star. This is expected to produce a flash of neutrinos with energies of 3–10 TeV. Alternatively, neutrinos can be produced in the same region as the γ -ray photons, within the jet. Here, the so-called prompt neutrino emission with energies of ~ 100 TeV should accompany the γ -rays. The detailed timing of neutrino and γ -ray emission can constrain the physics of the GRB emission.

Despite sophisticated searches, neutrinos from GRBs have not been detected so far. While our best hope for neutrino detection is with the continued operation of IceCube until (at least) 2020, the follow-up project KM3NeT is in its extended design phase, with the implementation of the first phase of the infrastructure being immanent. The neutrino detection from GRBs would clarify the hadronic content in GRB jets. Moreover, systematic measurements of the neutrino energies, in particular if they peak at certain key energies, could help discriminate between models, even more so when combined with properties of the measured γ -ray spectrum. A neutrino detection from GRBs would also directly prove GRBs as sources of ultra-high energy cosmic rays. The ratio of neutrinos to γ -rays, typically produced in similar numbers, would provide indications, otherwise difficult to obtain, on the attenuation of γ -rays in the early stages of the fireball.

Detection of neutrinos from cosmologically remote GRBs (i) provides limits on the lifetime of the dominant mass eigenstate by a factor >200 better than for SN 1987A; (ii) is a testbed of neutrino properties with an unprecedented accuracy; (iii) tests if neutrinos follow the weak equivalence principle; (iv) facilitates the exploration of quantum-gravity-induced Lorentz invariance violation; (v) provides tremendous advantage over other methods of studying cosmology, as neutrino flavor ratios should be independent of any evolutionary effects.

In addition to those microphysics-related goals, the detection of high-energy neutrinos from GRBs aims at answering the astrophysical questions of the (i) identification of the sources of ultra-high energy cosmic rays; (ii) determination of the ratio of accelerated electrons to protons in GRBs, (iii) proper treatment of the GRB jet physics, including hadronic cosmic rays. In order to achieve those central goals, neutrino telescopes rely heavily on satellites that trigger GRBs: neutrino analyses can improve their sensitivity by reducing the main background of atmospheric neutrinos to almost zero through the selection of events in space and time, according to the occurrence of GRBs. This makes the GRB analysis one of the most sensitive ones for cosmic neutrinos. Only with existing satellite triggers, we can answer the question:

How are γ -ray and neutrino flux in GRBs related, and how do neutrinos from long GRBs constrain the progenitor and core-collapse models?

3.2.2 GRBs and gravitational waves

Short GRBs (sGRBs) and GWs are linked by the common topic “compact binary mergers” and they nicely illustrate how complementary and mutually beneficial the information obtained in both channels is [118, 119]. Moreover, the additional EM signals expected from a compact binary merger provide a close link to cosmic nucleosynthesis.

About one quarter of the *CGRO*/BATSE and *Fermi*/GBM bursts are classified as short-duration (< 2 s), hard GRBs. As short GRBs are intrinsically less luminous in EM radiation than their long-duration cousins, the observed sample is dominated by relatively nearby sources. The presently known redshift distribution suggests that a detection rate of 1000 GRBs/yr corresponds to 10–15 short GRBs/yr at $z < 0.1$ (~ 450 Mpc) [11, 120], depending on the energy range of the trigger instrument.

The question of their central engine is a long-standing puzzle. Compact binary mergers (either NS-NS or NS-BH) are the prime suspects, but this connection is far from proven. The coincident detection of a sGRB and a GW signal could finally settle this issue. The network of the gravitational detectors Advanced *LIGO*/*VIRGO*, soon complemented by LIGO-India and KAGRA is expected to deliver the first direct GW detections within a few years from now. It

will be capable of identifying an optimally oriented NS-NS (NS-BH) merger out to ~ 450 (~ 900) Mpc, with a combined predicted rate of the order of 50 yr^{-1} [14]. The GW signal of a compact binary merger potentially delivers a wealth of information on the physical parameters of the binary system. For example, it provides the neutron star masses and radii, it carries the imprint of the equation of state at supra-nuclear densities, and it constrains the collapse stages, e.g., through a hypermassive NS or magnetar to a BH, information that is hardly accessible otherwise. Comparison of the rates of GW detections with and without sGRB counterparts may constrain the geometry of the relativistic outflow (“jet”), the source energetics and the physical emission processes. But while providing a clear view on the physics of the actually merging system the poor localizations by GWs of ≈ 10 –1000 square degrees [12, 121] leave us nearly blind with respect to the astrophysical environment in which the merger takes place.

A complementary EM detection can provide a wealth of additional information. Firstly, it locates the source for optical follow-up providing an accurate localization relative to the host galaxy, thus allowing us to study the environment of such evolved sources. This, in turn, constrains binary stellar evolution by providing information on kick velocities, initial separations etc. Secondly, a redshift and thus luminosity determination combined with the absolute source luminosity distance provided by the GW signal can deliver precise measurements of the Hubble parameter (10 GW+EM events in Advanced LIGO/Virgo may constrain the Hubble parameter to 2–3% [122], and ET will constrain it to $<1\%$ [123]), and hence help to break the degeneracies in determining other cosmological parameters via CMB, SNIa or BAO surveys. Thirdly, the detection of a radioactively powered transient [124] may provide an interesting link to cosmic nucleosynthesis: this could show the “r-process in action” and finally settle the question of where the heaviest elements around the platinum peak (nucleon numbers $A \sim 195$) come from. Neutrino-driven winds from a merger remnant [125] may lead to yet another radioactive transient, but with likely different properties. Once the matter that is dynamically ejected interacts with the ambient medium it may produce radio flares which independently would set a limit on the merger rate [126].

The localisation of GW events has another more subtle benefit: it improves the accuracy with which parameters can be estimated from the GW observation [127]. The covariance of angular errors with uncertainties in other parameters (distance, polarisation, stellar masses etc) is usually significant. Thus, a more accurate position through EM follow-up also improves the determination of all the parameters measured gravitationally. For short GRBs, several of the GW events will be near threshold, and because the

GW amplitude is peaked along the jet axis, the detection range increases by a factor of ~ 2 with coincident detection of a short GRB X-ray afterglow [128].

Only with a sensitive GRB detector in orbit, operating in conjunction with the gravitational wave detectors, can we answer the question:

Can short GRBs be unambiguously linked to gravitational wave signals, and what do they tell us about the neutron star merger scenario?

3.2.3 Gamma-ray polarisation

Until 5 years ago, the prompt 20–1000 keV emission was interpreted as a smoothly broken power law produced by synchrotron emission. Recent discoveries of additional spectral components at high and low energies with *Fermi*, as well as γ -ray polarisation measurements with *INTEGRAL* and *IKAROS* have dramatically challenged our view of the GRB emission process. Is the broken power law a Comptonised thermal component from the photosphere? Is the high-energy part produced by inverse Compton radiation and the low-energy component of synchrotron origin? Time-resolved γ -ray polarimetry of the GRB prompt emission would be a unique discriminant of the underlying physics. The level of polarisation will depend on the radiation mechanism as well as geometrical effects. In particular, it will probe the strength and scale of the magnetic field. A significant level of polarisation can be produced by either synchrotron emission or by inverse Compton scattering. The fractional polarisation from synchrotron emission in a perfectly aligned magnetic field can be as high as 70–75% [129, 130]. An ordered magnetic field of this type would not be produced in shocks but could be advected from the central engine [129, 131]. Strong correlations are predicted between the polarisation level, the jet Lorentz factor and the power-law index of the particle distribution [132]. Another asymmetry capable of producing polarisation, comparable to an ordered magnetic field, involves a jet with a small opening angle that is viewed slightly off-axis [133]. In the case of photospheric emission, as recently hotly debated based on *Fermi* data, polarisation can arise due to the multiple Compton scatterings before photons escape [134]. Measurements of the temporal evolution of both, the degree of polarisation as well as the polarisation angle have strong diagnostic power to constrain GRB models.

Recently some measurements of polarisation during the prompt emission of GRBs in the hundreds of keV energy range have been reported [135, 136, 137, 138, 139, 140, 141]. Although all these measures, taken individually, have not a very high significance ($\gtrsim 3 \sigma$), they indicate that GRBs may indeed be emitters of polarised radiation. In particular, the changing polarisation angle with time [138, 140] indicate a fragmented jet. This kind of polarisation measurements can shed

new light on the strength and scale of magnetic fields, as well as on the radiative mechanisms at work during the GRB prompt emission phase.

In addition, polarisation measures in cosmological sources are also a powerful tool to constrain Lorentz Invariance Violation (LIV), arising from the phenomenon of vacuum birefringence as shown recently [141, 142, 143, 144].

The next generation of instruments will be sensitive enough to not only provide averaged polarisation angles and degrees for each detected event (long and short bursts), but even more pulse-resolved measurements for the brighter events. The detailed analysis of the prompt emission polarisation properties in GRBs would lead to essential clues to the emission mechanism. In particular, an ordered magnetic field can be determined or ruled out.

3.3 Time-domain astrophysics

It is now widely accepted that the next astronomical discovery frontier is the time domain (as emphasised in the Astronet Roadmap and in the US Decadal Survey). Current time-domain experiments are extremely successful and the coming years will see a revolution in time-domain astronomy with many surveys in the optical and in the radio.

3.3.1 Other high-energy transient types

Besides GRBs, also other transient source classes can trigger instruments surveying for GRBs. Transient high-energy sources, watched in real-time, offer insight into the physics of accretion, the presence (and mass) of BH in galaxies, and the behaviour of matter under extreme gravitational and magnetic fields, to name but a few. While much science in these diverse subjects arises from detailed follow-up across the EM spectrum, many of the events are most dramatic at higher energies, and hence require high-energy triggers to identify, even in the era of LSST.

Within the Milky Way our proposed mission will be sensitive to emission from M-dwarf stars, mapping out the frequency of their activity and the implications for planet habitability (especially important as many next generation planet searches are targeting M-dwarfs due to improved contrast). We will pinpoint soft gamma-repeaters – highly magnetised neutron stars that are possible gravitational waves sources and which provide a test bed for physics in both strong gravitational and magnetic fields, and for models of the supernovae that may create them. Outbursts from X-ray binaries of various types are also likely to be discovered, potentially even from outside the Milky Way.

The breakout of the SN shock from the star might provide a short lived, but luminous X-ray burst, that has likely been observed in at least one case (SN 2008D). More generally, SN can create powerful X-rays via

their interaction with the circumstellar medium, offering a route to studying mass loss in the years before the stars demise. X-rays might also be generated from engine-like events deep within the ejecta that become visible at late times as the ejecta becomes optically thin. Of particular importance is the nature of the superluminous SN [145], whose origin may be similar to the dominant mechanism thought to operate in Pop III stars, and which have recently been claimed to be (at least occasionally) powerful X-ray emitters.

Moving further out into the Universe we can study more massive black holes in galactic nuclei. The recent discovery of hard high-energy emission from Swift J1644+57 [146] suggests that tidal disruption flares (TDFs) might be powerful high-energy transients, while it is also thought that all TDF produce softer thermal X-ray emission [147]. TDFs offer a unique route to probing BHs in galaxies, including their location and ubiquity within dwarf galaxies (where it is far from obvious they reside), hence they allow us to extend the relation between BH mass and stellar velocity-dispersion to much lower masses, providing strong constraints on galaxy evolution models. Finally, these events allow us to study accretion around supermassive BHs from switch on to switch-off in human timescales, much shorter than the millions of years in which active galaxies evolve.

3.3.2 Complementarity with other transient detection systems

The main reason for the community-wide focus on the transient and dynamic Universe is that it is most often associated with extreme physical phenomena: eruptions on a stellar surface, complete explosion of a star, "shredding" of a star by a supermassive black hole, merger of two extremely compact objects, etc. These phenomena most often emit non-EM signals, in particular cosmic rays, neutrinos and gravitational waves. Our proposed mission concept would detect and localise energetic phenomena that are most likely to be associated with non-EM signals.

By definition, the transient sky is unpredictable which is why all EM facilities have a very large field of view; the need for very wide area coverage cannot be overstated, and it is crucial to have an EM monitor that sees a large fraction of the sky all the time. We can do this with a dedicated γ -ray mission. Focussing on one wavelength range or one information carrier (e.g., EM, GW, ν) is like having a black-and-white picture: there is useful information but we are missing something. A range of instruments covering the whole EM spectrum in conjunction with other information carriers will give us a detailed *color* image, allowing us to see the whole physical picture, thus addressing the question:

What are the electromagnetic counterparts to gravitational waves and neutrino bursts?

4 Requirements for enabling instruments

Both, the use of GRBs as a tool as well as the simultaneous detection of an EM signal with a GW/neutrino signal, requires an in-orbit trigger and search facility (“GRB-finder”) that can simultaneously localise the event within the large error boxes provided by the GW (and neutrino) facilities.

In order to use GRBs as a tool, positions with arcsec accuracy are required. The localization accuracy of the GRB-Finder will not be sufficient, and thus an X-ray and/or optical/(N)IR telescope is required which is rapidly slewed to the position determined by the GRB-Finder. An X-ray telescope is preferred since the sky is too crowded at optical/(N)IR wavelengths. Finally, to tackle the early Universe questions and obtain decent statistics at $z > 10$, a next-generation GRB mission should detect of order 1000 GRBs/yr, providing 50 (5) GRBs at $z > 10(20)$ over a 5 yr mission lifetime. This high GRB rate requires a pre-selection of ‘interesting’ events, and therefore a (N)IR telescope is foreseen which will determine redshifts for the bulk of the high-redshift (e.g., $z > 7$) sources. Table 1 summarises these high-level requirements.

4.1 The GRB-finder

The localisation accuracy and timeliness are the crucial parameters when aiming at follow-up observations at longer wavelengths. We discuss in the following only concepts which provide localisations better than a few degrees within minutes after the GRB. In general, as the prompt GRB spectral slope is -1 in the 1–100 keV band, lowering the energy threshold allows for the detection of a larger number of GRBs.

Scintillation detectors: The use of simple scintillator detectors like *CGRO/BATSE* or *Fermi/GBM* has, in the past, only led to afterglow identifications for a handful of GRBs, due to their large localisation uncertainties. The systematic error for *GBM* bursts is 3.3° for $\sim 90\%$ of the cases, with a tail of 12.5° for the rest. The twelve NaI detectors on *Fermi/GBM* work in the 8–1000 keV band, and with an effective area of about 100 cm² each in the 20–50 keV band they detect ~ 270 GRBs per year [148]. Increasing the rate to our fiducial 1000 GRBs/yr can be achieved in different ways. Firstly, if flown in, e.g., a L2 orbit, the lack of the Earth occulting half the sky implies doubling the detection rate. Secondly, increasing the effective area by simply using larger crystals is straightforward. Scaling the background rate appropriately and assuming the same S/N ratio for triggering, an effective area of $10\times$ that of *GBM* would provide 1000 GRBs/yr in a low-Earth orbit (LEO).

Coded Mask Instrument: Such systems have been widely used in space for detection of GRBs (e.g. *Swift/BAT*), and work in the ~ 2 –200 keV band. Their advantages are: i) observe over a large solid angle;

ii) can use hard X-ray/ γ -ray detectors to cover quite large energy bandpass; iii) can give fairly good localisations (one to few arcminutes); iv) provide a large number of photons, allowing easier spectral analysis. Disadvantages are: i) they are non-focussing, so sky background prohibits the detection of faint sources or monitoring of fading emission from sources which trigger the instrument (i.e. this requires an additional focussing telescope which can create a data gap – as in the case of *Swift* – while the satellite slews); ii) while coded mask instruments can be used with large-area Si detectors to cover the X-ray band, they have a modest bandwidth (2–50 keV).

Simulations using the presently known $\log N$ - $\log S$ relation and luminosity function of GRBs [75] reveal the following trade between depth and area of a coded-mask similar to *Swift/BAT*: aiming at 2 (4) times the depth of *BAT* gives a similar number of high-redshift GRBs as increasing the detector area by a factor of 2.5 (5). In order to achieve ~ 1000 GRBs/yr, a system of seven *BAT*-like systems with only modestly increased ($1.4\times$) effective area would be necessary (or correspondingly enlarged versions of the advanced coded mask instruments *SVOM/ECLAIRS* or *UFO/UBAT*).

Lobster Optic Instrument: The use of a wide-field Lobster Eye (LE) Microchannel Plate (MCP) or Multi-Foil (MFO; [149, 150]) imaging instrument provides several advantages over traditional coded mask wide-field telescopes: i) one gets continuous monitoring in a single bandpass (i.e., no gaps due to slews) as the same telescope finds and then continues to monitor the transient; ii) the use of a focussing optic lowers the sky background against which sources are detected, increasing the sensitivity by about two orders of magnitude; iii) ability for good localisation ($< 1'$ down to about $10''$) particularly for higher focal lengths; iv) multiple, lightweight modules can be utilised to cover large solid angles. The principle disadvantage is the need of (modular) large area detectors (as for coded mask telescopes). LE instruments are restricted to low energies (of order 0.5–10 keV). At similar mask/optics area and FOV, a Lobster optics would detect about 3 – $4\times$ more GRBs than a coded mask system [151]. A detection rate of 1000 GRBs/yr could be reached with about 10 modules of the type proposed for the Lobster mission [152].

Compton Telescope: A Compton telescope would work at higher energies (~ 200 keV to ~ 50 MeV), and has the advantages of i) uniquely excellent gamma-ray polarisation capability, and ii) a wide energy band. The disadvantage is a localisation accuracy substantially poorer than a coded mask or Lobster optics instrument, of order 1° radius only, and a rather large mass. An existing concept of such a detector promises ~ 600 GRBs/yr [153], close to our 1000 GRBs/yr goal.

Table 1: Scientific requirements for a future GRB mission with assumed 5 yr lifetime.

Requirement	Goal	Detector ability
1. Detect 1000 GRBs/yr	obtain 50 (5) GRBs at $z > 10(20)$	large FOV, soft response
2. Rapid transmission to ground	allow timely follow-up observations	communication network
3. Rapid localization to few "	opt/NIR identification of 1000 GRBs/yr	slewing X-ray or opt/NIR telescope
4. Provide z -indication	allow selection of high- z objects	multi-filter or spectroscopic capability

4.2 The X-ray telescope for precise localisation (and spectroscopy)

The main driver for the design of the X-ray telescope (XRT) is the position uncertainty provided by the GRB-finder, such that the full GRB error circle can be covered. In addition, the sensitivity should allow all GRB afterglows to be detected. Scaling from the *Swift/XRT* detections of the faintest GRBs, and considering the goal of reaching substantially larger redshifts (and thus likely fainter afterglows), the XRT should be a factor ~ 3 more sensitive than *Swift/XRT* (Fig. 7). Such sensitivity requirement (of order 10^{-13} erg/cm²/s in 100 sec) excludes coded mask systems. In case of a Compton telescope as GRB-finder, a FOV of 3° diameter is needed. Combined with the sensitivity requirement, a single-telescope Wolter-I optics is problematic. A practical solution is to adopt the *eROSITA* scheme of 7 Wolter-I telescopes, and adjust their orientation on the sky such that they fill the required FOV. For the other two GRB-finder options a single *eROSITA* telescope would be sufficient, or alternatively the *XMM* flight spare (though larger and more massive). Simpler versions like an enlarged version of the *SVOM/MXT* or a long focal-length Lobster are possible as well, with the trade-off of less versatile auxiliary science options as compared to *Swift/XRT*.

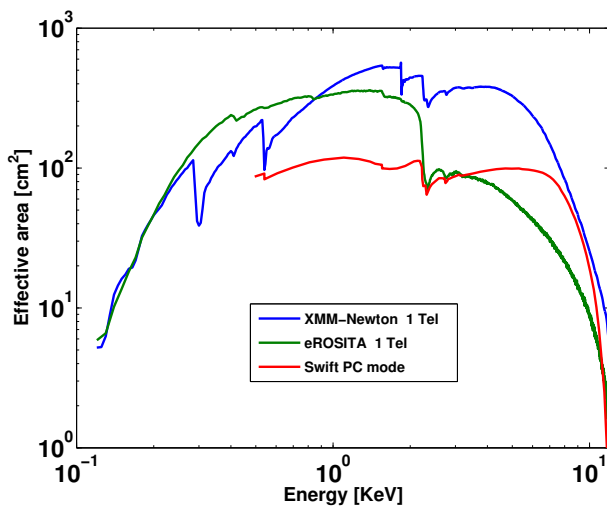


Figure 7: Comparison of the effective area of a modified *eROSITA* system (one telescope per sky position instead of all 7 telescopes co-aligned) with those of *XMM* and *Swift/XRT*. (From [153])

4.3 The (N)IR telescope

The main driver for the design of the Infrared Telescope (IRT) is the goal to (i) detect and accurately localise the counterpart, and (ii) measure the redshift to an accuracy of at least $\Delta z/z \sim 0.2$, so that high- z GRBs can be quickly identified for detailed follow-up study. Above redshift $z \sim 17$, Ly α is moving out of the K -band. This and the requirement to be sensitive up to redshifts of ~ 30 imply to observe in the L ($3.5 \mu\text{m}$) and M ($4.5 \mu\text{m}$) bands.

Based on a complete sample of GRB afterglow measurements obtained with the 7-channel optical/NIR imager GROND since 2007 (update of [77]), in particular the brightness distribution in each of the JHK channels, a minimum afterglow brightness of $M(\text{AB}) \approx 22$ mag at ~ 1 h after the GRB is deduced (Fig. 8). Such sensitivity will be reached with the future 30 m class telescopes, but since it is illusory to follow-up 3 GRBs/night with those instruments, we consider this an onboard requirement in the following.

Using standard parameters for the transmission of the optical components, read-out noise of the detector as well as zodiacal background light, a 1 m class telescope would achieve at least a 5σ M -band detection of each GRB afterglow with a 500 sec exposure when observed within 2 h of the GRB.

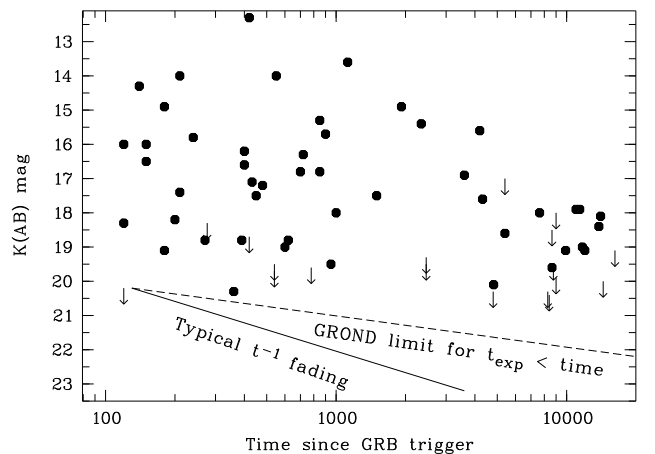


Figure 8: K -band photometry of a complete sample of GRB afterglows based on GROND data. At 1000 s after the GRB, 95% of the afterglows are brighter than $K(\text{AB})=22$ mag. With $K - M \sim 0.8$ mag for typical afterglow spectral slopes, we aim at $M(\text{AB})=21.2$ mag at 1000 sec, or $M(\text{AB})=22.2$ mag at 1 hr after the GRB.

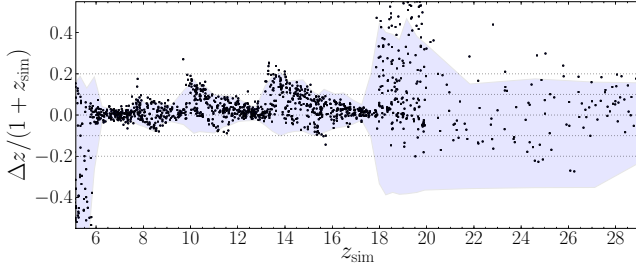


Figure 9: GRB afterglow photometric redshift accuracy of a $zYJHKLM$ filter set. Small black dots show a mock set of 900 simulated afterglow spectra and their corresponding photo- z . The blue-shaded area shows the quadratic sum of the typical difference to the input redshift and the 1σ statistical uncertainty of the photo- z analysis averaged over 30 afterglows in relative ($\eta = \Delta z/(1+z)$) terms. For the $7 < z < 17$ redshift range, the photo- z can be determined to better than 20%. At $z > 17.5$ (K -dropout), the error gets larger due to the gap above the K band and the widths of the L (M) bands; yet, the redshift accuracy is more than sufficient for any follow-up decision.

The inclusion of the LM bands into the IRT requires operating temperatures of about 37 (50) K for the M (L) channels. This will certainly require active cooling. In addition, several optical elements in the optical path will have temperature constraints, so that the thermal architecture of the instrument will need to be designed carefully, though it will be much less stringent than e.g. on Herschel.

After the slew to a GRB, the XRT will provide a position with an accuracy between 5-20'' depending on the details of the XRT and the off-axis angle of the GRB in the XRT FOV. This uncertainty is too large for immediate (low-resolution) spectroscopy, so two options are possible.

The simple and cheap option is a simultaneous multi-band imager in, e.g., the seven bands $zYJHKLM$ [153], thus covering the redshift range $7 \lesssim z \lesssim 30$. Since GRB afterglow spectra are simple power laws, and at $z > 3$ Lyman- α is the dominant spectral feature, relatively high accuracies can be reached even with broad-band filters (Fig. 9), as demonstrated in ground-based observations with GROND [154].

A more sophisticated, but also more sumptuous option is a combined imager and spectrograph, as proposed for the dedicated GRB mission ORIGIN [155]. Different areas of the detector are used for either imaging in (sequentially exposed) multi-band filters, low-resolution ($R=20$) spectroscopy, or high-resolution ($R=1000$) integral-field spectroscopy. Switching between these modes requires few arcmin re-pointings of the satellite, based on an accurate position derived from initial imaging data. The power of a $R=1000$ NIR spectrograph on a 1 m space telescope is demonstrated in Fig. 10, allowing unique absorption line diagnostics for $\sim 50\%$ of the GRBs up to the highest redshifts.

5 Strawman mission concepts

5.1 A Distributed Approach

As with other areas of research, the next step forward in understanding GRBs or using them as a tool requires a substantial larger effort on the instrumentation. The basic idea behind this distributed approach is our conviction that strategically linking together future large/expensive global facilities (both ground- and space-based) is of considerable importance to maximise the overall scientific return, in particular at the ever growing costs with more and more ambitious projects. In a perfect world, the different major funding agencies could be expected to seriously weigh up the possible synergy in making their selections.

Separating the tasks: The GRB-finder and the two narrow-field instruments do not have to share the same satellite. In fact, the rapid slewing of the X-ray and (N)IR telescopes is optimised if the flight configuration has low mass (angular momentum). A concept study with EADS Astrium indeed showed that a 2-satellite configuration flight would be preferable even in a LEO (at 500–2000 km distance, not requiring precision formation flying!) unless the GRB-finder is very simple and light-weight. Thus, a strawman concept would be (i) one satellite with a GRB-finder, and (ii) another satellite combining the *XMM* or *eROSITA* spare with an *EUCLID*-sized telescope (just M1 to M3 mirrors, and at largely reduced optical quality and alignment requirements). The GRB-finder with the largest impact on auxiliary science would be a “super-BAT”, i.e. an octahedron-shaped satellite where all but the Sun-facing direction contain a coded mask telescope with 2000 cm² detector area each. Being placed in L2, and with no slewing required, such a configuration would detect ~ 1200 GRBs/yr. The follow-up satellite would slew to each GRB and observe for ≈ 1 –2 h minimum time. This would allow up to 15 GRBs to be observed on a single day (occurring once or twice per year), but on average could leave about half the observing time of the X-ray/(N)IR telescopes to other science areas. Data could be sent to the GRB-finder (or other geostationary satellites) from where it would be much easier to rapidly down-link to Earth due to the fixed location in space.

Piggyback on ESA missions under consideration:

An alternative option could be to add a GRB-finder to one of the ESA missions already under discussion. This would provide Table 1 items (1) and (2), possibly even (3) for a subset of GRBs. Providing (3) and (4) would require either a dedicated mission or a smaller follow-up mission. We acknowledge that these are substantial modifications to the existing concepts.

(1) Adding a GRB-finder to ATHENA+ (or other L2-selected mission) and prepare for reasonably rapid (2 h) autonomous slewing capability: The presently planned ATHENA+ Wide-Field Imager has a FOV

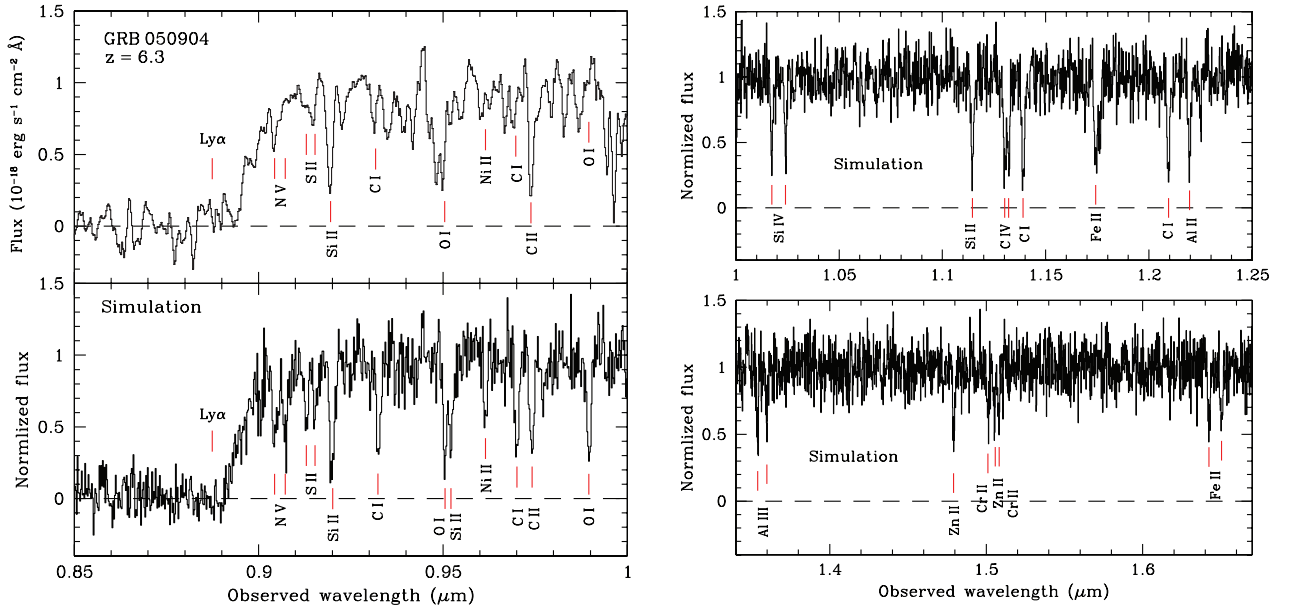


Figure 10: Spectrum of the afterglow of GRB 050904, taken with Subaru/FOCAS 3 days after the GRB (top left; [90]), and simulation of a $R=1000$ spectrum taken with a space-borne 1 m telescope for 1 hr exposure at an afterglow brightness of $J(AB)=21$ mag (lower left panel with the same wavelength range as the observed spectrum, and the J - and H -band regions in the right panels). A metallicity similar to that of GRB 050904 is assumed, with a ZnII column of $2.5 \times 10^{13} \text{ cm}^{-2}$, and ionised gas (e.g. AlIII , CIV , SiIV) with a column of $1/10$ of the neutral gas.

large enough to cover several-arcmin sized GRB error boxes, and the calorimeter would provide unique X-ray absorption spectra for the line-of-sights to GRBs. In such a scenario, a separate (N)IR telescope would be needed.

(2) Similarly, adding a GRB-finder to the Large Observatory For X-ray Timing, *LOFT* [156]: *LOFT* is expected to detect of order 150 GRBs/yr, which is too few for the purpose proposed here. Since autonomous slewing is not part of the *LOFT* concept, an enhancement of the GRB-finder capabilities would imply that a separate satellite with the X-ray and (N)IR telescope would be required.

5.2 All-in-One mission

We see the following options (though other mixing and matching of these components would also be possible), determined by the properties of the GRB-finder (field-of-view and localization accuracy). Most of these configurations, if not all, would benefit from an L2 orbit which therefore is taken as the default option. Depending on the combination, up to three autonomous slews will be needed to achieve (N)IR spectroscopy of the GRB afterglow.

Scintillator, single Lobster, NIR: Using a GBM-like detector with $4 \times$ larger effective area, and 20–24 modules will cover the full sky and return 1000 GRBs/yr with locations in the 1° – 4° range. Autonomously re-pointing a long focal length (~ 2 m), narrow-field ($8^\circ \times 8^\circ$) Lobster provides a 95% detection rate of the X-ray afterglows, and a position accu-

rate to <0.5 – $1'$. This position is good enough for the (N)IR telescope to slew and start 7-channel imaging and obtain a $1''$ position. Another slew would place the afterglow on the spectrograph.

All-sky Lobster, XRT, NIR: Using about a dozen short focal length (thus small effective area), large-FOV ($30^\circ \times 30^\circ$) Lobster modules would detect about 1000 GRBs/yr with locations accurate to few arcmin. Autonomously re-point a single *eROSITA*-like X-ray telescope to get a 99% X-ray afterglow detection rate, and localizations accurate to $\lesssim 30''$, accurate enough for (N)IR 7-channel imaging and/or grism spectroscopy. Possibly, a longer focal length Lobster could replace the *eROSITA*-like X-ray telescope.

Coded mask, XRT, NIR: Using eight Swift/BAT-like coded mask systems in octahedron orientation, and lowering the low-energy threshold from 20 to ~ 10 keV would provide 1000 GRBs/yr with locations accurate to few arcmin. Autonomously re-point a single *eROSITA*-like X-ray telescope to get a 99% X-ray afterglow detection rate, and localizations accurate to $30''$, accurate enough for (N)IR 7-channel imaging and/or grism spectroscopy.

Compton, XRT, NIR: Two systems of half a cubic-meter Compton telescopes (e.g. [153]), oriented in opposite directions, will detect about 1300 GRBs/yr, out of which about 900 will have localisations $<1^\circ$. Autonomously re-point of seven *eROSITA*-like X-ray telescopes, oriented to fill a 3° diameter FOV, will provide a 99% X-ray afterglow detection rate. The $30''$ localizations are accurate enough for (N)IR 7-channel imaging and/or grism spectroscopy.

References

- [1] Lamb D.Q., Reichart D.E., 2000, *ApJ* 536, L1
- [2] Naoz S., Bromberg O., 2007, *MN* 380, 757
- [3] Hjorth J. et al. 2003, *Nat* 423, 847
- [4] Bartos I., Brady P., Marka S. 2013, *Class. Quant. Grav.* 30, 123001
- [5] NASA, <http://ecuip.lib.uchicago.edu/multiwave-length-astronomy/gamma-ray/science/07.html>
- [6] Predehl P., Andritschke R., Böhringer H. et al. 2010, *SPIE* 7732E, 23
- [7] Khabibullin I., Sazonov S., Sunyaev R. 2012, *MN* 426, 1819
- [8] Amati L., Del Monte E., D’Elia V. et al. 2013, *Nucl. Phys. B* (in press; arXiv:1302.5276)
- [9] Gendreau K.C., Arzoumanian Z., Okajima T. 2012, *SPIE* 8443, 844313
- [10] Park I.H., Brandt S., Budtz-Jorgensen C. et al. 2013, *NJP* 15, 023031
- [11] Siellez K., Boër M., Gendre B. 2013, *MN* (subm.)
- [12] Aasi J., Abadie J., Abbott B.P. et al., 2013, arXiv:1304.0670
- [13] Sathyaprakash B. et al 2012, *Class. Quant. Grav.* 29, 124013
- [14] Abadie J., et al. 2010, *Class. Quant. Grav.* 27, 173001
- [15] Abbasi R., Abdou Y., Abu-Zayyad T et al. 2012, *Nat.* 484, 351
- [16] Whitehorn N., 2013, talk at IceCube Part. Astrophys. Symp., 2013 May 15, Madison
- [17] Hümmer S., Baerwald P., Winter W. 2012, *PRL* 108, 231101
- [18] Bouwens R.J., Illingworth G.D., Oesch P.A. et al. 2012, *ApJ* 752, L5
- [19] Tanvir N., Levan A.J., Fruchter A.S. et al. 2012, *ApJ* 754, 46
- [20] Johnson J.L., Greif T., Bromm V., 2008, *MN* 388, 26
- [21] Rossi E., Perna R., Daigne F 2008, *MN* 390, 675
- [22] de Souza R.S., Krone-Martins A., Ishida E.E.O., Ciardi B., 2012, *A&A* 545, A9
- [23] Zhu S., Racusin J., Kocevski D. et al. 2013, *GCN* #14471
- [24] Gilmore R.C., Bouvier A., Connaughton V. et al. 2013, *Exp. Astron.* 35, 413
- [25] Ghirlanda G. et al. 2013, *MN* (subm.)
- [26] Cucchiara A., et al. 2011, *ApJ* 736, C7
- [27] Tegmark M., et al. 1997, *ApJ* 474, 1
- [28] Yoshida N., Bromm V., Hernquist L., 2004, *ApJ* 605, 579
- [29] Kamada A., et al. 2013, *JCAP* 03, 008
- [30] Barkana R., Haiman Z., Ostriker J.P., 2001, *ApJ* 558, 482
- [31] Mesinger A., Perna R., Haiman Z. 2005, *ApJ* 623, 1
- [32] de Souza R.S., Mesinger A., Ferrara A. et al. 2013, *MN* (subm., arXiv:1303.5060)
- [33] Maio U., et al. 2012, *MN* 426, 2078
- [34] Bromm V., Yoshida N., Hernquist L., McKee C.F., 2009, *Nat.* 459, 49
- [35] Bromm V., Larson R.B. 2004, *ARA&A* 42, 79
- [36] Glover S. 2005, *SSRv* 117, 445
- [37] Ciardi B., Ferrara A. 2005, *SSRv* 116, 625
- [38] Hopkins A.M., Beacom J.F. 2006, *ApJ* 651, 142
- [39] Madau P., Haardt F., Rees M.J., *ApJ* 514, 648
- [40] Kistler M.D., Yüksel H., Hopkins A.M. 2013, arXiv:1305.1630
- [41] Abel T., Bryan G.L., Norman M.L. 2002, *Sci.* 295, 93
- [42] Bromm V., Coppi P.S., Larson R.B. 2002, *ApJ* 564, 23
- [43] Yoshida N., Omukai K., Hernquist L., Abel T. 2006, *ApJ* 652, 6
- [44] O’Shea B.W., Norman M.L. 2007, *ApJ* 654, 66
- [45] Clark P.C., Glover S., Simon C.O. et al. 2011a, *Sci.* 331, 1040
- [46] Greif T., Bromm V., Clark P.C. et al. 2012, *MN* 424, 399
- [47] Stacy A., Greif T., Klessen R.S. et al. 2013, *MN* 431, 1470
- [48] Turk M., Abel T., O’Shea B 2009, *Sci* 325, 601
- [49] Stacy A., Greif T., Bromm V. 2010, *MN* 403, 45
- [50] Clark P.C., Glover S., Simon C.O. et al. 2011b, *ApJ* 727, 110
- [51] Smith R.J., Glover S., Simon C.O. et al. 2011, *MN* 414, 3633
- [52] Beers T.C., Christlieb N., 2005, *ARA&A* 43, 531
- [53] Salvadori S., Schneider R., Ferrara A. 2007, *MN* 381, 647
- [54] Tumlinson J., 2007, *ApJ* 664, 63
- [55] Tominaga N., Umeda H., Nomoto K., 2007, *ApJ* 660, 516
- [56] Izutani N., Umeda H., Tominaga N. 2009, *ApJ* 692, 1517
- [57] Heger A., Woosley S.E. 2010 *ApJ*, 724, 341
- [58] Joggerst C.C. et al. 2010, *ApJ* 709, 11
- [59] Schneider R., Ferrara A., Salvaterra R. et al. 2003, *Nat.* 422, 869
- [60] Schneider R., Omukai K., Limongi M. 2012, *MN* 423, L60
- [61] Johnson J.L., Greif T.H., Bromm V. et al. 2009, *ASP Conf.* 419, p. 335
- [62] Bromm V., Loeb A., 2006, *ApJ* 642, 382
- [63] Campisi M.A., Maio U., Salvaterra R., Ciardi B., 2011, *MN* 416, 2760
- [64] Meszaros P., Rees M.J., 2010, *ApJ* 715, 967
- [65] Kommissarov S.S., Barkov M.V., 2010, *MN* 402, L25
- [66] Suwa Y., Ioka K., 2011, *ApJ* 726, 107
- [67] Greiner J., Krühler T., Fynbo J.P.U., et al. 2009, *ApJ* 693, 1610
- [68] Tanvir N.R., Fox D.B., Levan A.J. et al. 2009, *Nat.* 461, 1254
- [69] Salvaterra R., Della Valle M., Campana S. et al. 2009, *Nat.* 461, 1258
- [70] Fryer C., Woosley S.E., Hartmann D.H. 1999, *ApJ* 526, 152
- [71] Fynbo J.P.U., et al. 2008, *ApJ* 683, 321
- [72] Kistler M.D., et al. 2009, *ApJ* 705, L104
- [73] Elliott J., Greiner J., Khochfar S., et al. 2012, *A&A* 539, A113
- [74] Daigne F., Rossi E.M., Mochkovitch R. 2006, *MN* 372, 1034
- [75] Wanderman D., Piran T. 2010, *MN* 406, 1944
- [76] Ishida E.E.O., de Souza R., Ferrara A. 2011, *MN* 418, 500
- [77] Greiner J., et al. 2011, *A&A* 526, A30

- [78] Salvaterra R., Campana S., Vergani S.D. et al. 2012, *ApJ* 749, 68:
- [79] Wang F.Y., Bromm V., Greif T. et al. 2012, *ApJ* 760, 27
- [80] Schneider R., Ferrara A., Salvaterra R. et al. 2004, *MN* 351, 1379
- [81] Vreeswijk P.M., et al. 2004, *A&A* 419, 927
- [82] Black J.H., 1998, *Faraday Discuss.* 109, 257
- [83] Mao S., Mo H.J., 1998, *A&A* 339, L1
- [84] Choudhury T., Ferrara A. 2006, *MN* 371, 55
- [85] Alvarez M.A. et al., 2006, *ApJ* 639, 621
- [86] Mitra S., Choudhury T., Ferrara A. 2011, *MN* 419, 1480
- [87] Simcoe R.A. et al., 2004, *ApJ* 606, 92
- [88] Tornatore L., Ferrara A., Schneider R. 2007, *MN* 382, 945
- [89] Salvaterra R., Maio U., Ciardi B., et al. 2013, *MN* 429, 2718
- [90] Kawai N., et al. 2006, *Nat.* 440, 184
- [91] Ellis R.S., et al. 2013, *ApJ* 763, L7
- [92] Basa S., et al. 2012, *A&A* 542, A103
- [93] Savaglio S., Glazebrook K., Le Borgne D. 2009, *ApJ*, 691, 182
- [94] Bouwens R.J., et al. 2011, *ApJ* 737, 90
- [95] Noterdaeme P., Petitjean P., Carithers W.C. et al. 2012, *A&A* 547, L1
- [96] Jakobsson P., Fynbo J.P.U., Ledoux C., et al. 2006, *A&A* 460, L13
- [97] Fynbo J.P.U., Jakobsson P., Prochaska J.X., et al. 2009, *ApJS* 185, 526
- [98] Jakobsson P., Hjorth J., Malesani D., et al. 2012, *ApJ* 752, 62
- [99] Gallerani S., Ferrara A., Fan X., Choudhury T. 2008, *MN* 386, 359
- [100] Mesinger A., 2010, *MN* 407, 1328
- [101] McGeer I.D., Mesinger A., Fan X. 2011, *MN* 415, 3237
- [102] Ciardi B. et al. 2012, *MN* 423, 558
- [103] Barkana R., Loeb A., 2004, *ApJ* 601, 64
- [104] McQuinn M., et al. 2008, *MN* 388, 1101
- [105] Mesinger A. et al. 2004, *ApJ* 613, 23
- [106] Totani T., Kawai N., Kosugi G. et al. 2006, *PASJ* 58, 485
- [107] Mesinger A., Furlanetto S.R. 2008, *MN* 385, 1348
- [108] Inoue S., Salvaterra R., Choudhury T.R. et al. 2010, *MN* 404, 1938
- [109] Campana S., Thöne C.C., de Ugarte Postigo A. et al. 2010, *MN* 402, 2429
- [110] Campana S., Salvaterra R., Melandri A. et al. 2012, *MN* 421, 1697
- [111] Watson D., Jakobsson P. 2012, *ApJ* 754, 89
- [112] Behar E., et al. 2011, *ApJ* 734, 26
- [113] Starling R.L.C., Willingale R., Tanvir N.R. et al. 2013, *MN* (acc.; arXiv:1303.0844)
- [114] Nicastro F., et al. 2005, *ApJ* 629, 700
- [115] Campana S., Salvaterra R., Tagliaferri G. et al. 2011, *MN* 410, 1611
- [116] Becker J.K., 2008, *Phys. Rep.* 458, 172
- [117] Waxman E., Bahcall J. 1997, *PRL* 78, 2292
- [118] Bloom J.S. et al. 2009, *Astro2010 White paper*, arXiv:0902.1527
- [119] Phinney E.S. 2009, *Astro2010 White paper*, arXiv:0903.0098
- [120] Coward D.M., et al., 2012, *MN* 425, 2668
- [121] Klimenko S. et al., 2011, *PhRvD* 83, 102001
- [122] Dalal N. et al., 2006, *PhRvD* 74, 063006
- [123] Sathyaprakash B. et al 2010, *Class. Quant. Grav.* 27, 215006
- [124] Li L.-X., Paczyński B. 1998, *ApJ* 507, L59
- [125] Dessart L., et al. 2009, *ApJ* 690, 1681
- [126] Piran T., Nakar E., Rosswog S. 2013, *MN* 430, 2121 ack98
- [127] Sathyaprakash B.S., Schutz B.F., 2009, *Liv. Rev. Rel.* 12, No. 2
- [128] Cutler C., Thorne K.S., 2002, in: 16th GRG Conference, *World Sci. Publ.*, p. 72
- [129] Granot J. 2003, *ApJ* 596, L17
- [130] Toma K., et al. 2009, *ApJ* 698, 1042
- [131] Granot J., Königl A., 2003, *ApJ* 594, L83
- [132] Lyutikov M., Pariev V.I., Blandford R.D. 2003, *ApJ* 597, 998
- [133] Waxman E., 2003, *Nat.* 423, 388
- [134] Beloborodov A.M. 2011, *ApJ* 737, 68
- [135] Kalemci E., Boggs S. E., Kouveliotou C., et al. 2007, *ApJ Suppl.* 169, 75
- [136] McGlynn S., Clark D.J., Dean A.J. et al. 2007, *A&A* 466, 895
- [137] McGlynn S. et al., 2009, *A&A* 499, 465
- [138] Götz D., Laurent P., Lebrun F. et al. 2009, *ApJ* 695, L208
- [139] Yonetoku D. et al., 2011, *ApJ* 743, L30
- [140] Yonetoku D. et al., 2012, *ApJ* 758, L1
- [141] Götz D., Covino S. Fernandez-Soto A. et al. 2013, *MN* (in press, arXiv:1302.4186)
- [142] Fan Y.-Z., Wei D., Xu D. 2007, *MN* 376, 1857
- [143] Laurent P., Götz D., Binétruy P. et al. 2011, *Phys. Rev. D* 83, 12, 121301
- [144] Toma K. et al., 2012, *Phys. Rev. D* 109, 24
- [145] Gal-Yam A. 2012, *Sci.* 337, 927
- [146] Bloom J.S., Giannios D., Metzger B.D. et al. 201, *Sci.* 333, 203
- [147] Komossa S. 2012, *EPJ Web Conf.* 39, id. 02001
- [148] Meegan C., et al. 2009, *ApJ* 709, 791
- [149] Hudec R., Pina L., Marsikova V., et al. 2011, *AIP Conf. Proc.* 1358, p. 423
- [150] Tichy V. et al. 2011, *NIM Phys. Res. A.* 633, p. S169
- [151] Burrows D.N., Fox D., Palmer D. et al. 2012, *Mem. S.A. It.* 21, p. 59
- [152] Gehrels N., Barthelmy S.D., Cannizzo J.K., 2012, *Proc. IAU Symp.* 285, p. 41
- [153] Greiner J., Mannheim K., Aharonian F. et al. 2012, *Exp. Astr.* 34, 551
- [154] Krühler T., Schady P., Greiner J. et al. 2011, *A&A* 526, A153
- [155] den Herder J.-W., Piro L., Ohashi T. et al. 2012, *Exp. Astr.* 34, 519
- [156] Feroci M., Stella L., van der Klis M. et al. 2012, *Exp. Astr.* 34, 415

STELLAR IMAGER

Spokesperson: APPOURCHAUX, Thierry
Institut d'Astrophysique Spatiale, Orsay, France
+(33) 1 69 85 86 29
Thierry.Appourchaux@ias.u-psud.fr

Supporters

MOSSER Benoît
GARCIA, Rafael, A.
CHAPLIN William
CHRISTENSEN-DALSGAARD Joergen
GIZON Laurent
GAULME Patrick
JACKIEWICZ Jason

Observatoire de Paris, Meudon, France
Service d'Astrophysique, Saclay, France
Birmingham University, Royaume Uni
Aarhus Universitet, Aarhus, Danemark
Max Planck Institute, Lindau, Allemagne
New Mexico State University, USA
New Mexico State University, USA

1 Où en est l'astérosismologie?

L'évolution des étoiles est fondamentale pour comprendre l'Univers, en particulier la détermination des distances et des âges, ainsi que l'évolution chimique de la galaxie. Ce point a été identifié par le groupe d'études "Evolution des étoiles et des planètes" d'ASTRONET comme un axe majeur de recherche (A Science Vision for European Astronomy, Juillet 2007, P.T. de Zeeuw & F.J. Molster).

L'évolution stellaire a pu être partiellement testée sur le Soleil en utilisant les techniques d'héliosismologie qui permettent d'en déduire la structure interne.

Les missions CoRoT et *Kepler* ont ensuite permis de réaliser l'application de ces méthodes aux étoiles : c'est l'astérosismologie.

Depuis 2009, l'astérosismologie a considérablement évolué grâce à la grande quantité d'étoiles observées par les missions CoRoT et *Kepler*, ce qui a permis de développer des relations d'échelle pour remonter aux paramètres stellaires en utilisant les mesures astérosismiques (Bedding, 2011 et ses références; Mosser et al, 2012a).

La physique stellaire a pu faire de très grand progrès avec l'étude de la structure interne des

géantes rouges (Bedding, 2011; Mosser, 2012 et ses références).

La détection de modes dits mixtes a aussi permis de sonder l'intérieur profond des étoiles, le cœur des étoiles. La détection de ces modes permet notamment de mieux contraindre l'âge des étoiles ainsi que de permettre la mesure de la rotation interne des géantes rouges et sous-géantes (Mosser et al 2012; Deheuvels et al, 2012).

Les mesures sur la dynamique interne des géantes rouges ainsi que des sous-géantes vont aussi permettre de mieux comprendre le transport de moment angulaire dans différentes étoiles (Marques et al, 2013).

La découverte récente de plusieurs étoiles binaires physiques pour lesquels on dispose d'informations sismiques est en train de fournir aussi des contraintes sur les modèles d'évolution stellaire (Metcalfé et al. 2012; Appourchaux et al. 2012; White et al. 2012)

CoRoT et *Kepler* ont ainsi permis de paver le diagramme Hertzsprung-Russell.

La mission PLATO (Catala et al, 2011), en lice pour être la mission M3 du programme Cosmic Vision de l'ESA, a pour but de généraliser les

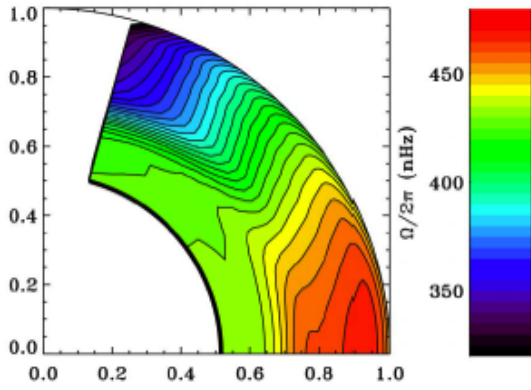


Figure 1 : Rotation solaire ($\Omega/2\pi$ en nHz) en fonction de la profondeur (la surface est à 1.0). La rotation différentielle observée en surface perdue dans la zone convective (d'après Schou *et al.*, 1998).

observations pionnières de CoRoT et *Kepler* en augmentant de plusieurs ordres de grandeurs leurs performances (en termes du nombre de cibles observées par exemple). Mais ces deux mission ne permettent de faire des mesures de la rotation interne des étoiles sous la surface comme nous pouvons le faire avec le Soleil (Figure 1)

C'est la connaissance des caractéristiques internes des étoiles qui permettra de répondre à des questions fondamentales comme la génération des champs magnétiques stellaires ou l'évolution de la rotation interne des étoiles, tous deux ayant une incidence importante sur l'évolution générale de l'étoile, ainsi que sur l'éventuel cortège planétaire qui les

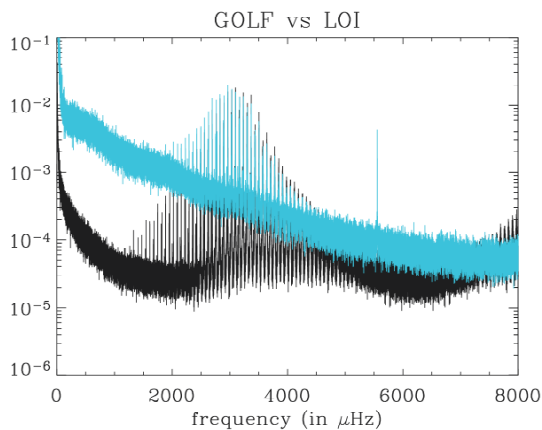


Figure 2 : Spectres de puissance observés par GOLF sur SOHO (vitesse radiale, noir) et par le LOI de VIRGO sur SOHO (intensité, bleu) obtenus en regardant le Soleil comme une étoile. Le bruit solaire en dessous de 2000 μHz est environ 100 fois plus fort en intensité qu'en vitesse radiale.

accompagne. Le projet STELLAR IMAGER se propose justement d'étudier les relations entre la rotation des étoiles et leur champ magnétique et leur influence sur les exoplanètes.

Dans ce contexte, les missions spatiales CoRoT, *Kepler*, TESS et peut-être PLATO ne pourront donner que des réponses partielles quant à la rotation interne des étoiles. Notamment, la connaissance de la rotation interne des étoiles de type solaire est limitée de par l'impossibilité d'explorer la structure interne dans les zones convectives de ces étoiles. En effet, les observations actuelles ne permettent pas de mesurer les degrés des modes supérieurs à $l=4$ qui sont nécessaires pour les inversions de structure interne dans ces étoiles (Voir la figure 1 pour le cas solaire).

STELLAR IMAGER se propose donc de permettre de mesurer les modes d'oscillations ($l \geq 4$) en faisant de l'imagerie de ces étoiles pour en déduire la dynamique interne des étoiles de type solaire (Figure 1). Ce projet est ambitieux et nous proposons différentes solutions pour y arriver.

2 Le futur

2.1 Vitesse radiale ou intensité ?

L'astérosismologie est rendue possible grâce à l'observation des perturbations de la surface de l'étoile dues à la propagation d'ondes à l'intérieur, soit en mesurant les fluctuations d'intensité lumineuse de l'étoile (comme sur l'instrument VIRGO de SOHO ou comme sur les missions CoRoT et *Kepler*), soit en mesurant les variations de vitesses radiales d'une raie d'absorption photosphérique (comme avec les instruments GOLF, MDI de SOHO, ou le spectromètre stellaire HARPS au sol). Les missions de sismologie à bord de SoHO ont montré que les modes d'oscillations sont plus facilement détectables en vitesse radiale qu'en intensité (Figure 2). Cela n'est en rien dû à des limitations instrumentales : c'est la contribution du bruit de fond stellaire (la granulation principalement) qui donne un rapport signal sur bruit bien meilleur en mesure de vitesse radiale qu'en photométrie.

Les observations en vitesse radiale permettent ainsi de mesurer beaucoup plus de modes (environ 2 fois plus), mais surtout de détecter des modes avec des durées de vie beaucoup plus

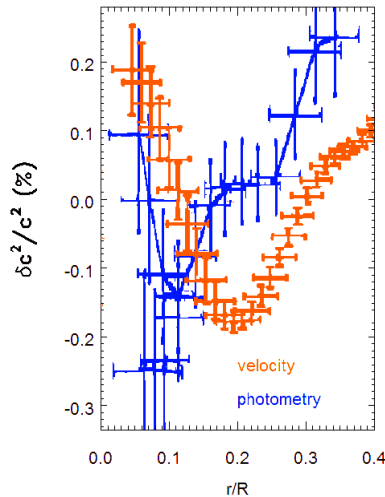


Figure 3: Inversion de la vitesse du son en fonction de la profondeur pour des observations de 6 mois obtenus avec VIRGO et avec GOLF (adapté de Gabriel et al 1997 et de Appourchaux et al 1997). La précision obtenue sur l'inversion de structure est 4 fois plus élevée.

longue et d'amplitude plus faible. La conséquence immédiate est que les modes de longue durée de vie sont mesurés avec des précisions beaucoup plus élevées (de l'ordre de l'inverse de la racine de la durée de vie). Les durées de vie détectées sont ainsi 10 fois plus longues, fournissant des fréquences avec une précision 3 fois plus élevée. L'impact sur la mesure de la rotation est immédiat : on peut ainsi obtenir une précision plus élevée sur les dédoublements rotationnels par exemple, et donc sur la rotation interne. La Figure 3 montre l'impact sur les inversions avec des observations de 6 mois contemporaines. L'intérêt des mesures de vitesse radiale est triple :

- Une durée de mesure réduite par typiquement au moins un facteur 5 (nombre de modes plus élevé et meilleur rapport signal à bruit)
- Une résolution spatiale des couches profondes des étoiles améliorée par un facteur 2 (nombre de modes plus élevé)
- Une rotation interne plus précise par un facteur 3 (détection de modes de durée vie longue)

En contrepartie, il faut noter que les observations Doppler sont plus difficiles à obtenir en mode multi-objets qu'en photométrie et ne pourront donc être menées que sur un moindre nombre d'objets.

Ces différents arguments plaident en faveur d'observations en vitesse radiale plutôt qu'en photométrie.

2.2 Imagerie ou non-imagerie ?

Les étoiles sont situées suffisamment loin pour que l'image de leur surface ne soit pas directement accessible avec un télescope monolithique de dimension raisonnable (diamètre < 20 m). L'héliosismologie a démontré que des mesures à résolution spatiale moyenne à élevée étaient fondamentales pour connaître la structure interne du Soleil et notamment la rotation dans la zone convective. L'imagerie de la surface des étoiles présenterait quantité d'informations des plus intéressantes, comme par exemple la quantité et la forme de taches magnétiques, la mesure directe de la rotation et sa variation avec la latitude ou de l'assombrissement centre-bord.

Il serait aussi extrêmement intéressant de pouvoir faire des images des étoiles dans le cadre d'analyses sismiques, pour en déduire avec une résolution sans précédent la dynamique interne des étoiles (voir Figure 1). De telles mesures auront un impact sur nos modèles de dynamo stellaire par exemple. Un concept d'imageur stellaire (Stellar Imager) était en cours d'étude à la NASA. Cette mission consistait en un réseau d'une trentaine de télescopes de 1 mètre de diamètre pour un diamètre équivalent de 500 mètres permettant d'obtenir environ un millier d'éléments d'image sur une étoile (Schrijver, Carpenter et Karovska, 2007). D'autres idées existent tel que le projet Luciola basé sur le concept d'hypertélescope qui permettrait des bases de l'ordre du km aussi (Labeyrie et al. 2009). Une feuille de route pour obtenir de tels hypertélescopes a été établie par Labeyrie et al. (2008). Derrière un tel Imageur Stellaire, la détection du signal d'oscillations stellaires pourrait se faire soit par détection de fluctuations d'intensité, soit par mesure de vitesse radiale.

Ces arguments plaident donc en faveur d'observations avec résolution spatiale.

3 Quelle instrumentation ?

Les instruments capables de détecter des fluctuations d'intensité de quelques parties par million à des périodes de quelques minutes sont réalisables facilement, soit avec des photodiodes (comme pour les instruments de VIRGO à bord de SoHO) ou avec des détecteurs à couplage de charge (CCD) tels que ceux de la mission CoRoT. Cette aisance a toujours été privilégiée

pour les différentes missions étudiées pour faire de l'astérosismologie ainsi que pour les missions en cours (*Kepler*), approuvée (TESS) ou à venir (PLATO).

Comme nous l'avons indiqué précédemment, les mesures de vitesses radiales fourniront beaucoup plus d'informations sur la structure interne des étoiles. Le développement de mesures spectroscopiques pour la détection des oscillations stellaires remonte à Connes (1985) avec l'idée du Absolute Astronomical Accelerometer. C'est d'ailleurs ce concept instrumental utilisé dans HARPS, quoi que simplifié, qui a été à la base de la détection des premiers modes stellaires sur α Cen (Bouchy et Carrier, 2001). Des instruments du type de HARPS sont capables de réaliser ces détections avec une précision suffisante (Pepe et al, 2000). C'est d'ailleurs sur ce concept qu'est basé le Stellar Oscillation Network Group (SONG) qui permettra d'observer depuis le sol les étoiles les plus brillantes avec des télescopes de 1 m de diamètre avec une précision de 1 m.s^{-1} (Grundahl et al. 2011)

Le concept de tachymètre de Fourier permet aussi d'obtenir de telles précisions avec une résolution spectrale plus faible (Mosser et al, 2003). Ce dernier concept a d'ailleurs été proposé pour des observations au Dôme C en Antarctique (Mosser et al, 2007). L'avantage de la solution avec un tachymètre de Fourier provient du rapport performance/masse. Un spectromètre échelle efficace pour les mesures de vitesse radiale doit avoir un pouvoir de résolution au moins supérieur à 50 000 (115 000 pour l'instrument HARPS, jusqu'à 120 000 pour SONG), ce qui nécessite simultanément un instrument de très grande taille et un CCD de très grand format. A l'opposé, un tachymètre Doppler est compact. Notons que le réseau d'héliosismologie au sol GONG, tout comme MDI à bord de SoHO et HMI à bord de SDO, sont des tachymètres de Fourier.

C'était également sur un tachymètre de Fourier que s'appuyait l'imageur Doppler Echoes / SYMPA (Soulat et al, 2012; Gaulme et al, 2008; Schmider et al, 2007) qui fut proposé pour l'étude des vents et de la sismologie joviens à bord de la mission JUICE de l'ESA. L'instrument proposé actuellement, dont le PI est N. Murphy (JPL), est basé sur l'utilisation d'une

seule raie d'absorption, ce qui n'est pas adapté pour l'astérosismologie.

Pour permettre l'imagerie, cette instrumentation serait donc placée *derrière* un télescope permettant de faire de la haute résolution spatiale (interférométrie à longue base): l'Imageur Stellaire.

Ces arguments plaident donc pour une instrumentation avec un niveau de spatialisation élevé (photomètre) avant d'utiliser une instrumentation plus performante (spectromètre) placé derrière le STELLAR IMAGER.

4 Pourquoi une mission spatiale?

L'héliosismologie en réseau et les observations astérosismiques avec des spectromètres échelle ont montré que le signal Doppler des oscillations de type solaire, de faible amplitude, peut être très valablement mesuré avec des observations au sol. Néanmoins, les observations au sol, qu'elles soient basées sur un réseau ou en Antarctique, sont limitées à 3 mois pour les étoiles (Mosser & Aristidi 2007).

Les mesures photométriques sur les étoiles sont en revanche impossibles à faire du sol essentiellement à cause du niveau de scintillation très élevé (Appourchaux et al, 1993).

En astérosismologie comme en héliosismologie, il est absolument essentiel d'avoir des mesures sur des durées longues (des mois, voire des années) et aussi ininterrompues que possible (cycle utile supérieur à 80 %). Ces durées d'observation plus longues, de l'année à quelques années, ne sont possibles que dans l'espace. Cette durée est absolument fondamentale pour la détection des modes de faible amplitude et de durée de vie longue, tels les modes mixtes présents dans les sous-géantes et géantes rouges, qui apportent directement l'information du cœur stellaire (Deheuvels et al. 2012; Mosser et al. 2012). C'est d'ailleurs cette propriété qui est déjà utilisée par *Kepler* et le serait par PLATO.

Bien sûr, si des observations de quelques mois sont possibles avec des instruments observant les étoiles sans résolution spatiale, il n'en va pas de même si on veut faire de l'imagerie. Comme

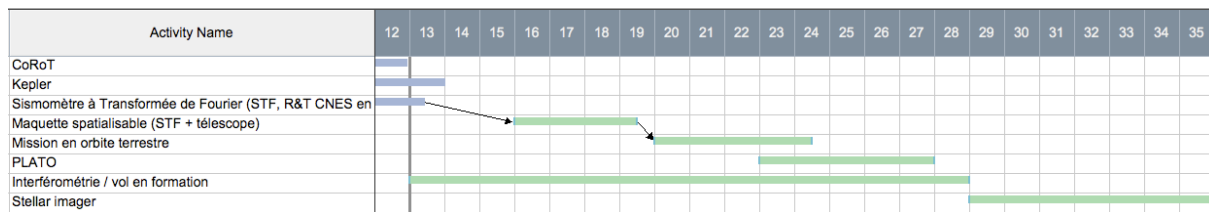


Figure 4: Implémentation temporelle de la feuille de route pour permettre des mesures de vitesses radiales de très longues durées pour l’astérosismologie. Les barres bleues représentent les projets en cours ou à lancer. Les barres vertes représentent les projets en étude. Les activités en rouge représentent les idées spécifiques de cette réponse.

nous l’avons vu précédemment, un *télescope* de diamètre équivalent à 500 mètres pointé pendant une année n’est concevable que dans l’espace.

Ce projet à long terme ne peut se faire que si une feuille de route est établie de manière claire et ciblée.

5 Une possible feuille de route

Le développement d’un instrument capable de faire des images des étoiles et des mesures astérosismiques peut se réaliser en plusieurs étapes (Figure 4 et Table 1).

Des études de faisabilité d’un concept d’Imageur Stellaire sont nécessaires pour permettre de mener à bien les objectifs envisagés. D’autres domaines scientifiques pourront bénéficier de tels développements (voir le projet Luciola; Labeyrie et al. 2009).

Les mesures photométriques dans l’espace ayant un héritage élevé, on peut d’abord valider la détection des modes d’oscillations avec un démonstrateur de l’Imageur Stellaire avec de la photométrie.

Le développement d’un spectromètre de type tachymètre Doppler spatialisable pour les années 2020-2030 demande des développements R&T qui pourront se mettre en place au sol d’abord, pour ensuite être utilisés à terme sur un possible projet spatial, dont l’ultime déclinaison serait donc utilisée sur un Imageur Stellaire

Instrumentation	Technologie	Support	Science
Sismomètre à TF au sol (STF)	Validation du concept de sismologie par TF	R&T CNES en cours	Sismologie d’une dizaine d’étoiles. Comparaison STF / CoRoT
STF sur une mission en orbite terrestre	Validation d’un STF spatialisé	R&T CNES, Financement CNES	Sismologie de quelques étoiles (durées longues)
Interférométrie spatiale	Validation des techniques de vol en formation et des d’interférométrie	Collaboration inter-agences	TBD
Démonstrateur de l’imageur stellaire	Interférométrie et sismologie en photométrie, hypertélescope	Collaboration inter-agences	Rotation interne des étoiles brillantes
Imageur stellaire	Interférométrie et sismologie par TF	Collaboration inter-agences, financement CNES	Etudes des zones convectives stellaires et de leur dynamique (rotation)

Table 1: Implémentation de la feuille de route en termes de R&T et de Science

Le concept de spectromètre serait très proche de celui des projet SIAMOIS ou Echoes (Mosser et al, 2007; Soulat et al, 2012). Cet instrument serait constitué d’un télescope de 40 cm de diamètre relié à spectromètre à Transformée de Fourier par une fibre optique. Cet instrument

pourrait être d’abord installé au sol. Ensuite, une version de cet instrument pourrait être spatialisée pour inclusion sur des missions interplanétaires et observer durant les phases de croisière, en ayant le souci d’impact faible voire nul sur les ressources du satellite (faible

puissance, pas de manœuvres spécifiques, autonomie opérationnelle). Une fois que le concept aura démontré ses capacités, il pourra être implémenté sur une mission telle que l'Imageur Stellaire dont les capacités de détection sismique auront été démontrées auparavant avec un instrument photométrique.

Pour la réalisation d'une telle feuille de route à aussi long terme, il serait envisageable d'avoir un soutien de type R&T et des études de Phase 0 avant une possible implémentation sur une mission spatiale.

6 La communauté intéressée

La communauté française a toujours été moteur dans le développement de l'héliosismologie et de l'astérosismologie avec la mise au point d'instrument sur SOHO et la mise en place de la mission CoRoT. Les laboratoires intéressés sont les suivants :

- Institut d'Astrophysique Spatiale
- Observatoire de Paris
- Observatoire de Toulouse
- Observatoire de Nice
- Service d'Astrophysique du CEA

Les laboratoires impliqués pourraient mettre en place la R&T et les développements spatiaux nécessaires à l'implémentation de la feuille de route.

La communauté européenne est très active en astérosismologie. L'Université d'Aarhus a un rôle majeur dans l'animation du Kepler Asteroseismic Science Consortium ou KASC (J.Christensen-Dalsgaard, H.Kjeldsen) ainsi que dans la mise du réseau d'observation stellaire SONG (Stellar Oscillation Network Group, dont le chef de projet est F.Grundahl). L'Université de Birmingham a aussi un rôle majeur dans le KASC est en train de développer un groupe qui attire de jeunes chercheurs autour de William Chaplin. Finalement le groupe d'héliosismologie et astérosismologie de Laurent Gizon à Katlenburg-Lindau consiste un autre grand pôle d'attraction de la recherche en astérosismologie.

La communauté s'est donc agrandi autour du noyau historique français et montre que la mise en place aussi bien technique que scientifique du Stellar Imager pourra se faire dans le futur à une échéance de 10 ans voire 20 ans.

Références

- Appourchaux, T. et al, 1993, PRISMA - Probing Rotation and Interior of Stars: Microvariability and Activity. Report on the phase A study., European Space Agency, Paris (France)
- Appourchaux, T. et al, 1997, Structure inversions with the VIRGO data, in *Sounding Solar and Stellar Interiors*, IAU Symp. 181, J. Provost and F.-X. Schmider eds., 159
- Appourchaux, T. et al, 2012, A serendipitously discovered seismic binary, KASCV, Lac Balaton, Hongrie
- Bouchy, F. & Carrier, F., 2001, P-mode observations on α Cen A, *A&A*, **374**, L5
- Catala, C. et al, 2011, PLATO: PLANetary Transits and Oscillations of stars, *Journal of Physics Conference Series*, **271**, 012084
- Connes, P., 1985, Absolute Astronomical Accelerometry, *Astrophysics and Space Science*, **110**, 211
- Bedding, T. R., 2011, Solar-like Oscillations: An Observational Perspective", ArXiv e-prints, arXiv:1107.1723
- Deheuvels, S. et al, 2012, Seismic Evidence for a Rapidly Rotating Core in a Lower-giant-branch Star Observed with Kepler, *ApJ*, **756**, 19
- Gabriel, A.H. et al, 1997, Performances and first results from the GOLF instrument on SOHO, in *Sounding Solar and Stellar Interiors*, IAU Symp. 181, J. Provost and F.-X. Schmider eds., 53
- Gaulme, P. et al, 2008, SYMPA, a dedicated instrument for Jovian Seismology. II. Real performance and first results, arXiv, *A&A*, **490**, 859
- Grundahl, F. et al, 2011, SONG - getting ready for the prototype, *Journal of Physics Conference Series*, **271**, 012083
- Labeyrie, A. et al, 2009, Luciola hypertelescope space observatory: versatile, upgradable high-resolution imaging, from stars to deep-field cosmology, *Experimental Astronomy*, **23**, 463
- Labeyrie, A. et al, 2008, Steps toward hypertelescopes on Earth and in space", *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, **7013**,
- Marques, J.P. et al., 2013, Seismic diagnostics for transport of angular momentum in stars. I.

Rotational splittings from the pre-main sequence to the red-giant branch, *A&A*, **549**, A74

Metcalfe, T.S. et al, 2012, Asteroseismology of the Solar Analogs 16 Cyg A and B from Kepler Observations, *ApJ*, **748**, L10

Mosser, B., Maillard, J.-P. & Bouchy, F., 2003, Photon Noise-limited Doppler Asteroseismology with a Fourier Transform Seismometer. I. Fundamental Performances, *PASP*, **115**, 990

Mosser, B. & Aristidi, E., 2007, Duty Cycle of Doppler Ground-based Asteroseismic Observations, *PASP*, **119**, 127

Mosser, B. & The Siamois Team, 2007, A Fourier Tachometer at Dome C in Antarctica, *CoAst*, **150**, 309

Mosser, B. et al, 2012a, Asymptotic and measured large frequency separations, accepté par *A&A*, ArXiv e-prints, arXiv:1212.1687

Mosser, B. et al, 2012b, Spin down of the core rotation in red giants, *A&A*, **548**, A10

Mosser, B., 2012, Red giant seismology: observations, ArXiv e-prints, arXiv:1210.7301

Pepe, F. et al, 2000, HARPS: a new high-resolution spectrograph for the search of extrasolar planets, *SPIE*, **4008**, 582

Schmider, F.-X. et al, 2007, SYMPA, a dedicated instrument for Jovian seismology. I. Principle and performance, *A&A*, **474**, 1073

Schou, J. et al, 1998, Helioseismic studies of differential rotation in the solar envelope by the solar oscillations investigation using the Michelson Doppler Imager, *ApJ*, **505**, 390

Schrijver, C.J., Carpenter, K. & Karovska, M., 2007, *CoAst*, **150**, 364

Soulat, L. et al, 2012, Echoes: a new instrumental concept of spectro-imaging for Jovian seismology, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 8442

White, T. et al, 2012, Kepler observations of the asteroseismic binary HD 176465, KASCV, Lac Balaton, Hongrie

Chronos :

A NIR Spectroscopic Galaxy Formation Survey

A white paper submitted to the European Space Agency
for the definition of the future L2/L3 missions.

POINT OF CONTACT:

Ignacio Ferreras

Mullard Space Science Laboratory

University College London

Holmbury St Mary, Dorking, Surrey RH5 6NT

United Kingdom

email: i.ferreras@ucl.ac.uk

phone: +44 (0)1483 204 125

fax: +44 (0)1483 278 312



Background Image: The NASA/ESA *Hubble Space Telescope* Ultra Deep Field (UDF) is one of the deepest exposures of distant galaxies. CHRONOS will observe equally faint galaxies, by exposing the equivalent of one UDF every week for five years, producing the deepest spectroscopic survey ever.

Chronos^{*}

A NIR Spectroscopic Galaxy Survey: From the formation of galaxies to the peak of activity

IGNACIO FERRERAS^{1†}, RAY SHARPLES²

Science Core Team

JAMES S. DUNLOP³, ANNA PASQUALI⁴, FRANCESCO LA BARBERA⁵, ALEXANDER VAZDEKIS⁶,
SADEGH KHOCHFAR³, MARK CROPPER¹, ANDREA CIMATTI⁷, MICHELE CIRASUOLO^{3,8},

RICHARD BOWER², JARLE BRINCHMANN⁹, BEN BURNINGHAM¹⁰, MICHELE CAPPELLARI¹¹,
STÉPHANE CHARLOT¹², CHRISTOPHER J. CONSELICE¹³, EMANUELE DADDI¹⁴, EVA K. GREBEL⁴,
ROB IVisON^{3,8}, MATT J. JARVIS¹¹, DAISUKE KAWATA¹, ROBERT C. KENNICUTT¹⁵,
TOM KITCHING¹, OFER LAHAV¹⁶, ROBERTO MAIOLINO¹⁵, MATHEW J. PAGE¹,
REYNIER F. PELETIER¹⁷, ANDREW PONTZEN¹⁶, JOSEPH SILK¹², VOLKER SPRINGEL^{4,18},
MARK SULLIVAN¹¹, IGNACIO TRUJILLO⁶, GILLIAN WRIGHT⁸

May 24, 2013



¹ Mullard Space Science Laboratory, University College London, UK; ² University of Durham, UK; ³ Royal Observatory, University of Edinburgh, UK; ⁴ ARI/ZAH, University of Heidelberg, Germany; ⁵ INAF/OAC, Naples, Italy; ⁶ Instituto de Astrofísica de Canarias, Spain; ⁷ Università di Bologna, Italy; ⁸ UK ATC, Edinburgh, UK; ⁹ Leiden University, the Netherlands; ¹⁰ University of Hertfordshire, UK; ¹¹ Oxford University, UK; ¹² Institut d'Astrophysique de Paris, France; ¹³ Nottingham University, UK; ¹⁴ CEA/Saclay, France; ¹⁵ Institute of Astronomy, University of Cambridge, UK; ¹⁶ Physics & Astronomy Department, University College London, UK; ¹⁷ Kapteyn Institute, University of Groningen, the Netherlands; ¹⁸ Heidelberg Institute for Theoretical Studies, Germany.

^{*}Chronos, the Greek god personifying time is the name we have given to this survey, as it is designed to understand the formation and evolution of galaxies across cosmic time

[†]i.ferreras@ucl.ac.uk

Executive Summary

In a decade or two from now, we will have made significant strides in our understanding of the early formation history of the Universe, through missions such as *Planck* and *Euclid*, and in its recent state, through *Gaia* and ground-based surveys such as SDSS and their more substantial successors. What will still be problematic is how we arrived here given the initial conditions. This is the realm of “baryon physics”, the nature of the formation and evolution of the galaxies. Understanding this is a colossal task, currently occupying a large fraction of the international astronomical community. It involves complex astrophysics at redshifts 1–6. To make the critical and substantial advances in our understanding of the essential nature of this process requires spectroscopy (for the astrophysics) in the infrared (because of the redshift) of a large volume of the Universe (to examine the critical effects of environment). This is what the concept presented here sets out to achieve.

We propose *Chronos*, an L-class mission to understand the formation and evolution of galaxies, by collecting the deepest NIR spectroscopic data, from the formation of the first galaxies at $z \sim 10$ to the peak of formation activity at $z \sim 1-3$. The strong emission from the atmospheric background makes this type of survey impossible from a ground-based observatory. The spectra of galaxies represent the equivalent of a *DNA fingerprint*, containing information about the past history of star formation and chemical enrichment. The proposed survey will allow us to dissect the formation process of galaxies including the timescales of quenching triggered by star formation or AGN activity, the effect of environment, the role of infall/outflow processes, or the connection between the galaxies and their underlying dark matter haloes. To provide these data, the mission requires a 2.5m space telescope optimised for a campaign of very deep NIR spectroscopy. A combination of a high multiplex and very long integration times will result in the deepest, largest, high-quality spectroscopic dataset of galaxies from $z=1$ to 12, spanning the history of the Universe, from 400 million to 6 billion years after the big bang, i.e. covering the most active half of cosmic history.

The highly demanding requirements results in a mission that is the spectroscopic equivalent of a *Hubble Space Telescope* obtaining one Ultra Deep Field (the deepest exposure of distant galaxies ever attained) every fortnight for five years. A two-tiered survey will provide a high quality stellar mass limited dataset of about 2 million spectra from galaxies covering the most important epochs of structure formation, back to the early phases soon after recombination. Although missions such as *Euclid* or *WFIRST* will provide low-resolution spectra in the NIR, the requirements of resolution and SNR in the continuum for the analysis of the properties of the underlying stellar populations at $z \lesssim 3$ are too demanding for them. Therefore, *Chronos* is the link between cosmology-orientated missions, such as *Planck* or *Euclid* and *Gaia*’s targeted exploration of our own galaxy. Our mission will gather key spectroscopic indicators of the properties of the underlying stellar populations in galaxies. This will be complemented by *Herschel*’s view of the “dusty” side of the Universe.

The survey will allow us to understand the most fundamental open questions in galaxy formation today: the connection between the star formation and the mass assembly history of galaxies; the interplay of star formation and activity from a central supermassive black hole in shaping the properties of galaxies; the connection between chemical composition and the formation histories of galaxies (“extragalactic archaeology”); and the contribution of the environment and the pervading dark matter halos to the formation of galaxies.

The main science questions that the mission will answer are:

- The connection between the star formation history and the mass assembly history.
- The role of AGN and supernova feedback in shaping the formation histories of galaxies, with a quantitative estimate of quenching timescales.
- The formation of the first galaxies.
- The source of reionization.
- Evolution of the metallicity-mass relation, including $[\alpha/\text{Fe}]$ and individual abundances.
- Initial Mass Function as a tracer of star formation modes.

I. A NATURAL CHOICE FOR THE NEXT L-CLASS MISSION

Even though astronomy as a human activity goes back to the origins of civilization, and the understanding of celestial bodies has been responsible for the scientific method that led to our technology-based society, some of the most fundamental aspects of this scientific discipline remain largely unsolved. It was less than a century ago that we began to understand the true meaning of galaxies as “Universe islands”, and even more recently that we could put them in context with the underlying dark matter backbone and the evolution of the Universe as a whole. The luminous component of galaxies is dominated by stars, gas and dust. This L-class mission is designed to address one of the major questions in ESA’s “COSMIC VISION 2015-2025” BR-247 document, namely question 4: **How did the Universe originate and what is it made of?**, and especially topic 4.2: **The Universe taking shape.**

Two of ESA’s recent cornerstone missions concentrate on very specific – and essential – aspects of this problem, namely the evolution of the dust in star forming galaxies (*Herschel*) and the history of our own Milky Way galaxy (*Gaia*). The global structure and evolution of the Universe is the main target of ESA’s *Planck* and *Euclid* missions. Now, the next Large mission should tackle the much wider problem of understanding the “baryon physics” of galaxy formation, namely the highly complex set of physical mechanisms responsible for the transformation of the primordial hydrogen and helium gas mixture into the galaxies we see today. At a more fundamental level, it is possible to understand the nature of dark matter only if we have a comprehensive understanding of galaxy formation. Such an endeavour requires the highest quality spectroscopic dataset over a large survey, probing two major cosmic epochs: a) the epoch at the peak of galaxy formation activity, namely between redshifts $z \sim 3$ and 1 (i.e. 2 and 6 billion years after the Big Bang), and b) the formation of the first galaxies, at redshifts between $z=12$ and 6 (between 400 and

900 million years after the Big Bang). The spectroscopic analysis of galaxies constitutes the equivalent of DNA fingerprinting, allowing us to determine the composition of the galaxy and its past formation history.

Even though ongoing and future surveys, such as *Euclid*, *WFIRST*, *LSST* or *SKA*¹ will give insightful clues about the origins of galaxies, high quality spectroscopic data at moderately high resolution ($R \equiv \lambda/\Delta\lambda \approx 1,000 - 2000$) is the *only* way that this problem can be solved. While ground-based NIR spectroscopy will undoubtedly make important strides in the coming decades, OH and other backgrounds inherent to ground based NIR spectroscopy will ensure that the spectroscopic performance of *Chronos* – at the faintness levels and spectral coverage required for this science – will remain unchallenged. The behemoths of astronomical science in the coming decades (*JWST* from space or the *E-ELT* from the ground) will, in all likelihood, target in detail specific issues of this project, but their very limited field of view makes them incapable of gathering anywhere near the scale of dataset required to make the extensive advances *Chronos* will deliver in the area of galaxy formation and evolution. Rather than a competitor, ultra-large observing facilities will be a complement to *Chronos*.

Chronos in a nutshell

Chronos is a dedicated 2.5m space telescope optimized for ultra-deep NIR spectroscopy at moderate resolution ($R=1500$) in the 0.9-1.8 μ m range. The 5-year long, two-tiered survey will reach $H_{AB}=26$ over a 100 deg², and $H_{AB}=27.2$ over 10 deg² at a 5σ level in the continuum. The two main science drivers are the formation of galaxies at the peak of activity ($1 < z < 3$) and the first galaxies and the source of reionization ($z > 6$).

II. THE EVOLUTION OF GALAXIES AT THE PEAK OF ACTIVITY

II.1 Introduction

Over the past two decades, advances in detector technology have allowed us to probe the evolution of star formation with cosmic time (Fig. 1, left panel). Between the formation of the first galaxies and the present time, there was an

epoch when the global star formation history was at its peak, in the redshift interval between $z \sim 3$ and 1. During this epoch, the majority of the stars in the present Universe were formed and assembled into galaxies, making it – along with the very first epoch of galaxy formation (see Sec. III) – the most important interval of cosmic history. In

¹Although *SKA* will deliver an unprecedented map of the first stages of structure formation, its design will only target the gas component through observations of the HI 21 cm line.

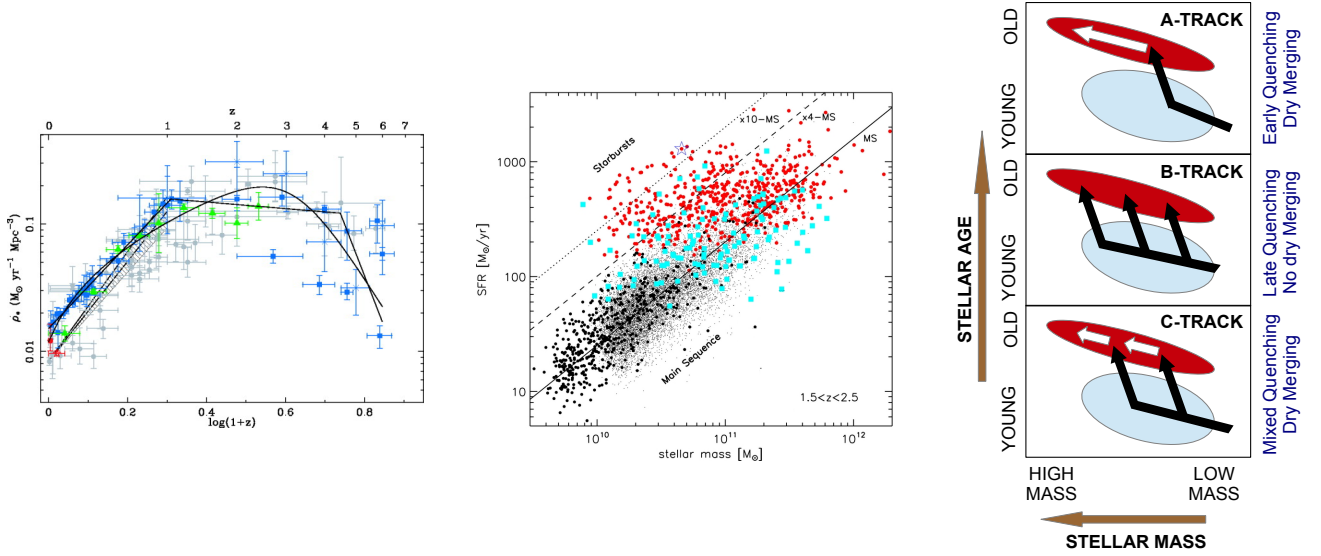


Figure 1: **Left:** Cosmic star formation history (from Hopkins & Beacom, 2006). **Middle:** Different modes of star formation (from Rodighiero *et al.*, 2011)). **Right:** Schematics of galaxy evolution from the blue cloud to the red sequence (adapted from Faber *et al.*, 2007).

the “local” Universe (i.e. $z \lesssim 0.1$), large spectroscopic surveys, most notably the Sloan Digital Sky Survey (hereafter SDSS, York *et al.*, 2000) provided enough data to trigger a quantum leap in our understanding of galaxy formation. In the post-SDSS era it is possible to dissect datasets according to various properties such as stellar mass, velocity dispersion, morphology, environment, enabling us to “ask the right questions”. SDSS has shown that gathering a *complete* database comprising up to a million spectra is necessary to constrain in detail the many aspects that contribute to the process of galaxy formation. The spectroscopic analysis of the stellar populations in the SDSS Universe confirmed in exquisite detail the strong bimodality between passive, and predominantly massive galaxies (the so-called *red sequence*) and a *blue cloud* of star forming systems (Baldry *et al.*, 2004). Furthermore, SDSS also revealed the presence of a characteristic mass scale in the local Universe, around 3×10^{10} solar masses in stars (Kauffmann *et al.*, 2003), which marks a clear transition in the baryon content of galaxies (Moster *et al.*, 2010), reflecting different modes of formation, either at a fundamental level down to star forming regions, and/or at a global level reflecting the contribution from a supermassive black hole, from feedback associated to star formation, or even from the environment where galaxies live.

Star formation in galaxies seems to have two possible channels: “normal” star forming galaxies, such as our own Milky Way galaxy; and so-called starburst galaxies, where an intense rate of formation implies a much higher efficiency in the conversion of gas into stars. Fig. 1 (middle panel) shows that galaxies in the $1 < z < 3$ redshift window have a wide range of star formation efficiencies, from the “quiescent” main sequence phase to intense starbursts. Various processes involving star formation, quenching and

mergers have been invoked to explain the observed trends (Fig. 1, right panel). A very large number of papers have been devoted to this open problem, at the observational, theoretical and modelling levels. However, a definitive answer beyond simple sketches of the evolution is possible only with detailed spectroscopic information about the stellar content of galaxies over the peak of formation activity, i.e. $z \lesssim 3$. *Chronos* will obtain detailed formation histories over the range of stellar mass, redshift, and environment, required to decipher the mechanisms that control the efficiency of star formation.

... in a nutshell

- Complete sample of galaxies out to $z \sim 3$ down to $10^{10} M_\odot$ in stellar mass.
- Accurate assessment of environment over the most active period of galaxy formation, probing the interplay between dark matter and baryons.
- Detailed age, metallicity, abundance ratios, IMF: extragalactic archaeology.
- Understanding the mechanisms controlling the growth of galaxies: infall, outflows, AGN feedback, supernovae-driven winds.

Why do we need a new spectroscopic survey?

In the local Universe, SDSS provides detailed high quality spectroscopic information only out to a relatively modest apparent magnitude, and samples used for detailed spectroscopic analysis are often restricted to $z \lesssim 0.1$. In addition, the passband-shifting effect as we move into high redshift implies that the region around the 4000\AA break – which is highly sensitive to the properties of the stellar

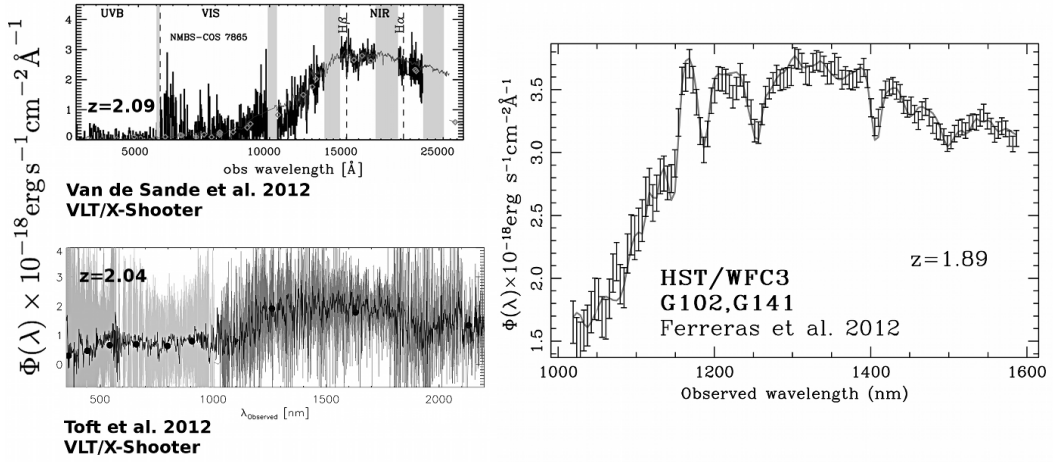


Figure 2: Comparison of NIR spectra from the ground (VLT/X-Shooter, left, comprising integration times of 5-7 hours) and from space (HST/WFC3 slitless grisms, with an integration of just over 1 hour). These galaxies are very massive, with an apparent magnitude $H_{AB} \sim 20$. *Chronos* will extract ~ 1 – 2 million spectra down to $H_{AB} \sim 25$ – 26 .

populations – moves into the Near Infrared (NIR), so that future spectroscopic surveys in the optical region, such as MS-DESI (restricted to $\lambda < 1 \mu\text{m}$) will not be able to target galaxy formation at the peak of activity. In the NIR, ground-based observations are hampered by the high atmospheric background. Fig. 2 compares state of the art ground-based NIR spectroscopy of $z \sim 2$ galaxies taken by the X-Shooter instrument at ESO’s *Very Large Telescope* with a simple slitless grism spectrum taken by the WFC3 on board the *Hubble Space Telescope* at a similar redshift. The ground-based data was obtained at a higher spectral resolution, nevertheless, the difference in SNR is remarkable. Even though the field of view and spectral resolution of the *HST* data falls far below our target specifications, the figure illustrates that a 2.5m telescope in space is capable of superb deep NIR spectroscopy unrivalled from the ground. In addition, the restriction of ground-based observations to the allowed bands *J, H, K* introduce “redshift gaps” that will prevent a comprehensive study. Quoting Silk & Mamon (2012): “Ultimately, one needs a spectroscopic survey akin to SDSS at $z=1$ – 2 ”. However, even with the large field of view provided by Subaru’s prime focus instruments (e.g. PFS), the quality required for a comprehensive analysis of the stellar populations of galaxies over the $z \sim 1$ – 3 range requires a space telescope. Over a 5 year period, *Chronos* will deliver millions of high quality spectra plunging down to a flux level between 100 and 500 times lower than those shown in Fig. 2. In the future, observatories such as the *E-ELT* from the ground or *JWST* from space will be capable of achieving such low flux levels. However, the very small field of view covered by these facilities will make surveys over many square degrees unfeasible. As a consequence, neither *E-ELT* nor *JWST* will be capable of investigating the large scale environment of high redshift galaxies. *Chronos* will be the only facility able to provide a large dataset of deep, high quality spectroscopic data in the NIR over large areas

of the sky, required to properly assess the role of environment on the physical properties of galaxies and on their evolution. Neither broad-band nor medium-band photometric surveys such as DES, J-PAS or LSST can give enough “spectral resolution” to answer the key open questions of galaxy formation and evolution. Even at moderate resolution (e.g. ESA’s *Euclid* – and possibly NASA’s *WFIRST* – will provide $R \lesssim 600$ slitless grism spectroscopy, where the effective resolution is limited by the extent of the surface brightness profile of the galaxy), it will not be possible to obtain accurate constraints on the processes underlying the formation of stars in galaxies. Furthermore, as cosmology-orientated surveys, they are not designed to achieve high enough signal-to-noise ratio in the continuum for faint sources, a strict requirement in our mission. Nevertheless, for our purposes, *Euclid* is, rather, a valuable complement, as it will greatly help in the selection of targets for detailed spectroscopy with *Chronos*, providing in addition morphological information and photometry in several bands.

II.2 Probing galaxy formation through their stellar content

The redshift range of $z \sim 1$ – 3 is a fundamental epoch of galaxy formation for several reasons:

- (A) It is the peak of the cosmic star formation history (Hopkins & Beacom, 2006).
- (B) It is the peak of the AGN activity (Richards *et al.*, 2006).
- (C) It is the peak in the merger rate (Ryan *et al.*, 2008).
- (D) It is the epoch when hosting haloes of massive galaxies allows for cold accretion via cosmic streams (Dekel *et al.*, 2009).

Looking at the star formation history of galaxies today via moderately high resolution spectra reveals the integrated

star formation history of these galaxies over all their progenitors (Thomas *et al.*, 2005; De la Rosa *et al.*, 2011). The hierarchical paradigm of structure formation predicts that the number of such progenitors can be quite significant (e.g., Khochfar & Silk, 2006). However, based on observational constraints at $z=0$, it is not possible to estimate the importance of the role of such progenitors and hence the role of merging. This is mainly due to the ‘cosmic conspiracy’ of the star formation main sequence, which, to first order, shows a linear relation between star formation rate and stellar mass of the galaxy (Daddi *et al.*, 2007). Thus, by knowing the star formation rate of a present-day galaxy at any higher redshift, it is not possible to determine whether it formed all its stars in one main progenitor or in many. The *only* viable option to observationally relate the star formation history and the mass assembly history involves deep spectroscopic observations, probing the underlying stellar populations during the peak of activity. A complete mass-limited sample will serve as an fundamental, unbiased benchmark to relate galaxies at different redshifts with their merging histories.

Chronos will provide a definitive statement on the formation timescale of galaxies with respect to morphology, mass and environment. It will also deliver information about the velocity dispersion and chemical enrichment of the populations. Such studies not only constrain but also motivate significant developments in numerical simulations in a cosmological context, to achieve a more consistent view of how galaxies form and evolve. For instance, the evidence that massive galaxies are old and enhanced in Mg over Fe (e.g., Renzini, 2006, and references therein) points towards an early and rapid formation, thus constraining the timescales in which haloes in regions of the Universe that are destined to form a cluster collapse (e.g., De Lucia *et al.*, 2006). A powerful test of these models is to study differences in the stellar content of galaxies in different environments. However, the evolutionary trends of stellar populations can be hidden due to the age-metallicity degeneracy, which not only affects the colours but also to a great extent the absorption line-strength indices of the old stellar populations (e.g., Worthey, 1994), if there is a relation between the age and the metallicity of the galaxies (e.g., Ferreras *et al.*, 1999). Much progress has been made during the last decade to lift this degeneracy, however, the chief aspect of this spectroscopic survey is that the targeted redshift interval represents a range in lookback time that resolves the “age axis” directly. Furthermore, studying the galaxies when they were younger allows to derive much more accurate ages as, in this regime, the spectral indicators have much larger variations for smaller changes in the mean age (Jørgensen *et al.*, 2013).

Constraining the characteristic timescales for the formation of the bulk of the stellar populations has been a major endeavour. This is performed through the study of the chemical composition of galaxies derived from their spec-

tra. As different elements are released to the interstellar medium by stars of different masses and, therefore, over different timescales, stellar and gas abundance ratios (once suitably calibrated) provide potential cosmic ‘clocks’ capable of eliciting the timescale of star formation within a galaxy. It is important to understand how non-solar abundance-patterns might affect the main conclusions derived from the stellar population analysis. These timescales can be fine tuned if, apart from [Mg/Fe], other abundance ratios, such as [CN/Fe], are included in this analysis (e.g., Carretero *et al.*, 2004). However, the different behaviour of elements released by massive supernovae (e.g., Woosley *et al.*, 2002) remain unclear, among other reasons, because these studies are still in their infancy. The dearth of quality spectroscopic observational datasets over a wide range of cosmic time explains the deficiencies of our understanding on the distribution of chemical elements in galaxies.

Efficiency of star formation

The efficiency of converting baryons into stars within given dark matter haloes is of prime interest. Theoretical predictions of LCDM-based models suggest a state of self-regulation in which the star formation rate is controlled by the growth rate of dark matter haloes (see, e.g., Bower *et al.*, 2006; Bouché *et al.*, 2010; Bower *et al.*, 2012; Guo *et al.*, 2013). The unprecedented sample of galaxies that *Chronos* will provide over the $z \sim 1-3$ range will allow us to construct high precision correlation functions for the galaxy population, including detailed information about their star formation histories, relating galaxy growth with the underlying distribution of dark matter structure. In this way, it will be possible to link measured star formation histories to theoretically predicted growth rates of dark matter haloes. It has been shown in studies at low redshift, that by only using abundance matching techniques it is not possible to obtain robust constraints on the galaxy halo occupation function (Neistein *et al.*, 2011).

Galaxy Formation and the Initial Mass Function

One of the most fundamental properties of star formation, the stellar Initial Mass Function (IMF), which describes the mass distribution of stars at birth, is assumed to be universal and constant with cosmic time. Recently, it has been argued that the stellar initial mass function may not be universal; differences have been hinted in the most massive galaxies at $z \sim 0$ (van Dokkum & Conroy, 2010; Cappellari *et al.*, 2012). A combination of a large, high-quality spectroscopic dataset from SDSS and detailed population synthesis models (Vazdekis *et al.*, 2012) enabled the confirmation of a systematic trend with velocity dispersion (Ferreras *et al.*, 2013; La Barbera *et al.*, 2013). Such non-universality reflects fundamental differences in the mode of star formation with respect to galaxy mass (Hopkins, 2012), that need to be addressed in ab initio simulations of star formation (e.g. Bate *et al.*, 2003). Furthermore, recent theoretical developments suggest that the IMF shape

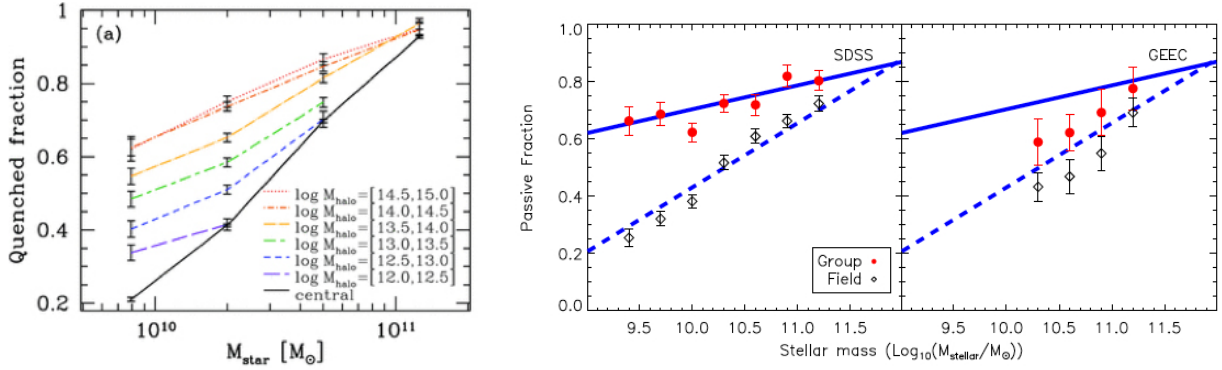


Figure 3: **Left:** The fraction of quenched galaxies (whose specific star formation rate is lower than 10^{-11} yr^{-1}) as a function of their stellar mass M_* , for galaxies residing in environments of different content of dark matter (M_{halo}). The solid line shows the variation of the fraction of quenched central galaxies (the most massive galaxy within each environment) in M_* (from Wetzel *et al.*, 2012). **Right:** The fraction of quenched galaxies as a function of their stellar mass in galaxy groups and in the field, as measured at $z \simeq 0$ in the SDSS survey and at $z \simeq 0.4$ in the GEEC survey (from McGee *et al.*, 2011).

and mass-cutoffs might depend on the star formation rate (Weidner *et al.*, 2011). However, the imprints of a varying IMF on spectra might be coupled to a variation of the abundance ratio of certain chemical species (Conroy & van Dokkum, 2012). Future population synthesis models (see II.6) along with high quality spectroscopic data of galaxies probing a wide range of cosmic time will allow us to disentangle these effects. With *Chronos*, it will be possible to probe the evolution of the IMF during the most important epoch of star formation, spanning a range of mass, velocity dispersion and metallicity. This issue is fundamental for an accurate assessment of the cosmic star formation history – which depends on the assumptions made for the underlying stellar populations. Moreover, derived star formation histories may have to be revised depending on these results, as a systematic change in the IMF can affect the model predictions relating the distribution of stellar ages and metallicities. We emphasize that such studies require high quality NIR spectroscopic data of very faint sources, such as those that *Chronos* will provide, beyond the capabilities of any spectroscopic survey in the coming decades.

II.3 The ecology of galaxies

In addition to the intrinsic mechanisms described above, the environment where galaxies live plays a fundamental role in shaping their evolution, as it is capable of quenching their star formation by removing their hot and cold gas reservoirs and to literally disrupt them by removing their stars. From the observed properties of galaxies at $z \simeq 0$ we have collected a large body of evidence for the occurrence of such environmental processes, but the determination of their timescales and amplitudes remains at a qualitative level. We have not yet established in a quantitative way how these parameters depend on environment and redshift, i.e. on the assembly history of a galaxy cluster or galaxy group, through cosmic time.

The unprecedented statistical power of SDSS, in terms of the photometric and spectroscopic properties of galaxies measured at optical wavelengths, has allowed us to describe the behaviour of the star formation activity of galaxies across many orders of magnitude with respect to their stellar mass and environment at $z \simeq 0$. We know that the population of quenched galaxies – not forming new stars any longer – increases with their stellar mass for a given kind of environment, and with environment size (from small galaxy groups to large clusters) at fixed stellar mass (see Fig. 3, left; Weinmann *et al.*, 2006; van den Bosch *et al.*, 2008; Pasquali *et al.*, 2009; Wetzel *et al.*, 2012).

Galaxies become increasingly older (in terms of the mean age of their stars) as their environment becomes more massive (from galaxy groups to clusters), suggesting that galaxies in today's clusters were accreted at earlier times (i.e. at a higher redshift of infall) than galaxies in today's groups and had their star formation activity suppressed for longer times (Pasquali *et al.*, 2010). Most likely the quenching of their star formation activity happened while these galaxies were still living in smaller groups, which merged at later times with bigger structures like galaxy clusters.

Unfortunately, observations of $z \simeq 0$ galaxies can not constrain their redshifts of infall, or the time when they were subjected to environmental effects for the first time. Both *Euclid* and *Chronos* will trace the assembly history of environments with cosmic time and provide us with a direct measurement of the redshift of infall of galaxies as a function of their stellar mass. In addition, the lensing information from *Euclid* will be combined with the spectroscopic information produced by *Chronos* to probe the dependence of the star formation histories on the dark matter halos. However, while *Euclid* will only trace the assembly of the very massive end, with a significant bias towards star-forming galaxies, *Chronos* will extend the study to smaller masses, including old populations, and therefore, avoiding

the selection bias of the *Euclid* sample. A comparison of the $z \approx 0$ data with the predictions of semi-analytic models of galaxy evolution indicates that galaxies should quench their star formation over a few billion years; this is the only available and indirect estimate of the timescale for environmental quenching of star formation and it is unclear how much it depends on environment and whether it has changed with redshift. When and in which environments did the quenching of the star formation activity of galaxies start? How fast did it proceed? The quantitative and direct answers to these questions come from the measurements of star formation rates, star formation histories and chemical enrichment of galaxies of different stellar mass, in different environments at different epochs, from $z \sim 1-3$ to $z=0$. Only these observables provide a direct estimate of the typical timescales of star formation in galaxies and hence a model-independent estimate of the timescales with which different environments succeeded in quenching their member galaxies of different stellar mass, and gave rise to the present-day galaxy populations.

With increasing redshift such measurements move to infrared wavelengths and become challenging even for modern ground-based telescopes. The intervening Earth atmosphere offers us only a partial disclosure of galaxies properties at $z > 0.5$; we can mostly measure emission lines (hence star formation rates), while absorption lines (age and metallicity indicators) become less and less accessible. From the data collected so far on galaxies at $0.3 < z < 0.8$, we know that the fraction of quenched galaxies is larger in galaxy groups than in the field, but definitively lower than the fraction of quenched galaxies in groups at $z \approx 0$ (see Fig. 3, right; Wilman *et al.*, 2005; McGee *et al.*, 2011). At intermediate redshifts, the fraction of star forming galaxies decreases from 70-100% in the field to 20-10% in the more massive galaxy clusters (Poggianti *et al.*, 2006). Nevertheless, in terms of their star formation rates, galaxies in groups are not significantly different from those in the field; only star forming galaxies in clusters exhibit star formation rates a factor of 2 lower than in the field at fixed stellar mass (Poggianti *et al.*, 2006; Vulcani *et al.*, 2010; McGee *et al.*, 2011).

In the highest redshift range probed for environment at present, $0.8 < z < 1$, the more massive galaxy groups and clusters are populated mostly by quenched galaxies along with a 30% fraction in post-starburst galaxies (i.e. with a recently truncated star formation activity; Balogh *et al.*, 2011). The fraction of post-starburst galaxies is a factor of 3 times higher in clusters than in the field. Cluster and field galaxies are instead very similar in terms of the strength of their star formation activity and the amount by which their star formation has been quenched. These results have led Muzzin *et al.* (2012) to conjecture that either the quenching of star formation due to the secular evolution of galaxies dominates over the quenching induced by galaxy environment, or both mechanisms occur together with the same

timescale. Which timescale? We do not currently know. In order to make further progress, we require a facility such as *Chronos* to observe a complete stellar-mass limited sample of environments at $z \gtrsim 1$, and to measure the star formation histories of their galaxies with an unprecedented accuracy, thus providing the fading timescales of star formation of galaxies of different stellar mass inhabiting different environments. This is not simply an incremental step in our knowledge of environment-driven galaxy evolution. This is the *fundamental quantitative change* from the simple head-count of quenched or star forming galaxies to the measurement of physical properties of galaxies in environments at $z \gtrsim 1$, during the peak of galaxy formation activity. Such a step makes it possible to compare for the first time the same physical properties of galaxies at fixed stellar mass in different environments between $z=0$ and $z \gtrsim 1$, and to firmly quantify the extent to which environment regulates and modifies galaxy evolution across cosmic time.

II.4 Revealing the stellar population content of $z \sim 1-3$ galaxies

Understanding the nebular and stellar population properties of high redshift galaxies is an essential step towards a self-consistent picture of galaxy formation and evolution. The study of strong emission lines in the spectra of galaxies at $z \gtrsim 1$ has recently led to important results on the gas-phase properties, like the fact that metallicity exhibits a sharp transition towards subsolar values at $z \gtrsim 2.5$ (e.g., Möller *et al.*, 2013), and gas-rich disks are more dispersion dominated than in the nearby Universe (e.g., Förster-Schreiber *et al.*, 2011). Both results point to a major role of the accretion of unprocessed gas during the assembly of galaxies. In stark contrast, little is known about the stellar population content (i.e. the overall star formation history, metallicity, and IMF) of galaxies at $z > 1$.

Extracting star formation histories from spectra: An example

Fig. 4 (left) illustrates the reason for this impasse, that will render problematic our understanding of galaxy formation and evolution in the coming two decades. The upper panel plots a synthetic model spectrum, resembling the progenitor of a nearby early-type galaxy with mass $3 \times 10^{10} M_{\odot}$, as seen at $z=2$ (grey spectrum). The bottom panel shows the typical emission spectrum from the night sky (blue) as well as telluric absorption (grey). Such a high background – several orders of magnitude higher than the signal – makes the spectrum intrinsically inaccessible from a ground-based observatory, regardless of its photon collecting power (see also Fig. 2). The synthetic spectrum of Fig. 4 is obtained as a linear combination of simple stellar population models from the MILES synthetic library (Vazdekis *et al.*, 2010), assuming that the galaxy starts forming stars at $z \approx 5$ (cor-

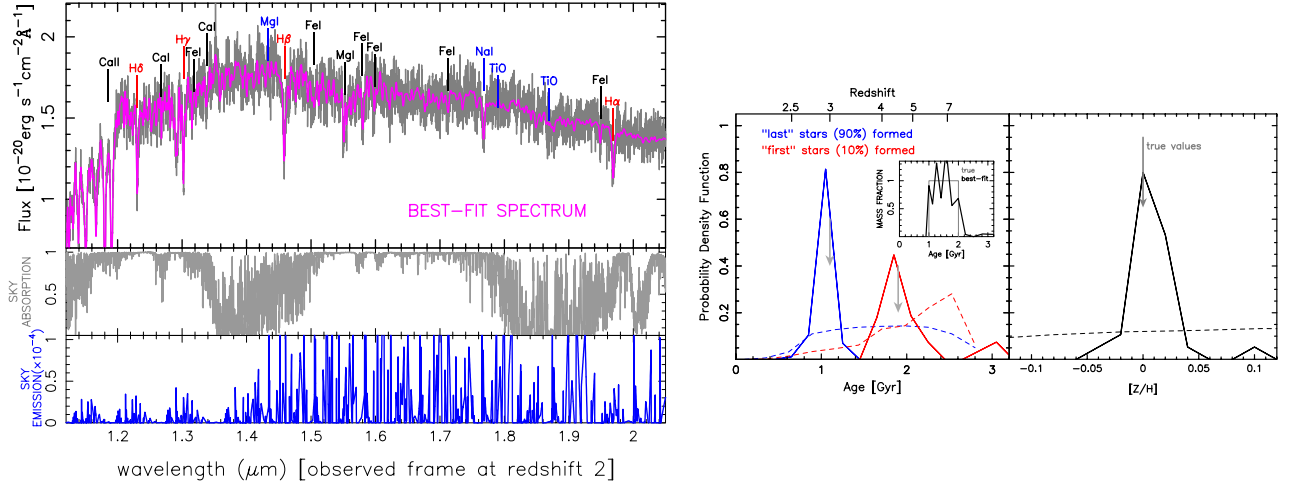


Figure 4: **Left-Top:** Synthetic spectrum resembling the progenitor of a nearby early-type galaxy, with a stellar mass of $3 \times 10^{10} M_{\odot}$, at $z \sim 2$ with a SNR of 20 per resolution element (grey). The best-fit model – obtained by direct spectral fitting – is overplotted in magenta. Some spectral features, sensitive to the star formation history (red), metallicity/abundances ratios (black), and IMF (blue) are included. **Left-Bottom:** sky emission (blue) and telluric absorption (grey), make the target spectrum very challenging from the ground. **Right:** Constraints on the timescales of star formation (left) and metallicity (right), derived from the spectrum on the left. The vertical grey arrows mark the input values. On the left, marginalized probability distribution functions (PDF) of the first (blue) and last (red) time of star formation (corresponding to 10% and 90% of total stellar stars formed). The inset shows the “true” star formation history of the system (grey histogram), and one typical best-fitting estimate (black curve). Note that a deep photometric survey cannot constrain the SFH or the metallicity in detail (dashed lines in both panels)

responding to an age of ~ 2 Gyr at $z = 2$, at a constant rate, with solar metallicity and a Kroupa-like IMF, down to $z \sim 3$ (Age ~ 1 Gyr), when star formation is suddenly quenched (because of, e.g., internal and/or environmental processes). The SNR of the spectrum (~ 20 , per resolution bin at $R=1500$) corresponds to a deep exposure, as planned for the *Chronos* ultra-deep survey (see Sec. V). Notice the strong Balmer lines in the spectrum (e.g. $H\beta$ and $H\delta$), that reflect the recent (~ 1 Gyr) quenching of star formation – a fact only recently observed at $z \sim 1.5$ (Bezanson *et al.*, 2013; Ferreras *et al.*, 2013), and eventually attributable to the presence of an AGN through suitable diagnostic lines that, at $z \gtrsim 1$, are hard to observe from the ground (e.g. $H\alpha$ and the companion [NII] line). A survey like *Chronos* is therefore required for the redshift range corresponding to the peak of galaxy formation activity.

Several absorption lines can be measured in the spectrum, most being sensitive to total metallicity and to the chemical abundances of individual elements (e.g. Mg, Si, Ti, Ca, Na), and some of them also to the fraction of dwarf-to-giant stars in the stellar IMF (e.g. the TiO features, see blue hatched regions). While the measurement of (total) metallicity and, to a lesser degree, that of $[\alpha/\text{Fe}]$ abundance ratio are a common practice in the case of low-redshift galaxies (e.g., Gallazzi *et al.*, 2005; Thomas, Maraston, Johansson, 2011), abundance estimates of single chemical species at low- z has become feasible only recently (e.g., Johansson, Thomas, Maraston, 2012; Conroy *et al.*, 2013), thanks to the rapid development of stellar population models and

spectral fitting techniques. During the next two decades we will develop superb stellar population models and software tools, to constrain also the star formation histories, abundance patterns, and the stellar IMF in high redshift galaxies, provided that their spectra will become accessible. Estimating metallicity and abundance ratios for galaxies at $z > 1$ would therefore give us crucial insights into the chemical enrichment history of galaxies (infall versus outflows of cold gas, and preferential loss of metals from supernovae-driven winds), nucleosynthesis yields, and the time-delay distributions of different types of SNe (e.g. Type Ia relative to core-collapse, driving the $[\alpha/\text{Fe}]$ of a stellar population at different epochs), as absorption features at high redshift would reflect the abundance of different elements by the time and immediately after they are synthesized in a galaxy.

During the next decade, direct fitting of stellar population models to data will likely become the “standard” tool to optimally extract the spectral information, compared to other well consolidated approaches like the analysis of line strengths (e.g. Lick-system indices). The magenta curve in the top panel of Fig. 4 (left) shows the result of fitting the ultra-deep-survey-like synthetic spectrum with a linear combination of 50 simple stellar populations, with different ages and metallicities. Notice that internal reddening – certainly important at high redshift – is set to be zero when synthesizing the spectrum, while it is included as a free parameter in the fitting procedure. The superb quality of the fit will be typical for data of the quality we envisage

for *Chronos*, with flux calibration accuracy better than a few percent.

Fig. 4 (right) illustrates the possibility of constraining timescales and metallicity for targets in the ultra-deep survey. One hundred noise realizations of the above mock spectrum are fitted as shown in Fig. 4, deriving from each best-fitting mixture some relevant, illustrative, parameters, i.e. the first and last epochs of star formation as well as the average metallicity. The first and last epochs are defined by the times when the system formed 10% and 90% of its stellar mass, respectively. Hence, NIR spectra with the given combination of SNR and resolution allows us to constrain sufficiently the formation timescales of a stellar population at redshifts corresponding to the peak of formation activity, as well as its total metal content (at less than 10%).

Velocity dispersion (not shown in the plot) can also be constrained with a $<10\%$ accuracy. Notice that the last epoch of star formation is connected to the quenching mechanism (e.g. AGN and/or environment), while the first epoch is ultimately driven by the initial conditions of density perturbations, along with other subtle physics (like reionization preventing star formation in small halos) illustrating the possibility to finally understand galaxy evolution in a full cosmological framework. In comparison, deep photometric surveys will not be able to compete on this front: the dashed lines in Fig. 4 (right) show the constraining power when only using broadband photometry, simulating data for the same galaxy, at $H_{AB}=25$, from a survey 1 mag deeper than the *Euclid* wide survey, and using a wide photometric coverage: $rizYJH + K$.

II.5 Towards the first galaxies

Chronos will be designed primarily in order to obtain a complete mass-limited sample of galaxies down to $\sim 10^{10} M_{\odot}$ over the $1 \lesssim z \lesssim 3$ redshift interval. In addition, this redshift range allows us to observe the rest-frame spectral window around the 4000\AA region, a highly sensitive area to the age distribution of stars and their chemical composition. Beyond this, though, the capabilities of the instrumentation also opens up the $3 \lesssim z \lesssim 6$ redshift range. Although at those redshifts it will not be possible to observe mass-limited samples, *Chronos* will obtain SFRs from the NUV emission, which – complemented with stellar mass estimates from additional photometry from *Euclid* and future ground-based NIR photometric surveys will give a snapshot of the evolution of the efficiency of star formation between the first phases of galaxy formation at $z \gtrsim 6$ (the topic of the next section), and the epoch at the peak of activity (this section), acting as a bridge between these two fundamental stages of cosmic evolution. In addition, rest-frame NUV spectral features such as Mg_{UV} (Daddi *et al.*, 2005) will help characterize the properties of the stellar populations, although with a significantly lower precision with respect to the $z \lesssim 3$ sample.

II.6 Population Synthesis in 2030

The most common methodology for deriving relevant stellar population parameters from the integrated light of galaxies consists in confronting observational data to predictions from stellar population synthesis models (Tinsley, 1980). However this approach is hampered by various fundamental degeneracies such as that between the age and the metallicity (Worthey, 1994). There are also other limitations that entangle derivations of burst-age and burst-strength (Leonardi & Rose, 1996) or effects from the IMF (Vazdekis *et al.*, 2010). Such degeneracies are commonly tackled with targeted spectral indices, direct spectral fitting, or a combination of both. However, as the quality of these models rely on the employed ingredients, great efforts are being put to develop stellar models and spectral libraries. This goal is being achieved in part by means of new stellar evolutionary calculations, with updated input physics, which might eventually include Helium and atomic diffusion, and higher mass/metallicity/age resolution (Pietrinferni *et al.*, 2004). Stellar libraries at moderately high spectral resolution with varying abundance ratios, either theoretical (Coelho *et al.*, 2005) or empirical (Milone *et al.*, 2011) as well as theoretical stellar evolutionary tracks with varying element mixtures (Pietrinferni *et al.*, 2006) are being developed. These libraries will lead to new generations of stellar population synthesis models that are better suited to estimate the observed abundance patterns, including the measurement of *individual* abundance ratios, opening the field of extragalactic archaeology. This information puts the mass-metallicity relation “under the microscope”, allowing us to quantify in detail the various aspects of galactic chemical enrichment, including the effect of infall and outflows, and its connection with environment (Kawata & Mulchaey, 2008); or the dispersion of metals into the intergalactic medium (Pontzen *et al.*, 2008). In addition, developments in stellar libraries in the NUV (e.g., Koleva & Vazdekis, 2012) will optimise the methodology to extract information from the $z \gtrsim 3$ sample (see II.5).

The advent of models predicting galaxy spectra at moderately high resolution (e.g., Bruzual & Charlot, 2003; Vazdekis *et al.*, 2010) has opened the possibility of establishing robust constraints on the star formation histories, via a variety of full spectrum-fitting methods (e.g., Koleva *et al.*, 2008). There is a growing body of publications based on this approach as it allows us not only to attempt to estimate the star formation history (see II.4) but also to interpret better the results based on line-strength indices, which are more biased towards recent bursts and therefore hide the contributions weighted by mass. These estimates are particularly relevant for assessing the various mechanisms proposed for the assembly of galaxies and the role of environment.

II.7 Synergy with *Herschel*

The synergies with *Euclid* are obvious, and throughout this white paper there are abundant references to the use of the *Euclid* surveys to aid in the target selection, analysis and interpretation of the data. We devote this subsection to another of ESA's flagship missions: *Herschel* has provided, for the first time, efficient imaging of large areas of the sky in the far-IR window, from 70 to 500 μm . In particular, the Herschel Multi-tiered Extra-galactic Survey (HerMES, Oliver *et al.*, 2010) – the largest project on *Herschel* at 900 hrs – mapped over 70 deg^2 , tracing dust-enshrouded star-formation sources during the peak of galaxy mass assembly. *Chronos* will be essential to take full advantage of

the legacy of the *Herschel* surveys providing unique spectroscopic follow-up. Indeed, a detection in the FIR implies large amounts of dust emission, i.e. reprocessed light from the UV stellar emission. As a result the optical/near-IR SEDs are often very red and a large fraction of luminous *Herschel* galaxies are very faint or undetected in the optical bands, requiring deep NIR spectroscopy for their redshift measurements. However, over 90% of the 250 μm -detected sources have a counterpart at $K_{\text{AB}} < 24$. The *Chronos* surveys are thus optimally designed to exploit in full the investment of ESA in HerMES, providing redshifts, dynamical masses, stellar population properties, local environment and clustering for essentially all of the sources detected in the HerMES survey.

III. COSMIC REIONIZATION & GALAXY/BLACK-HOLE FORMATION

III.1 Introduction

Cosmic reionization is a landmark event in the history of the Universe. It marks the end of the “Dark Ages”, when the first stars and galaxies formed, and when the intergalactic gas was heated to tens of thousands of degrees Kelvin from much lower temperatures. This global transition, during the first billion years of cosmic history, had far-reaching effects on the formation of early cosmological structures and left deep impressions on subsequent galaxy and star formation, some of which persist to the present day.

The study of this epoch is thus a key frontier in completing our understanding of cosmic history, and is currently at the forefront of astrophysical research (e.g. Robertson *et al.*, 2013). Nevertheless, despite the considerable recent progress in both observations and theory (e.g. see recent reviews by Dunlop 2013 and Loeb 2013) all that is really established about this crucial era is that Hydrogen reionization was completed by redshift $z \simeq 6$ (as evidenced by high-redshift quasar spectra; Fan *et al.* 2006) and probably commenced around $z \sim 15$ (as suggested by the latest WMAP9 microwave polarisation measurements, which favour a ‘mean’ redshift of reionization of 10.3 ± 1.1 ; Hinshaw *et al.* 2013). However, within these bounds the reionization history is essentially unknown, and new data are required to construct a consistent picture of reionization and early galaxy formation/growth.

Unsurprisingly, therefore, understanding reionization is one of the key science goals for a number of current and near-future large observational projects. In particular, it is a key science driver for the new generation of major low-frequency radio projects (e.g. LOFAR, MWA and SKA) which aim to map out the cosmic evolution of the *neutral atomic* Hydrogen via 21-cm emission and absorption. However, such radio surveys cannot tell us about the sources

of the ionizing flux, and in any case radio observations at these high redshifts are overwhelmingly difficult, due to the faintness of the emission and the very strong foregrounds. It is thus essential that radio surveys of the neutral gas are complemented by near-infrared surveys which can both map out the growth of ionized gas, and provide a complete census of the ionizing sources. A genuinely multi-wavelength approach is required, and cross-correlations between different types of observations will be necessary both to ascertain that the detected signals are genuine signatures of reionization, and to obtain a more complete understanding of the reionization process.

It has thus become increasingly clear that a *wide-area, sensitive, spectroscopic* near-infrared survey of the $z=6-12$ Universe is required to obtain a proper understanding of the reionization process and early galaxy and black-hole formation. Such a survey cannot be undertaken from the ground, nor with *JWST* (inadequate field-of-view), or *Euclid* (inadequate spectroscopic sensitivity). Only a mission such as *Chronos* can undertake such a survey and simultaneously address the three, key, interlated science goals which we summarize below.

... in a nutshell

- Charting the progress of reionization through the clustering of Ly- α galaxies.
- Determining the source of reionization.
- Studying the emergence of the first galaxies over cosmologically representative volumes.

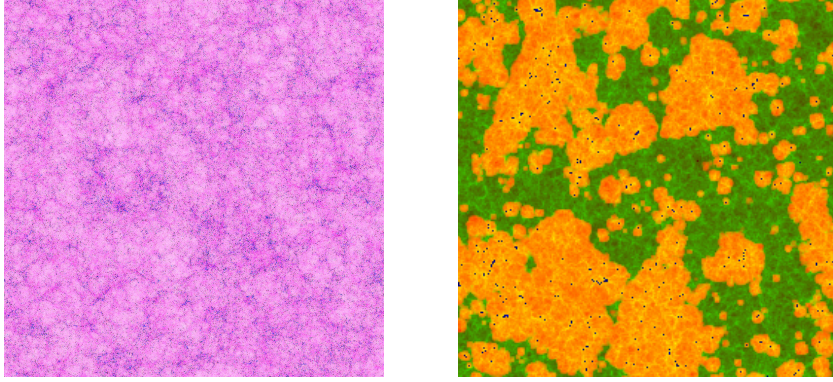


Figure 5: **Left** Early structure formation in Λ CDM (at $z = 6$) from an N -body simulation with 5488^3 (165 billion) particles and a volume $425 h^{-1} \text{Mpc}^3$. Shown are the dark-matter density (pink) and halos (blue). This synthetic image corresponds to 3.5×3.5 degrees on the sky. **Right:** The geometry of the epoch of reionization, as illustrated by a slice through a $(165 \text{Mpc})^3$ simulation volume at $z = 9$. Shown are the density (green/yellow), ionized fraction (red/orange), and ionizing sources (dark dots) (Iliev *et al.*, 2012). The necessity of a deep, near-infrared spectroscopic survey covering many square degrees is clear.

III.2 The clustering of Lyman- α emitters as a probe of reionization

Cosmological simulations of the reionization process predict that the highly-clustered, high-redshift sources of Lyman-continuum photons will lead to an inhomogeneous distribution of ionized regions; the reionization process is expected to proceed inside-out, starting from the high-density peaks where the galaxies form. Thus, as demonstrated by the state-of-the-art simulations shown in Fig. 5, reionization is predicted to be highly patchy in nature. This prediction is already gaining observational support from the latest large-area surveys for Lyman- α emitters at $z \simeq 6.5$, where it has been found that, depending on luminosity, their number density varies by a factor of 2 – 10 between different $\simeq 1/4 \text{deg}^2$ fields (Ouchi *et al.*, 2010; Nakamura *et al.*, 2011). It is thus clear that surveys over many square degrees are required to gain a representative view of the Universe at $z > 6$. Crucially, with such a survey, the differential evolution and clustering of Lyman-break galaxies and Lyman- α emitting galaxies can be properly measured for the first time, offering a key signature of the reionization process.

As has been well-demonstrated over recent years, galaxies at high-redshift can be very effectively selected on the basis of *either* their redshifted Lyman break (the sudden drop in emission from an otherwise blue galaxy, due to inter-galactic absorption at wavelengths $\lambda_{\text{rest}} < 1216 \text{\AA}$), or their redshifted Lyman- α emission. The former class of objects are termed Lyman-Break Galaxies (LBGs) while the latter are termed Lyman- α Emitters (LAEs). In principle, LAEs are simply the subset of those LBGs which display detectable Lyman- α emission, but the current sensitivity limitations of broad-band near-infrared imaging over large areas has meant that narrow-band imaging has been successfully used to yield samples of lower-mass galaxies which are not usually identified as LBGs (e.g., Ono *et*

al., 2010). Nevertheless, as demonstrated by spectroscopic follow-up of complete samples of bright LBGs (e.g., Stark *et al.*, 2010; Vanzella *et al.*, 2011; Schenker *et al.*, 2012), the fraction of LBGs which are LAEs as a function of redshift, mass, and environment is a potentially very powerful diagnostic of both the nature of the first galaxies, and the physical process of reionization.

With the unique combination of deep, wide-area near-infrared imaging, provided by surveys such as *Euclid*, and deep, complete follow-up near-infrared spectroscopy, made possible with *Chronos*, we now propose to fully exploit the enormous potential of this approach. The essential idea of using *Chronos* to constrain reionization is as follows: while the Lyman- α luminosity of LAEs is affected both by the intrinsic galaxy properties, *and* by the H I content (and hence reionization), the luminosity of LBGs (which is measured in the continuum) depends only on the intrinsic galaxy properties. Thus, a deep, wide-area, complete survey for LBGs at $z \simeq 6\text{--}12$ with accurate redshifts secured by *Chronos* will deliver a definitive measurement of the evolving luminosity function and clustering of the emerging young galaxy population, while the analysis of the follow-up spectroscopy will enable us to determine which LBGs reside in sufficiently large ionized bubbles for them to also be observed as LAEs. In order to prevent strong damping wing absorption of Ly α photons, a galaxy must carve out a bubble of radius R_I corresponding to a redshift difference with respect to the source of $\Delta z > 0.01$, or around 250 physical kpc at $z \approx 8$. According to the most recent reionization history predictions from cosmological simulations, consistent with the various reionization constraints, the H I fraction at this redshift is around $\chi \approx 0.4 - 0.7$. It is easy to show that R_I for a typical galaxy with a star-formation rate of $\dot{M}_* = 1 M_\odot \text{yr}^{-1}$ is of the same order or smaller, depending on poorly established values of the ionizing photon escape fraction. Thus, such galaxies will be only marginally detectable in the Ly α line if they are

isolated. In practice, some of these galaxies will be highly clustered and therefore will help each other in building a H II region which is large enough to clear the surrounding H I and make it transparent to Ly α photons.

This argument emphasizes the importance of clustering studies of LAEs, for which *Chronos* is optimally designed. A key aim is to compute in great detail the two-point correlation function of LAEs and its redshift evolution. For the reasons outlined above, reionization is expected to increase the measured clustering of emitters and the angular features of the enhancement would be essentially impossible to attribute to anything other than reionization. In fact, under some scenarios, the apparent clustering of LAEs can be well in excess of the intrinsic clustering of halos in the concordance cosmology. Observing such enhanced clustering would confirm the prediction that the H II regions during reionization are large (McQuinn *et al.*, 2007).

As required to meet our primary science goals, the *Chronos* surveys will result in by far the largest and most representative catalogues of LBGs and LAEs ever assembled at $z > 6$. Detailed predictions for the number of LBGs as extrapolated from existing ground-based and *HST* imaging surveys are deferred to the next subsection. However, here we note that the line sensitivity of the 100 deg² spectroscopic survey will enable the identification of LAEs with a Ly α luminosity $\geq 10^{42.2}$ erg s⁻¹, while over the smaller, ultra-deep 10 deg² survey this line-luminosity limit will extend to $\geq 10^{41.6}$ erg s⁻¹. Crucially this will extend the Lyman- α detectability of LBG galaxies at $z \simeq 8$, with brightness $J \simeq 27$ (AB mag), down to “typical” equivalent widths of $\simeq 15 \text{ \AA}$ (Stark *et al.*, 2010; Vanzella *et al.*, 2011; Curtis-Lake *et al.*, 2012; Schenker *et al.*, 2012).

The total number of LAEs in the combined *Chronos* surveys will obviously depend on some of the key unknowns that *Chronos* is designed to measure, in particular the fraction of LBGs which display detectable Ly α emission as a function of redshift, mass and environment. However, if the observed LAE fraction of bright LBGs at $z \simeq 7$ is taken as a guide, the *Chronos* surveys will uncover $\sim 10,000$ LAEs at $z > 6.5$.

III.3 The emerging galaxy population at $z > 7$, and the supply of reionizing photons

Chronos will provide a detailed spectroscopic characterization of an unprecedentedly large sample of LBGs and LAEs. Crucially, as well as being assembled over representative cosmological volumes of the Universe at $z \simeq 6-12$, these samples will provide excellent sampling of the brighter end of the galaxy UV luminosity function at early epochs. As demonstrated by the most recent work on the galaxy luminosity function at $z \simeq 7-9$ (McLure *et al.*, 2013), an accurate determination of the faint-end slope of the luminosity function (crucial for understanding reionization) is in fact currently limited by uncertainty in L^* and ϕ^* .

Consequently, a large, robust, spectroscopically-confirmed sample of brighter LBGs over this crucial epoch is required to yield definitive measurements of the evolving luminosity functions of LBGs and LAEs.

Leaving aside the uncertainties in the numbers of LAEs discussed above, we can establish a reasonable expectation of the number of photometrically-selected LBGs which will be available for *Chronos* spectroscopic follow-up by the time of the mission. For example, scaling from existing *HST* and ground-based studies, the ‘Deep’ component of the *Euclid* survey (reaching $J \simeq 26$, 5- σ over $\simeq 40 \text{ deg}^2$), is expected to yield $\simeq 6000$ LBGs in the redshift range $6.5 < z < 7.5$ with $J < 26$ (selected as “Z-drops”), $\simeq 1200$ at $7.5 < z < 8.5$ (“Y-drops”), and several hundred at $z > 8.5$ (“J-drops”) (Bouwens *et al.*, 2010; Bowler *et al.*, 2012; McLure *et al.*, 2013).

Therefore, the planned spectroscopic follow-up over 10 deg^2 , will be able to target (at least) $\simeq 1500$ LBGs in the redshift range $6.5 < z < 7.5$, $\simeq 300$ in the redshift bin $7.5 < z < 8.5$, and an as yet to be determined number of candidate LBGs at $8.5 < z < 9.5$. The proposed depth and density of the *Chronos* near-infrared spectroscopy will allow detection of Ly α line emission from these galaxies down to a 5- σ flux limit $1 \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$, enabling rejection of any low-redshift interlopers, determination of the LAE fraction down to EWs of $\simeq 10 \text{ \AA}$, and accurate spectroscopic redshifts for the LAE subset.

III.4 The contribution of AGN to reionization & the early growth of black holes

SDSS has revolutionised studies of quasars at the highest redshifts, and provided the first evidence that the epoch of reionization was coming to an end around $z \gtrsim 6$ (Becker *et al.*, 2001). As with the studies of galaxies discussed above, pushing to higher redshifts is impossible with optical surveys, regardless of depth, due to the fact that the Gunn-Peterson trough occupies all optical bands at $z > 6.5$. Therefore, to push these studies further in redshift needs deep wide-field surveys in the near-infrared.

The wide-area, ground-based VISTA near-infrared public surveys such as VIKING and the VISTA hemisphere survey are slowly beginning to uncover a few bright quasars at $z \simeq 7$ (e.g., Mortlock *et al.*, 2011), and it is to be expected that *Euclid* will be able to provide a good determination of the very bright end of the QSO luminosity function at $z > 6$. However, the shape of the QSO luminosity function at these redshifts can only be studied with detailed near-infrared spectroscopy over a significant survey area. This is the only direct way to properly determine the contribution of accreting black holes to the reionization of the Universe and constrain the density of black-holes within the first Gyr after the Big Bang; *Chronos*’s combination of depth and area provides the ideal way in which to measure the evolving luminosity function of quasars at $6.5 < z < 10$.

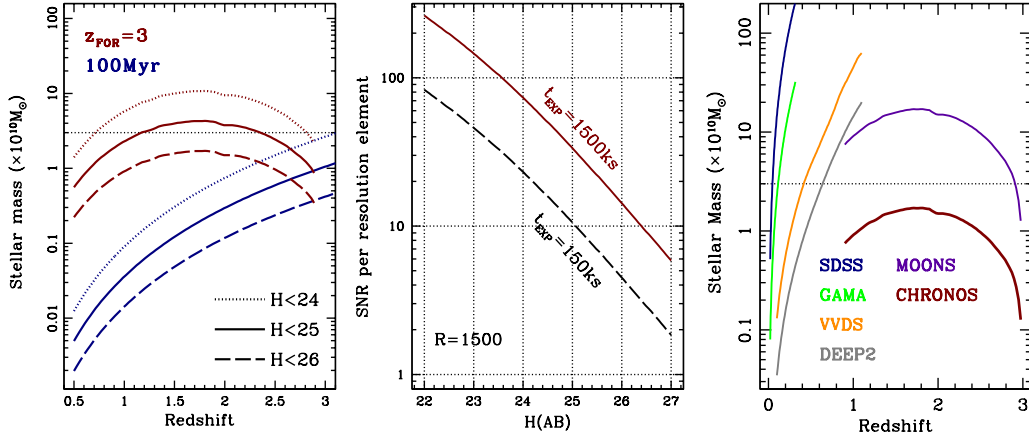


Figure 6: **Left:** Stellar mass versus redshift for three limiting-magnitude surveys, as labelled. The red lines correspond to old stellar populations, whereas the blue lines trace a more luminous, young population, corresponding to the typical age of a star forming galaxy. **Middle:** SNR per resolution element at $R \equiv \lambda/\Delta\lambda = 1500$ for a 2.5m (diameter) collecting area, 20% total efficiency, a zodiacal light (minimal) level of $H_{AB}=25$ and typical readout noise and dark current of cutting-edge NIR arrays ($RN=4e$, 0.01 e/s, respectively). We note that the SNR is estimated in the continuum, a much more stringent requirement than emission line estimates, typical of cosmology-orientated surveys. **Right:** Comparison of *Chronos* (wide survey) with a range of optical and NIR spectroscopic surveys with similar spectral resolution. The curves track the lower limit in stellar mass for an old population (i.e. the worst case scenario).

IV. ADDITIONAL SCIENCE CASES

Although this white paper deals with the core science driver of galaxy formation and evolution in the spectral window $1 \lesssim z \lesssim 12$, the legacy side of *Chronos* is immense, and additional science projects can be addressed. Among them, we list a few relevant cases below:

- i) Transients: the planned 5-year mission can accommodate the spectroscopic follow up of transients, most notably high redshift supernovae, allowing us not only to confirm the type of supernova, but spectroscopic features could be used to understand their properties and evolution, relevant to precision cosmology studies (Foley & Kasen, 2011).
- ii) Cosmology: *Chronos* will enable cosmological model testing beyond *Euclid*. As an example, measurements

of the velocity field, galaxy bias, and lensing potential simultaneously will enable a measurement of general single scalar-field models (e.g., Amendola *et al.*, 2012). The deep redshift range would also constrain early-dark energy models, complementing the *Euclid* cosmology objectives using techniques such as those used by Mandelbaum *et al.* (2012) in SDSS.

- iii) Brown dwarves: The *Chronos* survey could include a programme to explore cool T-dwarves out to a few hundred pc, allowing us to determine the scale height of this population. By targeting nearby star forming regions, we can probe the IMF down to Jupiter-size masses.

V. SURVEY REQUIREMENTS

The core requirement of the survey is the apparent magnitude limit to obtain a complete sample selected in stellar mass, with acceptable SNR per resolution element for the study of the underlying stellar populations. Any other criteria commonly approached in surveys (e.g. selection in luminosity or colour) will bias the sample. We use stellar population synthesis models (Bruzual & Charlot, 2003) to estimate the apparent magnitude with respect to

stellar mass and redshift for two extreme scenarios (Fig. 6, left panel). An old population (red lines) will lack the most massive and luminous stars, therefore appearing significantly fainter than a younger population with the same mass (blue lines). The age range used in the figure covers a conservative interval as obtained, e.g., from spectroscopic observations of the local Universe (Gallazzi *et al.*, 2005). The figure shows that at $H_{AB} \sim 26$, we will

obtain a complete sample out to $z \sim 3$ for galaxies with stellar mass around $1 - 2 \times 10^{10} M_{\odot}$. This is a *conservative* estimate, as observations suggest that there is a strong trend towards younger ages in low mass galaxies, with the characteristic star-forming galaxy appearing more massive at higher redshift (Pérez-González *et al.*, 2008). This trend implies that the majority of low-mass galaxies will be younger, allowing us to reach a completeness level – if we relax the constraint regarding old populations – at lower stellar masses, possibly around $10^9 M_{\odot}$, covering an unprecedented range of galaxy mass over the $z=1-3$ redshift window. The middle panel of Fig. 6 gives an estimate of the SNR per resolution element (at $R=1500$) achieved for two exposure times, as labelled. We envisage a two-tiered survey, comprising a wide-deep survey covering 100 deg^2 (corresponding to the dashed black line) and an ultra-deep survey with a 10 times longer integration over 10 deg^2 (solid red line). We emphasize here that a proper characterization of the stellar populations from spectroscopic data requires $\text{SNR} \gtrsim 5 - 10$. This figure assumes a low zodiacal and thermal background level, with a spatial resolution of 0.3 arcsec , and typical detector noise for the type of available arrays (e.g., Teledyne Hawaii 4RG). The rightmost panel of Fig. 6 compares the ability of *Chronos* to obtain a mass-limited sample out to a chosen redshift with recent or planned spectroscopic surveys at similar resolution. We note that neither *Euclid* nor *WFIRST* estimates are included in the figure, as their low-resolution, slitless grism spectra are not capable of achieving the goals of this white paper². *MOONS* is clearly the best option at present in the $z=1-3$ redshift range, however, the signal in the continuum will be weak unless very young populations are considered. Only *Chronos* can provide the collecting power and wide field of view to tackle in an unbiased way the analysis of galaxies at the peak of activity.

As regards to the required total areal coverage on the sky, we use as reference the SDSS, whose high quality spectroscopic data can be extended out to, at most, $z \lesssim 0.2$, covering a comoving volume of $5.5 \times 10^{-5} \text{ Gpc}^3$ per square degree. Over the proposed $z \sim 1-3$ range, we have 0.02 Gpc^3 per deg^2 . Hence, in order to probe the environment in detail comparable to the $\sim 10^4 \text{ deg}^2$ of SDSS/DR7 (Abazajian *et al.*, 2009), we need around 30 deg^2 . In addition, we expect environment to evolve significantly between the SDSS baseline and the goal of *Chronos*. We use a large cosmological simulation (Millennium, Springel *et al.*, 2005) to find an evolution in the number of a factor of $\sim 2-3$ for groups with halo mass $M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$ at $z=0$. Therefore, the general survey should target around 100 deg^2 , putting this project outside of the reach of *JWST* or any of the extremely large telescopes on the ground. SDSS has also shown that datasets comprising ~ 1 million spectra are necessary to split the sample with respect to the many properties under consideration (velocity dispersion, luminosity, mass, environment, etc). Finally, an extrapolation of the Muzzin *et al.* (2013) data using a fit to a Schechter law gives a number density of 1.2×10^5 galaxies per square degree at the $H_{\text{AB}}=26$ level in the $z \sim 1-3$ range.

... in a nutshell

- At $H_{\text{AB}}=26$ we expect $\gtrsim 100,000$ galaxies per square degree at $z \sim 1-3$
- Completeness down to a stellar mass $\sim 10^{10} M_{\odot}$ ($z \lesssim 3$) for any population.
- Two surveys: 100 deg^2 and 10 deg^2 extending over the $z \gtrsim 1$ environments probed by SDSS at $z \lesssim 0.1$.
- The final dataset will comprise $\sim 1 - 2$ million high-quality spectra.

VI. STRAWMAN MISSION CONCEPT

VI.1 Mission Profile

As an infrared survey mission, the preferred orbit for *Chronos* is at the low background L2 point, with heritage from *Herschel* and *Planck* operations and, in the future, from *Gaia*, *JWST* and *Euclid*. An Ariane 5 ECA or ME launcher provides excellent payload margin with a limit to L2 in excess of 6.2 tonnes. The standard fairing has a length of 12.7 m and a diameter of 4.6 m which can easily accommodate the proposed *Chronos* spacecraft configuration; these dimensions would allow a full 2.5m diameter $f/1.2$ telescope to be deployed without any dynamic mechanisms. Other launcher options could be considered, depending

on launch date. Once L2 has been reached, the Δv requirements for orbit station and formation keeping are small ($< 75 \text{ m s}^{-1} \text{ year}^{-1}$).

The *Chronos* survey strategy will follow that adopted for the *Euclid* deep-field programme, with frequent revisits to the same field centres to build up S/N on faint targets, ameliorate the contamination effects in crowded fields, and provide useful cadence for serendipitous studies of high-redshift supernovae and gamma-ray bursts. The regular layout of the proposed *Chronos* focal plane will allow a simple tiling strategy to cover contiguous areas of the 100 deg^2 (deep) and 10 deg^2 (ultra-deep) fields. For a 5 year mission,

²Nevertheless, as a reference, the *Euclid* Definition Study Report states that at $z \geq 1.5$ only galaxies with a stellar mass $> 4 \times 10^{11} M_{\odot}$ will provide useable spectra for the analysis of the populations (Laureijs *et al.*, 2011).

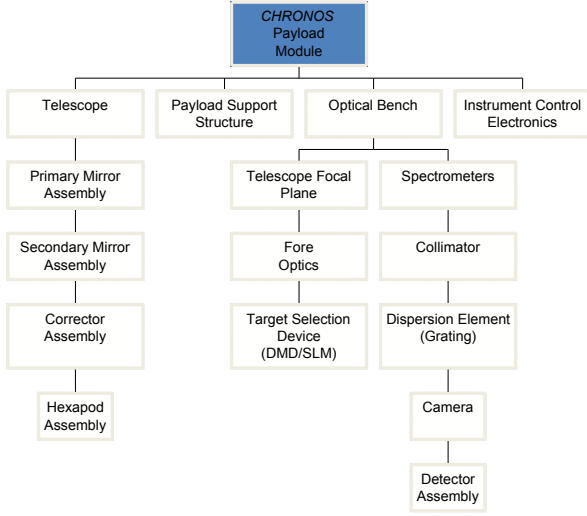


Figure 7: Subsystem breakdown of the payload module hardware.

at a 70% operational efficiency, the draft survey plan calls for visiting one pointing per week (deep) or one pointing per 10 weeks (ultra-deep with 10x longer exposure). Assuming that *Chronos* targets 30% of the available sample at a multiplex of ~ 5000 , will require ~ 8 mask configurations per pointing, giving a total exposure time of 150 ksec for the deep survey and 1500 ksec for the ultra-deep survey. The final galaxy samples would thus comprise ~ 1.5 million (~ 150 thousand) high quality spectra in the deep (ultra-deep) surveys.

As a dedicated survey mission, the ground segment can be kept relatively simple. Target definition for the survey will come from the optical-infrared imaging in the *Euclid* deep fields or, as a fallback, from *LSST* and *VISTA* ground-based deep survey ($AB \sim 27.5$ optical and $AB \sim 24.5$ infrared respectively). Fast data analysis will be required only for the transient detection programme. The downstream data rate will be approximately 50 GB/day after compression (depending on the number of intermediate detector samples are transmitted); assuming a typical K-band rate of transfer to the ground of 50 Mbit/s, all of the data can be transferred to the ground with a contact time of 3 hours per 24 hours.

VI.2 Payload Description

The science requirements for *Chronos* drive the choice of a telescope with a 2.5m aperture and a 1 deg field-of-view feeding eight identical multi-object slit-based spectrometers with moderate spectral resolution and good background subtraction. Selection of the science targets for the spectrometers can be achieved by using a digital micromirror device (DMD) or other form of spatial light modulator. There is no scientific need to make *Chronos* into a multi-purpose observatory so the spectrometers have a single

Item	Payload Requirements
Telescope	
Primary mirror	2.5 metre diameter
Field of View	1.0 degree diameter
Image quality	EE(80) < 0.3 arcsec
Spectrometers	
Target multiplex	~ 5000 objects per pointing
Field of view	0.2 deg ² (total, 8 spectrometers)
Spectral coverage	0.9 μ m to 1.8 μ m
Spectral resolution	R ~ 1500
Throughput	> 20% including detectors
Spacecraft	
Mass	< 4000 kg
Volume	
Diameter	3500 mm
Length	7000 mm

Table 1: Summary of the key performance requirements of the *Chronos* payload.

fixed resolving power (R ~ 1500). The DMD or spatial light modulator can be used to shut off the signal to the detectors, so no mechanisms are required. The payload module hardware breaks down into testable sub-systems as shown in Figure 7. The key performance parameters of the *Chronos* payload are shown in Table 1.

VI.2.1 Telescope Assembly

The 2.5-metre telescope could be a Korsch or Ritchey-Chrétien design for which a three-element field corrector which would give a field-of-view of 1 degree at $f/3$ to feed the eight spectrometers. Assuming a Ritchey-Chrétien design, this could be optimised to provide uniform image quality across the whole field with an image quality (EE80 ~ 0.3 arcsec) which is matched to the intrinsic size of galaxies at high redshift. Both the primary ($f/1.2$) and secondary ($f/3$) mirrors are hyperboloids. The secondary mirror (M2) is 0.9m in diameter and its mounting incorporates light sources for calibration of the spectrometer. While SiC will provide excellent performance if a 2.5m optical-quality mirror can be fabricated, lightweighted Zerodur would also be possible. The baseline design assumes that both M1 and M2 are fabricated from lightweighted Zerodur and are supported directly from the telescope support structure. A central light baffle incorporates the three fused silica corrector elements in the R-C design (two aspheric surfaces).

The telescope support structure is a SiC or CFRP space frame which supports all the hardware of the spacecraft, and under which is mounted the Instrument Optical Bench and the Instrument Service Module. The upper section of the telescope structure consists of a triangular frame, the corners of which act as structural nodes for the M1 backing structure, the M2 hexapod and the instrument optical

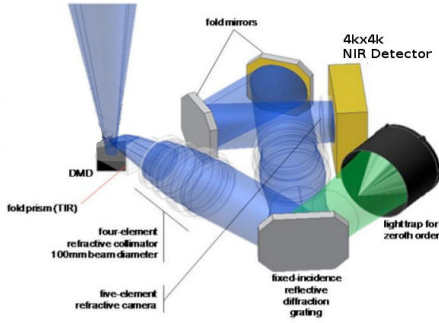


Figure 8: Close-up of a single spectrometer channel.

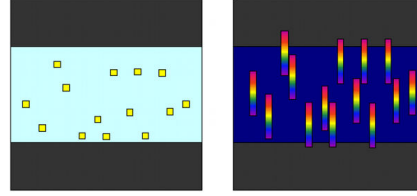


Figure 9: Schematic of multi object target selection using DMDs.

Phase A	30 Meuro
Phase B	200 Meuro
Phase C/D	600 Meuro
Launch	200 Meuro
Phase E/F	200 Meuro
Total	1230 Meuro

Table 2: *Chronos* lifecycle cost estimate, including national hardware contributions.

bench. The total length of the telescope is approximately 5 metres with a mass of 600 kg (excluding payload and service module).

VI.2.2 Spectrometer

At the heart of the *Chronos* concept are a set of eight multi-object spectrometers (Fig. 8), each capable of delivering complete samples of moderate resolution ($R \sim 1500$) near-infrared spectra for high redshift galaxies down to a magnitude limit of $H_{AB} \sim 26$ mag. To reach this faint limit requires 'multi-slit' spectroscopy with a target selection mechanism which is compatible with space operations. Our baseline approach is to use the Texas Instruments (TI) digital micro mirror devices (DMDs) which were originally proposed for the *SPACE* mission (Cimatti *et al.*, 2009). These are available in formats up to 2048×1080 pixels with a pitch of $13.68 \mu\text{m}$ and are currently at a technology readiness level of TRL ~ 4 (Zamkotsian *et al.*, 2010). Each of the individual micromirrors on the DMD can be switched into an 'ON' or 'OFF' position to define a *virtual* slit of 1.2×0.4 arcsec, centred on the target of interest, thus replicating the multislit masks used in ground-based spectroscopy of faint targets (Figure 9).

Simulations of the targeting efficiency of the DMD at the magnitude limit of the survey, indicate that each spectrometer can obtain the spectra for ~ 600 targets simultaneously, without spectral overlaps, giving a multiplex of ~ 4800 targets with eight spectrometers covering a total field of $\sim 0.2 \text{ deg}^2$.

To mitigate against qualification and availability of DMD devices in the timescale of an L2/L3 mission, a parallel technology development study should also be initiated early in the project to assess the technology readiness of other forms of target selection devices, including liquid crystal spatial light modulators and pupil beam steering devices.

The entrance apertures of the spectrometers will be positioned symmetrically within the telescope field of view to allow a simple step-and-stare operation to tile the sky contiguously. Each spectrograph will use refractive collimators and cameras, feeding a $4\text{k} \times 4\text{k}$ HgCdTe infrared array. A 60

lines/mm grating is used to produce a Nyquist sampled spectrum covering the range $0.9\text{--}1.8 \mu\text{m}$.

VI.3 Operational Model

The *Chronos* operational model follows the usual lines of a survey-type project. The satellite will operate autonomously except for defined ground contact periods during which housekeeping and science telemetry will be downlinked, and the commands needed to control spacecraft and payload will be uploaded. The data rate is around 50 GB/day which is easily handled with current data-processing systems. A data model for the mission will be developed in collaboration with ESA. Based on the data model, an archive system will be built, enabling data archiving, data processing and distribution of all *Chronos* observations with appropriate levels of processing, including all the necessary ancillary information.

VI.4 Programmatics and Cost

Chronos is envisaged as a typical science mission with ESA having overall control, but with a major contribution from a consortium of European institutes in the form of the science payload and ground segment. The mission has been designed to ensure that technologies with space heritage or high TRL are used where possible. The DMDs are an exception to this, although good progress has already been made in developing these for space application for the original *SPACE/Euclid* mission concepts. We believe that this technology can be further developed in good time for the *Chronos* mission and will thus not be a schedule driver. An approximate cost per phase is shown in Table 2.

... in a nutshell

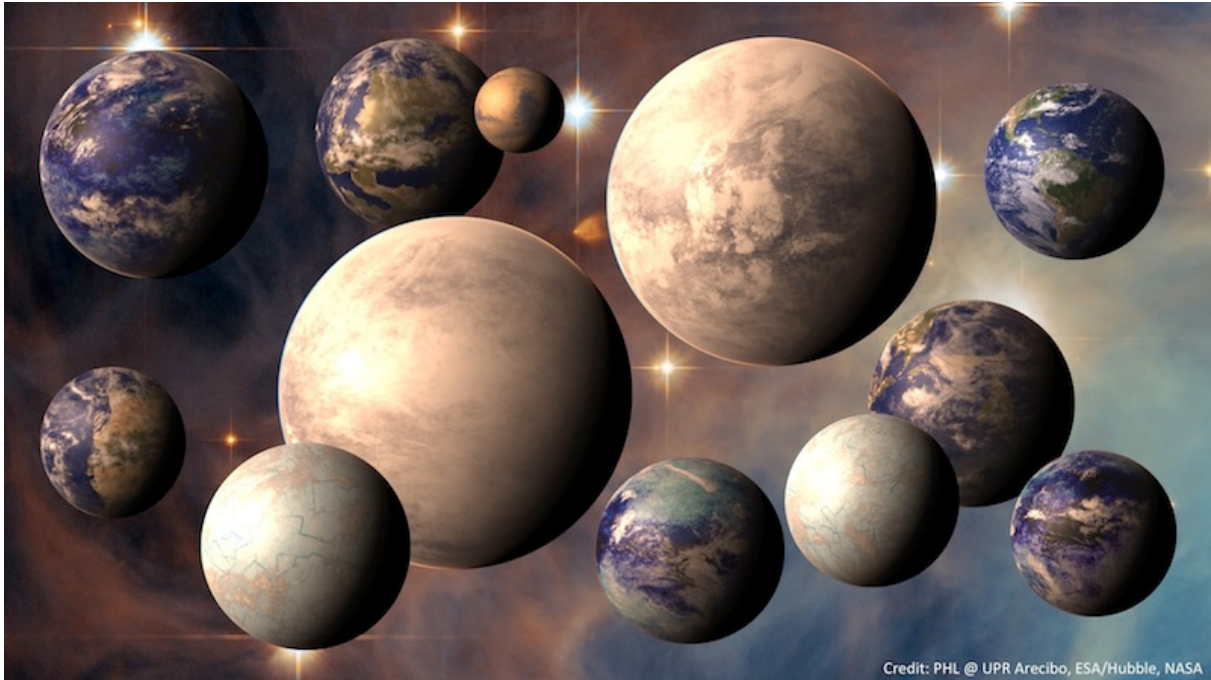
- 2.5m telescope, Korsch or Ritchey-Chrétien in SiC
- 8 spectrometers, $R=1500$, 4800 multiplex.
- 5 year mission at L2.
- Ariane 5 launcher.

REFERENCES

- Abazajian, K. N., *et al.*, 2009, *ApJS*, 182, 543
- Amendola, L., *et al.*, 2012, *Phys. Rev. D*, 85, 3008
- Balogh, M.L., *et al.*, 2011, *MNRAS*, 412, 2303
- Baldry, I. K., *et al.*, 2004, *ApJ*, 600, 681
- Bate, M. R., Bonnell, I. A., Bromm, V., 2003, *MNRAS*, 339, 577
- Becker, R., *et al.*, 2001, *AJ*, 122, 2850
- Bezanson, R., *et al.*, 2013, *ApJ*, 764, 8
- Bouché, N., *et al.* 2010, *ApJ*, 718, 1001
- Bouwens, R., *et al.*, 2010, *ApJ*, 709, 133
- Bower, R. G., *et al.*, 2006, *MNRAS*, 370, 645
- Bower, R. G., *et al.*, 2012, *MNRAS*, 422, 2816
- Bowler, R.A.A., *et al.*, 2012, *MNRAS*, 426, 2772
- Bruzual, G., Charlot, S., 2003, *MNRAS*, 344, 1000
- Cappellari, M., *et al.* 2012, *Nat.*, 484, 485
- Carretero *et al.*, 2004, *ApJ*, 609, L45
- Cimatti, A. *et al.*, 2009, *Exp. Astron.*, 23, 39
- Coelho, P., *et al.*, 2005, *A&A*, 443, 735
- Conroy, C, van Dokkum, P. G., 2012, *ApJ*, 747, 69
- Conroy, C., Graves, G., van Dokkum, P., 2013, *arXiv:1303.6629*
- Curtis-Lake, E., *et al.*, 2012, *MNRAS*, 422, 1425
- Daddi, E., *et al.*, 2005, *ApJ*, 626, 680
- Daddi, E., *et al.* 2007, *ApJ*, 670, 156
- Dekel, A., *et al.* 2009, *Nat.*, 457, 451
- de La Rosa, I. G., La Barbera, F., Ferreras, I., de Carvalho, R. R., 2011, *MNRAS*, 418, L74
- De Lucia, G., *et al.*, 2006, *MNRAS*, 366, 499
- Dunlop, J. S., 2013, *ASSL*, 396, 223
- Faber, S. M., *et al.*, 2007, *ApJ*, 665, 265
- Fan, X., *et al.*, 2006, *ARA&A*, 44, 415
- Ferrerias, I., Charlot, S., Silk, J., 1999, *ApJ*, 512, 81
- Ferrerias, I., Pasquali, A., *et al.*, 2012, *AJ*, 144, 47
- Ferrerias, I., La Barbera, F., de la Rosa, I. G., Vazdekis, A., *et al.*, 2013, *MNRAS*, 429, L15
- Foley, R. J., Kasen, D., 2011, *ApJ*, 729, 55
- Förster-Schreiber, N.M., *et al.*, 2011, *ApJ*, 739, 45
- Gallazzi, A., *et al.*, 2005, *MNRAS*, 362, 41
- Guo, Q., *et al.*, 2013, *MNRAS*, 428, 1351
- Hewett, P., *et al.*, 2006, *MNRAS*, 367, 454
- Hinshaw, G., *et al.*, 2013, *ApJS*, in press (*arXiv:1212.5226*)
- Hopkins, A. M., & Beacom, J. F. 2006, *ApJ*, 651, 142
- Hopkins, P. F., 2012, *arXiv:1204.2835*
- Iliev, I. T., *et al.*, 2012, *MNRAS*, 423, 2222
- Johansson, J., Thomas, D., Maraston, C., 2012, *MNRAS*, 421, 1908
- Jørgensen, I., *et al.*, 2013, *AJ*, 145, 77
- Kauffmann, G., *et al.*, 2003, *MNRAS*, 341, 54
- Kawata, D., Mulchaey, J. S., 2008, *ApJ*, 672, L103
- Khochfar, S., & Silk, J. 2006, *MNRAS*, 370, 902
- Koleva, M., *et al.*, 2008, *MNRAS*, 385, 1998
- Koleva, M., Vazdekis, A., 2012, *A&A*, 538, 143
- La Barbera, F., Ferreras, I., Vazdekis, A., *et al.*, 2013, *arXiv:1305.2273*
- Laureijs, R., *et al.*, 2011, *ESA/SRE(2011)12*, *arXiv:1110.3193*
- Leonardi, A. J., Rose, J. A., 1996, *AJ*, 111, 182
- Loeb, A., 2013, *ASSL*, 396, 3
- Mandelbaum, R., 2012, *MNRAS*, 420, 1518
- McGee, S., *et al.*, 2011, *MNRAS*, 413, 996
- Milone, A., *et al.*, 2011, *MNRAS*, 414, 1227
- McLure, R. J., *et al.*, 2013, *MNRAS*, in press (*arXiv:1212.5222*)
- Möller, P., *et al.*, *MNRAS*, in press, *arXiv:1301.5013*
- Mortlock, A., *et al.*, 2011, *Nature*, 474, 616
- Moster, B., *et al.*, 2010, *ApJ*, 710, 903
- McQuinn, M., *et al.*, 2007, *MNRAS*, 381, 75
- Muzzin, A., *et al.*, 2012, *ApJ*, 746, 188
- Muzzin, A., *et al.*, 2013, *ApJS*, 206, 8
- Neistein, E., Li, C., Khochfar, S., *et al.* 2011, *MNRAS*, 416, 1486
- Nakamura, E., *et al.*, 2011, *MNRAS*, 412, 2579
- Oliver, S., *et al.*, 2010, *A&A*, 518, L21
- Ono, Y., *et al.*, 2010, *ApJ*, 724, 1524
- Ouchi, M., *et al.*, 2010, *ApJ*, 723, 869
- Pasquali, A., *et al.*, 2009, *MNRAS*, 394, 38
- Pasquali, A., *et al.*, 2010, *MNRAS*, 407, 937
- Pérez-González, P. G., *et al.*, 2008, *ApJ*, 675, 234
- Pietrinferni, A., *et al.*, 2004, *ApJ*, 612, 168
- Pietrinferni, A., *et al.*, 2006, *ApJ*, 642, 797
- Poggianti, B.M., *et al.*, 2006, *ApJ*, 642, 188
- Pontzen, A, *et al.*, 2008, *MNRAS*, 390, 1349
- Renzini, A., 2006, *ARA&A*, 44, 141
- Richards, G. T., Strauss, M. A., Fan, X., *et al.* 2006, *AJ*, 131, 2766
- Robertson, B. E., *et al.*, 2013, *ApJ*, 768, 71
- Rodighiero, G., Daddi, E., *et al.*, 2011, *ApJ*, 739, L40
- Ryan, R. E., Jr., *et al.*, 2008, *ApJ*, 678, 751
- Schenker, M., *et al.*, 2012, *ApJ*, 744, 179
- Schenker, M., *et al.*, 2013, *ApJ*, 768, 196
- Silk, J., Mamon, G., 2012, *Res. in Astron. & Astrophys.*, 12, 917
- Springel, V., *et al.*, 2005, *Nature*, 435, 629
- Stark, D., *et al.*, 2010, *MNRAS*, 408, 1628
- Thomas, D., *et al.*, 2005, *ApJ*, 621, 673
- Thomas, D., Maraston, C., Johansson, J., 2011, *MNRAS*, 412, 2183
- Tinsley, B., 1980, *Fundam. Cosmic Phys.*, 5, 287
- Toft, S., *et al.*, 2012, *ApJ*, 754, 3
- van den Bosch, F.C., Pasquali, A., *et al.*, 2008, *arXiv:0805.0002*
- van de Sande, J., *et al.* 2012, *arXiv:1211.3424*
- van Dokkum, P.G., Conroy, C., 2010, *Nature*, 468, 94
- Vanzella, E., *et al.*, 2011, *ApJ*, 730, L35
- Vazdekis, A., *et al.*, 2003,
- Vazdekis, A., *et al.*, 2010, *MNRAS*, 404, 1639
- Vazdekis, A., *et al.*, 2012, *MNRAS*, 424, 157
- Vulcani, B., *et al.*, 2010, *ApJ*, 710, L1
- Weinmann, S.M., *et al.*, 2006, *MNRAS*, 366, 2
- Weidner, C., Kroupa, P., Pflamm-Altenburg, J., 2011, *MNRAS*, 412, 979
- Wetzel, A.R., Tinker, J.L., Conroy, C., 2012, *MNRAS*, 424, 232
- Wilman, D.J., *et al.*, 2005, *MNRAS*, 358, 71
- Willott, C., *et al.*, 2010, *AJ*, 139, 906
- Woosley, S. E., Heger, A., Weaver, T. A., 2002, *Rev. Mod. Phys.*, 74, 1015
- Worthey, G., 1994, *ApJS*, 95, 107
- York, D. G., *et al.*, 2000, *AJ*, 120, 1579
- Zamkotsian, F. *et al.*, 2010, *Proc. SPIE Vol. 7596*, 75960E

Exploring Habitable Worlds beyond our Solar System

White paper submitted in response to ESA's call for science themes for the L2/L3 missions of its Cosmic Vision program



Coordinator / spokesperson: Andreas Quirrenbach

Contact details:

Prof. Dr. Andreas Quirrenbach

Landessternwarte, Zentrum für Astronomie der Universität Heidelberg

Königstuhl 12

D-69117 Heidelberg

Germany

A.Quirrenbach@lsw.uni-heidelberg.de

+49-6221-54-1792

1. List of Supporters

632 supporting scientists world-wide, among them 424 faculty/staff, 119 postdocs, 46 PhD students and 561 persons with PhD.

ESA Member States (471): **Austria (9):** Baumjohann W., Funk B., Hron J., Kerschbaum F., Khodachenko M., Kupka F., Lammer H., Pilat-Lohinger E., Weiss W.; **Belgium (19):** Absil O., Aerts C., Carone L., Decin L., Dehant V., Gillon M., Javaux E., Lemaître A., Libert AS., Lombaert R., Paladini C., Papics PL., Raskin G., Rochus P., Surdej J., Triana S., Van Hoolst T., Van Winckel H., Vanaverbeke S.; **Czech Republic (3):** Civis S., Ferus M., Heyrovsky D.; **Denmark (4):** Andersen AC., Buchhave LA., Christensen-Dalsgaard J., Karoff C.; **Finland (1):** Lehto H.; **France (158):** Allard F., Almenara JM., Anthonioz F., Arenou F., Aristidi E., Arnold L., Astudillo-Defru N., Augereau JC., Barge P., Barros S., Baudino JL., Baudoz P., Beaulieu JP., Bendjoya P., Bénilan Y., Benisty M., Bertaux JL., Bertout C., Beust H., Beuzit JL., Bezard B., Boccaletti A., Bonfils X., Bordé P., Bouchy F., Bourrier V., Brack A., Briot D., Cabanac R., Caillat A., Cassaing F., Cassan A., Catala C., Cavalié T., Chaintreuil S., Charnay B., Chelli A., Chiavassa A., Cottin H., Coudé du Foresto V., Coustenis A., Creevey O., Crouzier A., Crovisier J., Daniel I., Dayan F., Deleuil M., Delfosse X., Delorme P., Delsanti A., Deschad J.E., Despois D., Dohlen K., Donati JF., Duvert G., Eggl S., Ertel S., Ferlet R., Ferrari M., Forget F., Forveille T., Galicher R., Gargaud M., Giard M., Gonzalez JF., Griessmeier JM., Guyot F., Hébrard G., Hénault F., Hersant F., Hervé G., Homeier D., Huby E., Jacquemoud S., Jolly A., Keller R., Kern P., Kervella P., Kiefer F., Kluska J., Labèque A., Labeyrie A., Lacour S., Lagage PO., Lagrange AM., Lainey V., Lambert J., Lamy L., Langlois M., Le Poncin-Lafitte C., Leach S., Lebreton J., Lecacheux A., Lecavelier des Etangs A., Leconte J., Léger A., Lellouch E., Léna P., Lestrade JF., Liliensten J., Lopez B., Maillard JP., Maire AL., Malbet F., Marcq E., Martic M., Martinez P., Mary D., Mathias P., Menut JL., Meunier N., Michel P., Milli J., Millour F., Monin JL., Montargès M., Morales JC., Mouillet D., Moutou C., Mugnier L., Ollivier M., Pantin E., Parmentier V., Perrier C., Perrin G., Petrov R., Piccoli R., Plez B., Rambaux N., Raulin F., Raymond S., Remus F., Renner S., Robert V., Rodriguez S., Roques F., Rospars JP., Rossi L., Rouan D., Rousset G., Samuel B., Schneider J., Selsis F., Soulez F., Tallon M., Tanga P., Thévenin F., Thiébaud É., Tisserand P., Toubanc D., Turck-Chièze S., Vakili F., Vidal-Madjar A., Vigan A., Vinatier S., Vincendon M., Wakelam V., Westall F.; **Germany (49):** Anglada-Escudé G., Berdyugina S., Bredehöft JH., Cabrera J., Crossfield I., Csizmadia S., Dreizler S., Eigtmüller P., Erikson A., Fossati L., Fruth T., Grenfell JL., Guenther E., Güngör D., Hatzes A., Heller R., Iro N., Jeffers S., Joergens V., Johann U., Kley W., Krivov A., Labadie L., Launhardt R., Lenzen R., Lobo Gomes A., Mancini L., Mandel H., Mohler-Fischer M., Mordasini C., Mueller M., Ofir A., Quirrenbach A., Raetz S., Rauer H., Redmer R., Reffert S., Reinert A., Schuller PA., Sonnabend G., Sornig M., Stangier T., Stepanovs D., Sterken V., Trifonov T., von Braun K., von der Lühe O., Weigelt G., Weiss A.; **Greece (3):** Chatzitheodoridis E., Georgantopoulos I., Varvoglis H.; **Ireland (2):** McKenna-Lawlor S., Ray T.; **Italy (28):** Adriani A., Antichi J., Barbieri M., Bellucci G., Boccato C., Bonavita M., Bonomo AS., Borsato L., Castronuovo MM., Ciavarella A., Cosmovici C., Damasso M., Desidera S., Di Mauro MP., Esposito F., Focardi M., Gai M., Gratton R., Lattanzi MG., Mesa D., Micela G., Nascimbeni V., Ortolani S., Pagano I., Piccioni G., Piotto G., Silvotti R., Sozzetti A.; **The Netherlands (25):** Birkby J., Bloemen S., Braam B., de Kok R., den Hartog R., Di Gloria E., Dominik C., Fridlund M., Gandolfi D., Hooijmakers J., Keller C., Kenworthy M., Roelfsema R., Rottgering H., Schwarz H., Snellen I., Snik F., Sodnik Z., Spaans M., Stam D., Thalmann C., Vermeersen B., Vicente S., Vosteen A., Waters R.; **Norway (1):** Werner SC.; **Poland (9):** Gawroński M., Grzedziński S., Konacki M., Nowak G., Rozyczka M., Siwak M., Smiljanic R., Smith A., Szuszkiewicz E.; **Portugal (16):** Adibekyan V., Boissé I., Brandão I., Coelho J., Delgado Mena E., Figueira P., Krone-Martins A., Luz D., Martins J., Moitinho A., Montalto M., Mortier A., Neves V., Santerne A., Santos N., Sousa S.; **Romania (4):** Belu AR., Fendrihan S., Pricopi D., Suran MD.; **Spain (44):** Aceituno Castro J., Alonso R., Amado PJ., Anglada G., Barrado D., Belmonte JA., Bernabeu G., Caballero JA., Claret A., Colomé J., Deeg H., Diaz-Sanchez A., Eiroa C., Esposito M., Fors O., Galadí-Enriquez D., Gutiérrez JL., Lara LM., Lopez-Puertas M., Maldonado J., Manteiga M., McCormac J., Modroño Z., Molina-Cuberos GJ., Montes D., Ortiz-Gil A., Palle E., Perez Cagigal M., Perez-Garrido A., Ramón Ballesta A., Régulo C., Ribas I., Roca Cortés T., Rodler F., Rodrigo R., Rodríguez E., Sanchez-Blanco Mancera E., Sanchez-Lavega A., Sanz-Forcada J., Skillen I., Valle PJ., Villaver E., Voss H., Zapatero Osorio MR.; **Sweden (15):** Bjerkeli P., Brandeker A., Carrera D., Davies M., Feltzing S., Johansen A., Justtanont K., Liseau R., Nordh L., Olberg M., Olofsson G., Persson CM., Rickman H., Romeo A., Wiström E.; **Switzerland (19):** Alibert Y., Bartholdi P., Ehrenreich D., Fortier A., Hagelberg J., Holl B., Lendl M., Lovis C., Mayer L., Mayor M., Mégevand D., Motaiebi F., Pfenniger D., Quanz S., Saesen S., Sahlmann J., Schmid HM., Udry S., Wytenbach A.; **United Kingdom (62):** Aigrain S., Astley J., Bagnulo S., Barstow J., Bochinski J., Bowles N., Bradshaw T., Brown D., Burleigh M., Cameron A., Catalan S., Cho J., Faedi F., Ferlet M., Fletcher L., Frith J., Galand M., Gomez Maqueo Chew Y., Goodwin S., Hall A., Hargreaves R., Haywood R., Hellier C., Hodgkin S., Hoekzema R., Irshad R., Irwin P., Ivison R., Jones H., Kolb U., Kraus S., Lim T., Lucas P., Mohanty S., Mueller-Wodarg I., Nayakshin S., Nelson R., Norton A., Ogilvie G., Pickering J., Pollacco D., Raven J., Rice K., Savini G., Serjeant S., Simpson F., Smalley B., Smith M., Stevens I., Stixrude L., Tinetti G., Tsapras Y., Turner O., Vidotto A., Walton N., Wheatley P., White G., Williams I., Wyatt M., Young J., Young R., Yurchenko S.

States in cooperation with ESA (14): **Canada (6):** Doyon R., Gregory P., MacPhee S., Marois C., Matthews J., Metchev S.; **Hungary (7):** Borkovits T., Dálya G., Kiss L., Klagyivik P., Regaly Z., Szabo R., Szabó G.; **Slovenia (1):** Zwitter T.

Non ESA States (147): **Argentina (3):** Beauge C., Giuppone C., Guilera OM.; **Australia (1):** Haubois X.; **Brazil (1):** Ferraz-Mello S.; **Chile (10):** de Graauw T., Grellmann R., Hales A., Ivanov V., Jenkins J., Kabath P., Mawet D., Minniti D., Muller A., Rabus M.; **China (5):** Gu S., Guo J., Wang T., Wang X., Zhang YH.; **India (1):** Vahia M.; **Israel (6):** Helled R., Mazeh T., Perets H., Podolak M., Vazan A., Zucker S.; **Japan (9):** Inaba S., Kawahara H., Kokubo E., Matsuo T., Murakami N., Narita N., Nishikawa J., Shibai H., Tamura M.; **Russia (6):** Bisikalo D., Demekhov A., Ksanfomalit L., Kulikov Y., Schematovich V., Tavrov A.; **Serbia (1):** Novakovic B.; **Taiwan (1):** Ip WH.; **United States of America (103):** Adams F., Albrecht S., Angerhausen D., Armstrong JT., Arras P., Aufdenberg J., Barman T., Baron F., Batalha N., Belikov R., Benneke B., Boley A., Boss A., Boyajian T., Breckinridge J., Brown R., Bryden G., Cabrera N., Caldwell D., Carpenter K., Catanzarite J., Chaboyer B., Choquet E., Clampin M., Cowan N., D'Angelo G., Danchi W., Davis P., de Val-Borro M., Defrere D., Dekens F., Memory BO., Deroo P., Doyle L., Duchene G., Dumusque X., Forrest W., Gaulme P., Gilliland R., Greene T., Groth E., Guyon O., Guzik J., Haghighipour N., Havel M., Huang X., Isella A., Janson M., Jenkins J., Jevtic N., Kane S., Karalidi T., Kasdin NJ., Kasting J., Kopparapu Rk., Lai D., Lawson P., Lay O., Lisse C., Livengood T., Lopez-Morales M., Lozi J., Lunine J., Lyon R., Mandell A., Marley M., Martinache F., Mather J., McElwain M., Mennesson B., Mitchell D., Monnier J., N'Diaye M., Orton G., Palmer P., Pipher J., Plavchan P., Pluzhnik E., Ragozzine D., Ridgway S., Rinehart S., Roberge A., Sahai R., Sanchis-Ojeda R., Sciamma-O'Brien E., Serabyn E., Sparks W., Stapelfeldt K., Su K., Taylor SF., ten Brummelaar T., Traub W., Trauger J., Tsvetanov Z., Turner N., Unwin S., Vakoch D., van Belle G., Vasisht G., Warner J., Woolf N., Wordsworth R., Yelle R.

For an updated list, with institutional details and more complete statistics, see www.blue-dots.net.

2. Executive Summary

We are now at a unique moment in human history. For the first time, we are able to build instruments that allow us to investigate directly how unique the Earth is and whether or not we are alone in the Universe. Discovering Earth's sisters and possibly life is the first step in the fundamental quest of understanding what succession of events led to the emergence and survival of life on Earth. For this, we need to know how, where and when stars form from gas and dust and how, where and when planets emerge from this process. This is certainly one of the most important scientific goals that ESA and Europe could set themselves. [ESA, Cosmic Vision, 2005]

Triggered by the discoveries of the first planets outside the Solar System twenty years ago, the study of planetary systems associated with other stars and of the properties of exoplanets has grown into one of the most vibrant fields of astrophysics. Surveys covering thousands of stars have yielded nearly 1000 confirmed planet discoveries, and steady progress is being made from Jupiter-size objects towards Neptune- and now Earth-size planets, driven by refinements in instrumentation and observing techniques.

Thanks to the ubiquity of planetary systems and the broad diversity of exoplanets, in terms of size, composition, temperature, and orbits, we have already begun the exploration of some distant worlds by remote sensing. The population of gaseous exoplanets at short orbital periods (the so called hot Jupiters or hot Neptunes depending on their size/mass) that includes a significant fraction of transiting objects has provided us with the data to expand comparative planetary science beyond our own Solar System. Transmission and emission spectroscopy achieved during primary transits and secondary eclipse, respectively, as well as orbital spectrophotometry for transiting and non-transiting planets have made possible the detection of atmospheric species and clouds, and the measurement of atmospheric temperatures, vertical/longitudinal thermal structures and wind speeds. First applied to the most favorable hot Jupiters, these techniques are now providing results on smaller and cooler planets. The trend towards the observation of terrestrial exoplanets will continue, but characterizing them, assessing their habitability and searching for signs of biological activity implies an ambitious space program that will aim, beyond the next decade, at the direct imaging of exoplanets that do not necessarily transit, which represent the vast majority of exoplanets and include our nearest neighbors.

The stunning progress in exoplanet science during the past years has broadened our view, and changed the perspective we had on these questions when they were framed within the Cosmic Vision program. Whereas previous proposals for large space missions, informed solely by our own Solar System, focused strongly on the possibility of detecting “Earth twins”, we are now in a position to formulate questions about habitability and ultimately extraterrestrial life in the more general context of comparative planetology, with a large number of systems available for study.

Among the remarkable feats of the exoplanet community has been the ingenuity with which new observing techniques have been invented and put into successful use over the past twenty years. We now have a diverse set of tools at our disposal, with which we can explore different aspects of exoplanetary systems. A number of complementary approaches have been identified that can address habitability from different angles. Coronagraphs and infrared interferometers have been studied at some level of detail, and other more recent concepts (external occulter and integrated-light telescopes) also show considerable promise. While none of these is ready yet for flight, the rapid progress over the past few years in the development of the key enabling technologies gives confidence that an exoplanet exploration mission will become viable technically and financially in time for implementation in the middle of the next decade.

Ever since the first discovery of a planet around a Sun-like star (Mayor & Queloz 1995), Europe has been playing a leading role in exoplanet science, with arguably the best ground-based instruments and the first dedicated exoplanet space mission (CoRoT). Future plans include the small mission CHEOPS, as well as two strong contenders for M3 (PLATO and EChO). The adoption of “Exploring Habitable Worlds beyond the Solar System” as the theme for a large mission will enable ESA to secure its leading role in this endeavor into and beyond the next decade.

3. The Science Case of Exoplanet Characterization: Atmospheres, Surfaces, Habitable Environments

Probably for as long as humans have looked up to the night sky, they have tracked the wanderings of the brightest specks of light across the background of seemingly unmoving stars. Observed through the first telescopes, these planets appeared to be physical bodies with their own moons around them, orbiting the Sun just like the Earth; a discovery that triggered the upheaval of the commonly held view of the place of the Earth and mankind in the Universe. With the increasing quality of ground-based telescopes, and, since the 1960's, with space-based telescopes and dedicated spacecraft, the Solar System's planets and moons changed from merely slightly differently colored, fuzzy dots into unique and stunning worlds of their own.

3.1. The Diversity of Planets

One of the lessons of Solar System exploration is diversity. Since the 1990's, an even broader perspective has emerged with the discovery of exoplanets, as planets around other stars are called. And these worlds present an even greater diversity than those of our Solar System, in terms of observed masses (from $0.67M_{\oplus}$ to $30M_{\text{Jup}}$ and beyond), orbital range (from 0.006 AU to more than 1000 AU), orbital eccentricity, and host star properties.

One of the key drivers for Solar System exploration has been the search for life elsewhere. Finding life forms on another planet or moon would help to shed light on the formation and evolution of life on Earth. A prerequisite for life appears to be liquid water. The search for life is therefore closely linked with the search for liquid water and habitable conditions. So far, we haven't found liquid water on other planets or moons. Venus was long suspected to harbor water, because while this planet is closer to the Sun than the Earth, and thus receives a much larger solar flux, its clouds reflect most of this flux back to space. It was only in the 1960's that Venus's surface temperature was found to be close to 500°C , due to the extreme greenhouse effect in its thick carbon dioxide atmosphere, and in the 1970's, it was discovered that its clouds consist of sulfuric acid instead of water (Hansen and Hovenier 1974).

Mars orbits on the outer edge of the Sun's so-called habitable zone – the region around a star where the stellar flux that is incident on a planet would allow liquid surface water to exist. While water-ice is abundant on Mars (e.g. Plaut et al. 2007), the surface pressure and the temperature are too low for liquid surface water to be in equilibrium with the atmosphere. Mars shows evidence for the geologically recent presence of liquid surface water, indicating that in the past, the Martian atmosphere might have been much thicker, the climate much warmer and wetter, and more favorable for life (Solomon et al. 2005). Traces of ancient life, and even subsurface pockets of current life, will be searched for by ESA's upcoming Exomars mission. ESA's L-class mission JUICE will get a close-up of moons of the gas giant Jupiter where thick crusts of water-ice are predicted to cover deep, possibly habitable, oceans of liquid water (Grasset et al. 2013).

While the search for life elsewhere in the Solar System is still ongoing, it has become clear that there cannot be evolved life, beyond very primitive micro-organisms, and then only on a very few bodies (Mars and some of the giant planets' moons). However, we know that there are many more observable habitable exoplanets and that no present observation excludes evolved forms of life on them. Therefore, to find abundant life as we know it on Earth, we have to look beyond our own planetary system. Only a search for life on planets around other stars would answer the longstanding question whether we are alone.

3.2. Reaching beyond the Solar System

After two decades of hunting for exoplanets, we have identified almost 1000 of them, and we now know that there are at least as many planetary companions as there are stars in our galaxy. And thanks to the increased sensitivity of instruments and analysis methods, it has also become apparent that small exoplanets are in fact much more common than giant, gaseous ones. Indeed, according to recent estimates, at least 10% of stars could have a small planet orbiting in their habitable zone.

We know very little about these exoplanets, apart from their masses (and often only a lower limit is known), sizes, and the diameters of their orbits. As we know from the terrestrial planets in the Solar System, whether or not a planet in or near the habitable zone of a star has surface conditions compatible with life will depend strongly on the chemical composition and thickness of its atmosphere. The next step in the thriving field of exoplanet research should therefore be studying the physical properties, atmospheres and surfaces of exoplanets. This search for habitable conditions, i.e. conditions that are compatible with the presence of liquid water, or for actual signatures of life as we know it, i.e. spectral features due to vegetation (Fuji et al. 2010) or gases like O₂ and CH₄ (Rauer et al. 2011), does not have to be confined to small, rocky exoplanets around solar type stars. Other types of stars have planets, too (indeed, several of the known small exoplanets orbit red dwarfs). And, just like Saturn's largest moon, Titan, moons of gas giants that are located in or near a star's habitable zone could also have atmospheres and habitable conditions (Heller 2012).

Apart from the search for life outside the Solar System, studying exoplanets' physical properties, atmospheres and surfaces, will provide us with a wealth of knowledge on the formation and evolutionary processes that shape not only planetary systems as a whole, but also the interiors, atmospheres, and surfaces of individual planets. The large diversity exhibited by the Solar System planets in, amongst others, their atmospheric chemical composition and structure, the radiative and dynamical processes governing their climates and weather patterns, their internal composition and structure, their magnetic fields, and even in the properties of their moons and ring systems, has allowed us to significantly broaden our understanding of how planets work. These fields of study are now being enriched even further by discoveries of exoplanets with properties that are not found within the Solar System. We are thus at the dawn of a new science: comparative exoplanetology.

3.3. Physical Properties of Exoplanets: A Rich Field for Exploration

The diversity in physical properties of exoplanets, such as size, mass, composition, and orbits, and the paucity of information about the formation and evolution of these planets and their atmospheres, provide many opportunities for discoveries and new insights. Meaningful comparative planetology that connects these systems to the Solar System is now becoming possible. Indeed, more detailed observations, observational baselines long enough to cover several orbital periods and/or seasonal changes, and the relentless growth in computing power for data analysis and numerical modeling of physical processes have revealed significant and important gaps in our knowledge and understanding. As an example, a general circulation model that satisfactorily simulates the Earth's current climate and weather patterns will not do the same for Venus upon changing the solar irradiation or the planet's obliquity, rotation period, or atmospheric thickness and composition. As another example, we don't know whether Venus and the Earth started off with very different atmospheres, or with similar ones. If they were similar: when and why did the divergence start? How stable are such atmospheres anyway? And what was the Earth's atmosphere like when there was liquid surface water while the Sun was young and faint?

Exoplanet characterization will shed new light on these important questions and enable new approaches to the open problems (e.g., Medvedev et al. 2013). Although the detection methods that have harvested the vast majority of exoplanets known today all have peculiar biases towards planetary sizes, orbital distances, or temperatures, it can safely be concluded that exoplanets cover a huge parameter space: from young and hot to old and cold, from small and solid to giant and gaseous, from tight to wide orbits, and from circular to eccentric orbits that give rise to extreme temperature changes and hence to extreme dynamical processes in planetary atmospheres. Differences in types of parent stars – their composition, size, and activity – could also result in differences in types of planets. Exoplanet characterization will fill the gaps in our knowledge and understanding because it changes the universe around us into a huge physics laboratory where arguably enough planets can be probed to tackle a range of variables, including habitability and life (e.g. Grenfell et al. 2007).

Comparative exoplanetology is thus undoubtedly among the most exciting areas in all of science in the 21st century. The questions raised by the extreme complexity of this field will keep challenging space- and ground-based technologies for decades to come. By scheduling a large mission for launch in 2034, ESA will further energize the field and take a large step forward towards a fuller understanding of habitable worlds beyond our Solar System.

4. Exoplanet Characterization: Present and Near Future

The study of extrasolar planets is presently one of the fastest-growing areas of astrophysics. While surveys with different techniques (radial velocities, transit photometry, microlensing, coronagraphic imaging, and soon astrometry with GAIA) are discovering planets and planetary systems at an accelerating pace, we are also moving progressively into the era of exploration and characterization with photometric and spectroscopic methods. With increasing instrumental sophistication, each technique progresses from large to small planets, and most of them from hot to cool. This is particularly important in the context of the quest for habitable planets, which are within reach of the discovery programs now, but whose characterization will require larger and more advanced observing tools.

4.1. *Studies of Exoplanet Atmospheres with Transit Spectroscopy*

Out of the nearly 1000 exoplanets discovered so far, the ~300 transiting planets represent a unique opportunity to access spectral features of exoplanetary atmospheres. Transiting exoplanet properties for which we have spectroscopic information are particularly diverse, with star-planet distances ranging from 0.014 to 0.45 AU, equilibrium temperature ranging from 540 to over 3000 K and orbital eccentricity ranging from 0 (for the circularized hot Jupiters) to 0.93. The Spitzer Space Telescope has led to great advances in the understanding of the composition of transiting giant planets. The infrared spectra as observed with secondary eclipse data of hot Jupiters are believed to be shaped predominantly by water absorption (Burrows et al. 2005, Seager et al. 2005), but other molecules such as methane also play a role (Swain et al. 2008). While methane in particular could become more important for cooler planets, its abundance in GJ 436b is still controversial (Line et al. 2011, Stevenson et al. 2010, Knutson et al. 2011, Beaulieu et al. 2011). For close-in planets orbiting luminous stars, strong irradiation could flatten the temperature gradient and weaken absorption features in the spectrum at the time of eclipse (Fortney et al. 2006). The results on HD 189733b from Spitzer/IRAC (Tinetti et al. 2007, Ehrenreich et al. 2007, Beaulieu et al. 2008, Désert et al. 2009) and HST/NICMOS (Swain et al. 2008, Sing et al. 2009, Gibson et al. 2011) provide the first glimpses at the atmospheric composition of this hot Jupiter, revealing the signatures of molecules and the presence of haze. Important observations of primary transits have also been made using Spitzer (Richardson et al. 2006, Gillon et al. 2007, Nutzman et al. 2009). Large ground-based telescopes have also been used successfully to obtain spectra of a few bright exoplanets (Snellen et al. 2008), recently even of some that do not transit (Brogi et al. 2012).

A new chapter in exoplanetary science began with the discovery of the first transiting super-Earths (Léger et al. 2009; Charbonneau et al. 2009), for which measurements of mass and radius are possible. GJ 1214b is an especially interesting object since its spectrum has also been measured, giving us constraints on the nature of its atmosphere. Spectra have been obtained in transmission during primary transit and in emission during secondary eclipse, from the ground and from space. The flatness of the spectra can have several interpretations, but it definitely rules out a clear atmosphere with solar composition. Possible explanations are depletion of CH₄ or a dense cloud layer (Bean et al. 2010, Croll et al. 2011, Crossfield et al. 2011, Berta et al. 2012). Alternatively, the planet might be significantly smaller than indicated by the best present estimates and not possess a substantial atmosphere at all (Bean et al. 2011). These conclusions emphasize the diversity of the planetary conditions but also the need for consistent, reliable observational constraints.

Transit surveys like NGTS from the ground, and TESS and PLATO (if selected for M3) from space, will discover new interesting targets for transit spectroscopy. With missions such as the James Webb Space Telescope (JWST), we should be able to acquire better quality spectra of transiting Hot Jupiters and Hot Neptunes, to access atmospheric signatures of a few super-Earths, and to start characterizing such planets. Additionally, the EChO mission (currently under consideration by ESA as a candidate M3 mission) aims at measuring the largest technically and financially feasible part of the planetary spectrum (from 0.4 to 11 or even 16 μm). This mission is also optimized for time-sequence studies of these planets, thus providing access to meteorological phenomena through observations of temporal variability. High-resolution near- and mid-infrared spectroscopy with the next generation of extremely large telescopes will provide further insight into the composition and dynamics of giant and possibly even Super-Earth planet atmospheres (Hedelt et al. 2013,

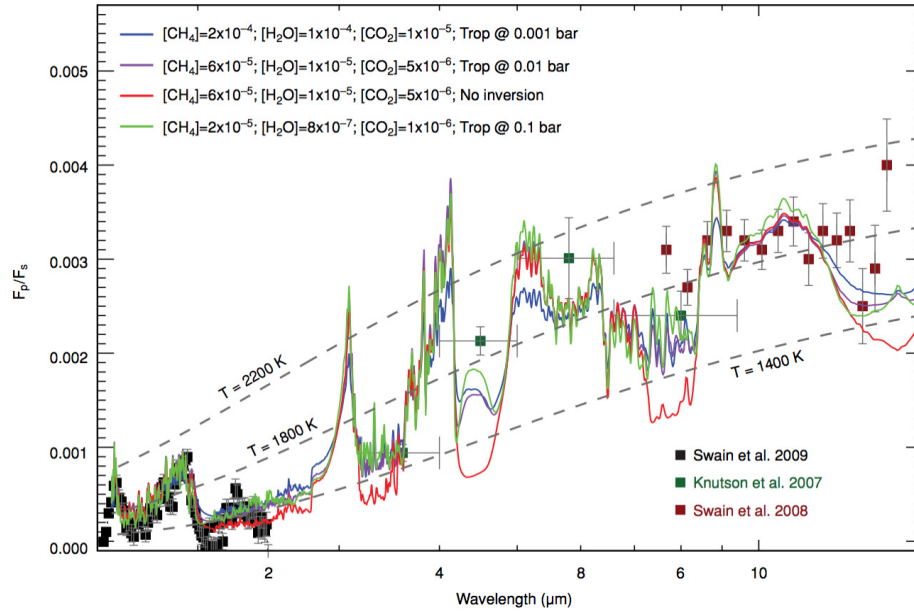


Figure 1: Synthetic emission spectra of the hot Jupiter HD209458b together with various observational data points (Swain et al. 2009). The data are from various sources, and there are gaps in the spectral coverage. Despite over a decade of study, the sparse data shown here represent the highest-quality exoplanet spectrum obtained to date.

Snellen et al. 2013). By the end of the next decade we will thus have acquired experience with the observational and theoretical tools needed for the analysis of exoplanet atmospheres. The logical next step will be the extension of such observations to potentially habitable planets.

4.2. Direct Detection with Coronagraphic Imaging

While radial velocity (RV) and transit searches are pushing towards the discovery of lighter planets down into the Super-Earth regime, direct imaging so far has revealed only a handful (about 30) of planetary-mass objects due to the high contrast that is needed at small separations (less than 1 arcsec). Nevertheless, a few emblematic objects have been discovered and studied, like β Pic b (Lagrange et al. 2009, 2010), the four planets around HR 8799 (Marois et al. 2008, 2010), and the intriguing supposedly planetary object in the Fomalhaut system (Kalas et al. 2008, 2013). In this respect, young systems offer a reduced star-to-planet contrast, as the planet's high, early luminosity decays slowly with age. The imaged planets are all located at physical separations larger than about 10 AU (β Pic b being the closest) and have masses (estimated from their luminosities) larger than 5-10 M_{Jup} . These systems are also very young, with ages ranging from a few Myr to a few hundred Myr. Overall, current ground-based instruments are now able to reach contrasts as large as 10^6 at typically 10 AU for the closest stars (10-20 pc), while the detection of more mature planets would require a dynamic range of more than 10^9 .

Still, these few objects provide crucial information for understanding the physics of exoplanets, in particular the diversity with respect to the planets found by RV and transits, which has important implications regarding their formation and evolution. We have learned that planets can be much more massive than those in the Solar System, that they form relatively quickly (β Pic has a well-constrained age of 12 Myr), and that different mechanisms could be required to explain their formation at large distances (gravitational instabilities as opposed to core accretion). This knowledge is inaccessible by indirect detection because RV and transit photometry concentrate on the inner part of old systems, which have certainly lost memory of initial conditions due to migration and/or planet scattering. In addition, we are starting to obtain atmospheric properties through photometry (temperature and surface gravity, Bonnefoy et al. 2011, 2013) and low resolution spectroscopy (composition, Janson et al. 2010, Konopacky et al. 2013, Oppenheimer et al. 2013). Finally, imaging is the only technique that provides a global picture of planetary systems including the distribution of the dust in protoplanetary and debris disks, which allows us to study the disk-planet connection (Lagrange et al. 2012) and to infer the presence of planets (Wyatt 2003).

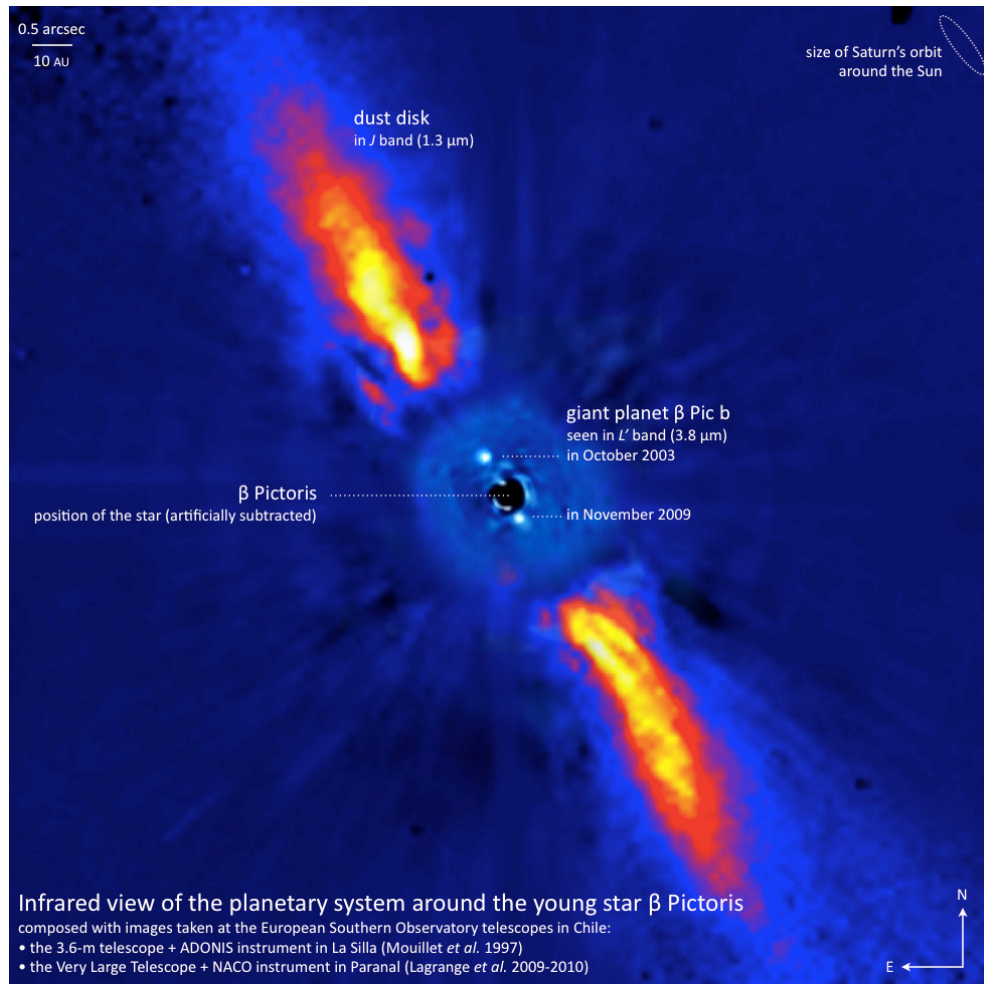


Figure 2: View of the β Pic system showing the debris disks superposed with the planet images in Nov. 2003 and Nov 2009 (Mouillet *et al.* 1997, Lagrange *et al.* 2009, 2010).

In the coming decade, several instruments optimized for direct imaging (extreme adaptive optics and coronagraphs) will be installed at large ground-based telescopes, starting with the series of planet finders like GPI (Macintosh *et al.* 2008) and SPHERE (Beuzit *et al.* 2008). These are designed to reach contrasts of $10^6 - 10^8$ very close to the star (>1 AU) and detect a new population of young giant planets with masses similar to that of our Jupiter. Characterization of their atmospheres will become feasible with low resolution spectrographs in the near IR ($1 - 2.3\mu\text{m}$). SPHERE and GPI will also put important constraints on the frequency of giant planets at large orbital periods. Towards the end of this decade, JWST (and possibly SPICA) will come with a suite of IR instruments, all having coronagraphic observing modes. JWST's NIRCAM, MIRI and NIRISS will allow detailed atmospheric characterization (mostly photometry but also low resolution spectroscopy) of the planets discovered by SPHERE but at longer wavelengths (2.5 to $16\mu\text{m}$), and will likely push the detection limit to the range of ice giants with long periods.

5. The Landscape in 2034

5.1. Further Developments in Exoplanet Detection and Characterization

Over the past two decades, the field of exoplanet research has grown faster than any other in astrophysics, both in terms of objects to be studied (from zero to at least one thousand), and in terms of active scientists (from a handful of part-timers to a large vibrant community). This growth has been driven by a strong diversification and many refinements of the available observing techniques

and facilities. As it is likely that many of these developments will continue in the foreseeable future, it is not easy to extrapolate the state of the field over another twenty years, or to predict some of the major discoveries that will undoubtedly be made during that period. Nevertheless, we can foresee major features of the landscape of exoplanet exploration around the envisaged launch date of the L3 mission:

- Radial-velocity surveys will have performed exhaustive searches for terrestrial planets orbiting nearby stars. ESPRESSO at the 8.2m VLT, CARMENES at the Calar Alto 3.5m and SPIRou at the 3.6m CFHT will be able to detect $1M_{\oplus}$ planets in the habitable zones of “quiet” Sun-like stars and M dwarfs.
- TESS will have detected the brightest targets harboring transiting terrestrial planets in orbits up to ~ 30 days, and a small sample of longer-period planets. PLATO, if selected for M3, will expand the parameter range with a catalog of temperate terrestrial planets in orbits up to the habitable zone of Sun-like stars, where they are likely able to retain their atmospheres, and to develop habitable conditions.
- Thanks to the combination of radial velocities with transit observations (CHEOPS), the mass-radius relation will have been established down to Earth-size planets; consequently the bulk composition of these planets will be understood (Sohl et al. 2012). Note, however, that strong degeneracies exist for planets with an atmosphere (Adams et al. 2008).
- High contrast imaging surveys (SPHERE, GPI, JWST, SPICA, ELTs) will have discovered many young Jupiters and Neptunes; spectroscopic follow-up will have provided more detailed information on a subset of them.
- ALMA and LBTI will have imaged debris disks and determined the prevalence of zodiacal dust disks around solar-type stars.
- Astrometry with GAIA will have discovered most of the giant planets between 15pc and 150pc, but it will still miss nearby low-mass planets. The latter could be found by a dedicated astrometric mission, which would be an attractive candidate for a medium-size mission (M4).
- Transit spectroscopy is already a rich field for hot giant planets orbiting close to their stars, and will expand towards cooler and smaller planets with data from many complementary facilities, including JWST, EChO (if selected for M3), and infrared instruments at extremely large telescopes (including the E-ELT).

However, and significantly, spectroscopic investigations of potentially habitable planets will still be lacking, because none of the facilities that are presently foreseen for construction during the next twenty years will provide data with sufficient scope and quality to make meaningful statements about habitability. Thus, even if the status of exoplanet exploration around 2034 cannot be foreseen in detail at present, **it is logical for ESA to focus its L3 mission on the characterization of habitable worlds**. Missing this chance would in fact endanger the leading role that Europe has been able to establish in the field of exoplanets from its inception. In contrast, an early adoption of this topic for L3 would provide a framework in which the scientific focus and output of intermediate investments in space and on the ground can be optimized.

5.2. Targets for Habitable World Exploration

The design and optimization of any exoplanet exploration mission depends critically on the number and properties of the targets it is to observe. From the preceding discussion it follows that important progress will be made in this regard within the coming years:

- Kepler and microlensing surveys are establishing η_{\oplus} , the fraction of stars with planets in their habitable zones. This will tell us on a statistical basis the number of planets available for exploration within a certain volume.
- Next-generation RV surveys (e.g. ESPRESSO in the visible, CARMENES and SPIRou in the near-IR) will discover Earth-like planets in the habitable zones of “favorable” Sun-like stars and M dwarfs, i.e., in a subset of stars with rather quiet photospheres. This will provide an actual sample of target planets within 15pc that are amenable to spectroscopic

characterization. PLATO, which has been proposed as the M3 mission, would provide a sizeable sample of additional targets (the closest transiting planets).

- An astrometric mission, which could for example be flown as ESA's M4 mission, could conduct an exhaustive search of the habitable zones of all nearby stars down to 1 M_{\oplus} , thus establishing the "ultimate" target sample for further exploration.
- ALMA and LBTI will characterize debris disks, which could manifest themselves as "noise" for planet characterization missions.
- Ground-based long baseline interferometry (VLTI, CHARA, NPOI), as well as asteroseismology from space with TESS, PLATO (if selected as the M3 mission) and from the ground, combined with parallaxes from GAIA, will establish precise values for the most important properties of the host stars (mass, radius, distance, age), which are needed to determine the corresponding properties of their planets.

In summary, one can be confident that the present uncertainties about the number and properties of potential targets, which are sometimes seen as impediments for the implementation of a cost-effective planet characterization mission, will largely be removed within the next decade. Furthermore, specimens representing different categories of exoplanets – including potentially habitable ones – will be known 20 years from now. While an exhaustive census of the solar neighborhood would certainly be desirable (and possible), it is by no means a prerequisite for starting the in-depth characterization of those planets that we know.

While it is thus still premature to define a possible mission target catalog, we can estimate the number of potentially habitable planets using current Kepler and Corot results that suggest a value of η_{\oplus} of 10 to 20% for F, G, and K stars (Batalha et al. 2013, Fressin et al. 2013). The corresponding value for M dwarfs may even be as high as ~50%. Based on these numbers, a variety of mission architectures are capable of characterizing samples of tens of potentially habitable planets.

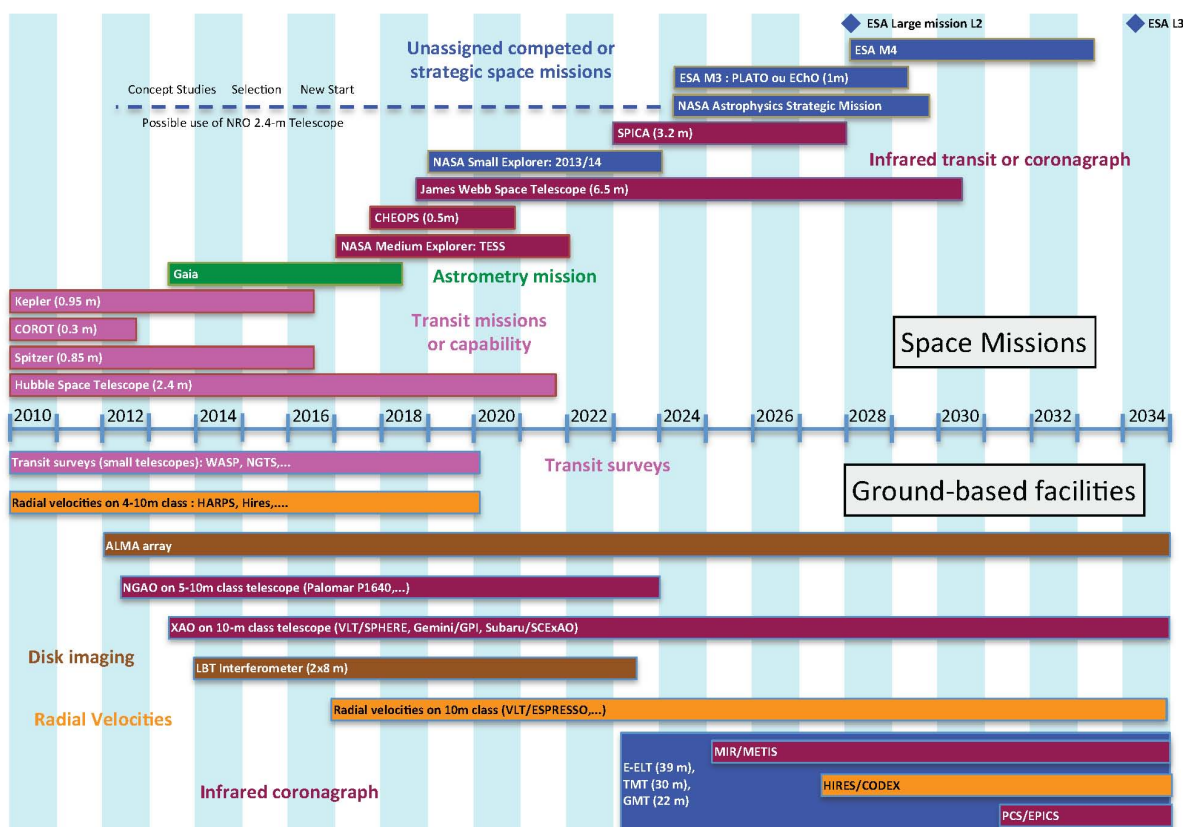


Figure 3: Chart summarizing the instrumental landscape up to 2034 both in space (upper part) and on the ground (bottom part). Bars of same color correspond to similar techniques.

6. Key Science Questions and Required Observations

The in-depth characterization of the physical and chemical properties of terrestrial exoplanets is a long term goal. It will require spectroscopy with sufficient resolving power and SNR at both thermal and scattered wavelengths. A broad spectral coverage from visible to mid-IR is necessary to assess the radiative budget of the planet, which is the key to understanding its climate. In addition, it also enhances the number of observable spectral molecular signatures, making the identification of molecules more robust and minimizing the uncertainty on their abundances. Observations will also have to be spread over several orbital periods, with different sampling frequencies, in order to characterize the signal variability associated with climate, rotation, seasons, phases and variations in the stellar luminosity. Polarimetry combined with visible spectroscopy would also constitute an additional way to derive the atmospheric gaseous/particle content. Techniques to constrain the mass and radius of the planet will not only contribute to understanding the nature of the planets but will also strongly increase the information content of the spectra whose interpretation depends on both the gravity and the radius. The radius, in particular, allows converting observed fluxes into albedos (scattered light) and brightness temperatures (thermal emission). The radii and masses of non-transiting planets can be constrained from spectra, although at reduced accuracy and with reliance on suitable models.

The general goals of obtaining an in-depth understanding of the physical and chemical properties of terrestrial exoplanets, and of developing the notion of habitability in the broader frame of comparative planetology, can be broken down into these more specific questions:

1. **What are the physical characteristics of the atmospheres (composition, temperature and pressure profiles, haze, clouds, winds)?**
2. **What is the internal structure of those planets?**
3. **What is the nature and composition of the surface (rocky, liquid, icy...)?**
4. **What is the time (and seasonal) variability of those features? Which roles do dynamics and photochemistry play?**
5. **What are the key processes which govern the chemistry in those exotic atmospheres?**
6. **If we discover chemical disequilibria – could they be caused by life?**

The information needed to address those questions can be provided by remote sensing observations. In particular:

Bulk planetary composition and internal structure:

The planetary composition and internal structure can be constrained in several ways. Measuring the planetary radius and mass will determine the mean density directly. Measuring the atmospheric composition will allow us to distinguish a Neptune-like planet from a terrestrial planet. In addition, finding traces of volcanic gases in the atmosphere may provide insight about the composition of the interior. Finally, an indirect estimate of the planet's surface gravity may be made through primary transit observations of the atmosphere, which give an indication of the atmospheric scale height.

Atmospheric composition:

The molecules which are most abundant, or have the strongest signatures, can be detected at low to medium spectral resolving power (e.g. H_2O , CO_2 , O_2 , O_3), from the UV to the IR, depending on the absorption properties of the molecular species (Des Marais et al. 2002). To detect less abundant or weaker molecular signatures a spectral resolving power of ~ 100 or higher is needed (e.g. C_2H_2 , HCN). Most atoms and ions can be found in the UV-VIS-NIR (e.g. Na , K , H_3^+). Very high spectral resolution is needed to resolve these lines. To estimate the elemental and molecular abundances, a combination of appropriate spectral resolving power R and wavelength coverage is desirable. The required R will mainly depend on the molecule/element, on the wavelength interval and atmospheric region we are probing. To estimate chemical gradients (spatial and temporal) we need to be able to spatially resolve the planet (e.g. through observations at different planetary

phases) and/or monitor the atmosphere with a cadence and integration time which are shorter than the specific chemical reaction rate.

Thermal structure:

The effective temperature of the planet can be calculated knowing the flux from the host star, the orbital parameters and the planetary albedo. However the planetary albedo depends not only on the reflectivity of the surface but also on the opacity of the atmosphere and on the cloud properties. The atmospheric opacity is also responsible for a greenhouse effect that increases the surface temperature beyond the effective temperature, with obvious consequences for the habitability of the planet.

The infrared is the best interval to probe the vertical thermal structure of a planetary atmosphere through spectroscopic absorption signatures of molecules. The higher the spectral resolution, the higher the altitude we can probe: for example with Spitzer and Hubble low resolution spectroscopy and photometry, we typically sound the atmospheric region between the bar and millibar levels.

Horizontal thermal gradients require the ability to probe the planet at different phases. This can be attempted by monitoring light curves of transiting and non-transiting planets.

Indirect constraints on the temperature can be obtained through the temperature dependence of molecular and elemental absorption properties, or through measurements of the atmospheric scale height with transit data (provided other parameters such as gravity and the main atmospheric components are known).

Atmospheric dynamics and variability:

Atmospheric dynamics and temporal variability can be monitored by repeated observations of the thermal structure of the atmospheres. Observations of variations can also provide information on the rotation rate and on seasonal changes. In any case, the integration time of the observations needs to be sufficiently short to sample the variations; in the case of periodic processes like the diurnal rotation of planets, phase binning can also be employed.

Clouds and aerosols:

The presence of clouds and/or hazes in a planetary atmosphere profoundly influences the radiation balance and the climate of a planet. The optical properties of clouds and hazes depend on the size, shape and distribution of the particles. Spectroscopic observations in the visible and the infrared of the planetary atmosphere can provide constraints on those parameters. Polarized light in the visible is well-suited for detecting and characterizing clouds and hazes. Repeated observations are necessary to detect temporal variability, formation processes, and typical patterns.

Albedo and surface:

Spectral and photometric observations of the planet in the visible and near-IR spectral range provide constraints on the planetary albedo and the surface type of a planet (provided there is a surface, and the atmosphere is transparent enough at least at some wavelengths to get a glimpse of it). Also in this case, polarization may be the key for retrieving the type of surface. The presence of liquid water at the surface might be detected thanks to its glint (Robinson et al. 2010).

Magnetic field and upper atmosphere:

Observations of ionized species mainly in the UV (notice though that H_3^+ is detectable in the NIR) offer the possibilities of sounding the upper atmospheres of exoplanets, exploring star-planet interactions, and investigating escape processes.

Moons and rings:

Moons of transiting exoplanets will probably be detected rather soon, as they induce characteristic distortions in the light curves as well as timing variations. Photometric observations might also reveal the presence of prominent Saturn-like ring systems. Large moons have been hypothesized to stabilize a planet's obliquity and improve climate stability (Williams & Kasting 1996).

Planetary system architecture:

For a full understanding of the conditions on the surface of an exoplanet one must also take into account the context provided by the whole planetary system. The dynamical stability of the orbital parameters and the obliquity, as well as impact rate and history, depend on the presence of other large bodies in the system, and may have profound implications for the planet's habitability. Ideally one should thus seek to obtain complete sets of orbital parameters for all planets in the system; the presence and distribution of interplanetary dust may also provide crucial information on these issues.

Links to the Solar System:

Exoplanet science has a strong link to Solar System research. However, one cannot use the Solar System as a blueprint for exosystems, because each system and its evolving planets has a unique history and different end products. The Solar System planets should be used as "test-cases" in exoplanet studies. They are the only planets that can be studied in situ. Data, expertise and sophisticated numerical models have been collected and developed. Therefore Solar System studies and exoplanetology will ultimately merge into the broader field of comparative planetology.

7. Strawman Concepts

From the discussion in the preceding section it is apparent that the characterization of exoplanets and the exploration of habitable conditions on them comprise a very rich diversity of specific questions that can be addressed by an equally diverse set of observational approaches. Consequently, one cannot define a single mission concept that will provide a comprehensive and definitive picture of the habitable worlds in the solar neighborhood. In this Section, we will therefore describe a number of very different concepts that approach the quest for habitable worlds with common objectives based on different observing techniques in different parameter ranges. Each concept is capable of advancing the field of exoplanet exploration in a very significant way, as each can discriminate between hostile and potentially habitable worlds. We will thus have a choice between these concepts that can be based primarily on technological maturity and financial considerations later in the mission definition process.

7.1. Viable Mission Architectures

The biggest challenge in exoplanet exploration is the enormous contrast between the planet itself and its host star. There are two basic approaches to distinguishing planetary and stellar photons: spatial and temporal separation.

An instrument that aims at spatially resolving the planet from its host stars needs to provide sufficient angular resolution, i.e. of order $0.1''$ for a habitable-zone planet at a distance of 10pc. At visible wavelengths, this corresponds roughly to the resolution limit of a meter-sized telescope; in the thermal infrared an interferometer is needed to keep the unit telescope size reasonable. Working at very high contrast means that the starlight has to be rejected efficiently with a coronagraph or nuller, and this in turn requires extremely precise control of the wavefront.

A variation on the concept of coronagraphic imaging is the idea of placing an occulter in front of the telescope, blocking the starlight even before it can enter the optical system. This obviates the need for precision wavefront control and decouples the inner working angle (IWA, the minimum accessible angular separation between star and planet) from the telescope size, but requires a complicated mission scenario with telescope and occulter spacecraft separated by thousands of km.

Concepts relying on temporal separation between star and planet obtain time series of the integrated light of the system. If the orientation of the observed system is nearly edge on, transits (when the planet is in front of the star) and eclipses (with the planet behind the star) lead to a dimming whose wavelength dependence contains information on the planetary spectrum. Even if a planet does not transit, its contribution to the integrated light varies with the orbital phase. Instruments aiming at detecting the ensuing changes of intensity, polarization, or wavelength of absorption lines are conceptually relatively simple, but require extremely high signal-to-noise and excellent stability.

7.2. Coronagraphs

High contrast imaging has been intensively developed in the last decade in particular for ground-based instruments like SPHERE and GPI (currently under construction for the ESO VLT and Gemini, respectively), but also for space with JWST. The fundamental challenge is obtaining high contrast at a small inner working angle (measured in multiples of the diffraction limit, λ/D). A lot of effort has been put into manufacturing various sorts of coronagraphs (apodization, phase masks, shaped pupils, among others; see Guyon et al. 2006 for a description). Most of these designs have been tested in the lab and some are currently implemented in real instruments. When used with good optics or with adaptive optics, these devices are able to deliver contrasts of 10^4 to 10^5 at a fraction of an arcsecond from the star, and soon planet finder instruments will reach even higher contrast (10^6 to 10^8) on the ground, which is sufficient for science programs focusing on young planets. The realm of mature planets, giants as well as terrestrial, will require even larger contrasts: 10^9 - 10^{10} in the visible and near-IR at closer separations of only a few times the telescope diffraction limit (corresponding to ~ 1 AU at 10-20pc). This means that a capability to suppress the starlight with an additional factor of 100 to 1000 with respect to SPHERE and GPI is needed. Such a challenging objective cannot be simply attained from an extrapolation of SPHERE and GPI, and hence calls for the development of new instrumental concepts and new strategies. In this respect, lab experiments have been built both in US and Europe to tackle coronagraphy and wavefront control with large spectral band passes (Trauger & Traub 2007, Guyon et al. 2012, Baudoz et al. 2012). The achieved contrasts are very close to the requirements but still require some efforts to increase achromaticity and performance at the system level. The key points that are at the focus of current research are: 1) the capacity to control the optical wavefront in real time along the whole optical path to the science image in both phase and amplitude, 2) the achromaticity of the coronagraph, and 3) the ability to recover the planetary signal embedded in the residual stellar light.

Moving from ground to space, in 2002 the NASA Terrestrial Planet Finder project studied a large coronagraphic telescope (6 to 8m in size) whose objective was to detect and characterize Earth twins in the solar neighborhood. Although abandoned in 2006, TPF-C has identified the key technologies for a space-based coronagraph (Quirrenbach 2005). Smaller, 2-4m class coronagraphic telescopes have also been studied, and SEE-COAST and SPICES (with a 1.5m telescope) were proposed as an M class mission for Cosmic Vision (Schneider et al. 2009, Boccaletti et al. 2012). Maire et al. (2012) investigated the astrophysical potential of such a small coronagraph in the context of exoplanets (Fig. 4). It can take spectra in the visible (0.45-0.90 μm) of mature planets from about 1 to 10 AU (depending on planet size and stellar spectral type) in the solar vicinity ($< 20\text{pc}$). A few Super earths (about $2.5 R_{\oplus}$), the most challenging targets, could be observed if present around the nearest targets (4-5 pc). To be efficient, such a mission needs an input catalog of targets, which will be provided by radial velocity surveys from the ground as well as astrometry with GAIA. Surprisingly, a 1.5m telescope with 10^{10} contrast capability can even perform low resolution spectroscopy of an Earth twin planet around the nearest star (about 1 pc).

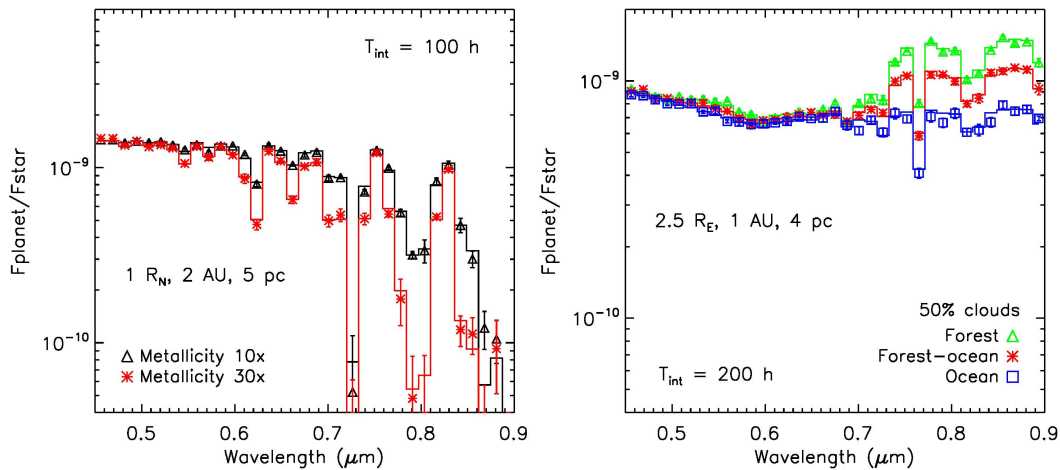


Figure 4: Simulated spectra across the SPICES spectral range, of Neptune-like (left) and Earth-like planets (right) with various properties demonstrating the ability of comparative planetology: spectral features vary at a detectable level with planet properties (Maire et al. 2012).

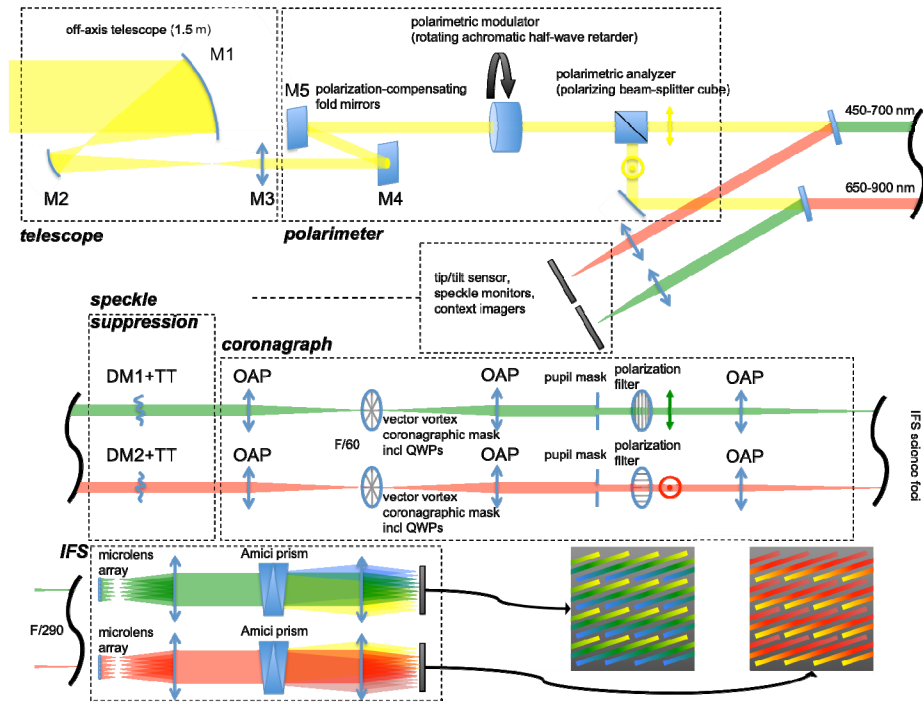


Figure 5: Conceptual design of the SPICES payload showing the main blocks: telescope, polarimeter, coronagraph and IFS. Only the main optics are shown here for sake of clarity.

To extend the mission capabilities to cover the appealing science case of spectroscopy of terrestrial planet atmospheres, beyond the marginal access provided by SPICES, a larger telescope (but with the same contrast capability) is required. A telescope diameter of 2-3 m is suitable for the science objectives proposed in this White Paper and is compatible with an ESA L class mission. In addition, an extension further into the near-IR is desired, although this imposes stronger limits on the angular resolution (equivalent to about 1 AU at 10 pc in the near-IR). With a suitable system design, it may even be possible to include the capability of sub-microsecond astrometry, and thus to measure the planetary orbit (Guyon et al. 2013).

The concept of SPICES (for details see Boccaletti et al. 2012; see also Fig. 5) can be considered as a starting point for a study of an L class mission and is described here to provide an example of the present state of coronagraphic concepts and technologies. SPICES is a 1.5-m off-axis telescope consisting of a coronagraphic system combined with wavefront correction, which feeds an integral field spectro-polarimeter covering the 450-900 nm band and measuring two linear polarizations (Stokes Q, U). The requirement on the telescope optical quality is not drastic, but 10nm rms must be achieved at mid-frequencies. The instrument consists of two channels to mitigate the problem of chromaticity. Each channel is assigned a direction of polarization and half of the spectral band while it contains a single deformable mirror and a polarimetry-compliant coronagraph. Wavefront sensing can be performed with a variety of techniques, such as the Self-Coherent Camera (Galicher et al. 2010), which also provides discrimination between speckles and planets (based on coherence) as a second stage (in the post processing). A deformable mirror (DM) provides a wavefront quality and stability on the order of tens of picometers. A Xinetics 48x48 actuator DM component has been tested at TRL5 by JPL (Trauger et al. 2010) and meets the contrast requirement. The coherent light is suppressed by a Vector Vortex Coronagraph, a derivation of the phase mask concept which can be made potentially achromatic with a 50% bandwidth (Mawet et al. 2009). A raw contrast of $\sim 10^9$ over 20% bandwidth has been demonstrated at JPL (Trauger et al. 2011, Mawet et al. 2011). The backend instrument is a microlens-based integral field spectrograph (IFS) similar to those being developed now for SPHERE and GPI. Polarimetry is implemented in this design by using a rotating half-wave retarder as a modulator and a polarizing beam-splitter cube as an analyzer. For the purpose of thermal stability, target accessibility, and high data rate for the full mission, the satellite must be on an orbit around the Sun-Earth L2 Lagrangian point.

7.3. External Occulters

A different type of starlight suppression, first proposed by Lyman Spitzer, combines a telescope and a starshade (or occulter) in space for discovery of planets (Spitzer 1960). The size of the shade and the inter-spacecraft separation were enormous and thus impractical, but over many years refinements in starshade design have reduced the required starshade dimensions and improved the level of suppression. The technology developments and mission studies for external occulters have mainly been done in the US; a serious effort to build up similar expertise will be required in European academia and industry.

Though the original concepts for the starshade have used transmitting sheets with graded transmission for apodization, the most recent work has focused on optimizing the shapes of serrated-edge binary masks. Petal shapes have been found (e.g. Vanderbei et al. 2007) that permit operation at IWA < 100 milli-arcsec at wavelengths from 0.5-1.1 μm , using a shade with a nominal diameter of 40 meters at a telescope-starshade separation of 40,000 km. The telescope can be an ordinary diffraction-limited space telescope, and its diameter is determined mainly by the integration time required to detect faint planets ($\propto D^{-4}$), and by the need for planet-star astrometry ($\propto \lambda / D$). It can observe the planet in the entire passband from 0.5 to 1.1 μm in a single integration. Slewing from one star to the next requires that the starshade travel several thousand kilometers. To accomplish this within a few weeks requires large starshade velocities and Δ -velocities. With conventional thrusters, this would take a hefty amount of fuel; advanced electric propulsion eliminates this concern, but requires substantial electrical power. A substantial engineering effort has been dedicated to minimizing the time between observations and the resources required, and some mission scenarios have been found that yield satisfactory efficiencies with one occulter, and much better with two occulters. For modest telescope sizes (up to $\sim 4\text{m}$), these mission concepts outperform internal coronagraphs in terms of the number of planets that can be observed, as the smaller IWA more than compensates for the poorer agility. External occulters are particularly suited for long integrations on a small number of “cornerstone systems” that can be studied in exquisite detail, including monitoring of seasonal changes.

The most difficult technological issue is ensuring that the edge shape of the occulter is made well enough and maintained that way. Managing scatter and diffraction of sunlight off the edges, and deployment of the large starshade also need to be addressed. A NASA-funded technology program has demonstrated the manufacture of an occulter with flight quality edges, and a current one is intended to demonstrate deployment. Conventional prelaunch end-to-end testing – i.e., demonstrating stellar suppression at typical mission distances – is impossible. Thus it will be necessary to rely on diffraction models validated by subscale testing.

7.4. Interferometers

A space-based interferometer with starlight rejection capabilities – i.e. nulling (Bracewell 1978, Angel & Woolf 1997) – offers simultaneously the sensitivity, angular resolution and dynamic range needed to isolate and spectroscopically characterize the light of an exo-Earth in the $\sim 6\text{--}20\mu\text{m}$ mid-infrared spectral domain.

As the faint planetary signal needs to be disentangled from the bright stellar one, the system must be spatially resolved typically at the 50–200 mas level. A space-based nulling interferometer is able to spatially resolve and discern the faint planetary photons from the 10^6 times brighter stellar flux, as well as from spurious sources like stellar leaks (due to resolved stellar disk), our own local Zodi-
acal cloud, the exozodiacal light, and the thermal emission produced by the instrument. Luckily, the mid-IR range is also where the otherwise huge flux contrast of the system is reduced.

A 10-year long activity on both sides of the Atlantic to select the optimal array geometry converged in 2005 into the so-called *Emma X-array* configuration. The baseline concept is an X-shape configuration of four 2-m collectors flying in formation at L2 over a 5-year duration. The beams are combined within an additional centrally positioned spacecraft, where destructive interference cancels out the light from the central star. The long and short baselines of the rectangular configuration are tunable from tens to hundreds of meters in order to uniquely optimize the transmission map of the interferometer to the size of the habitable zone, which directly depends on a given stellar spectral type.

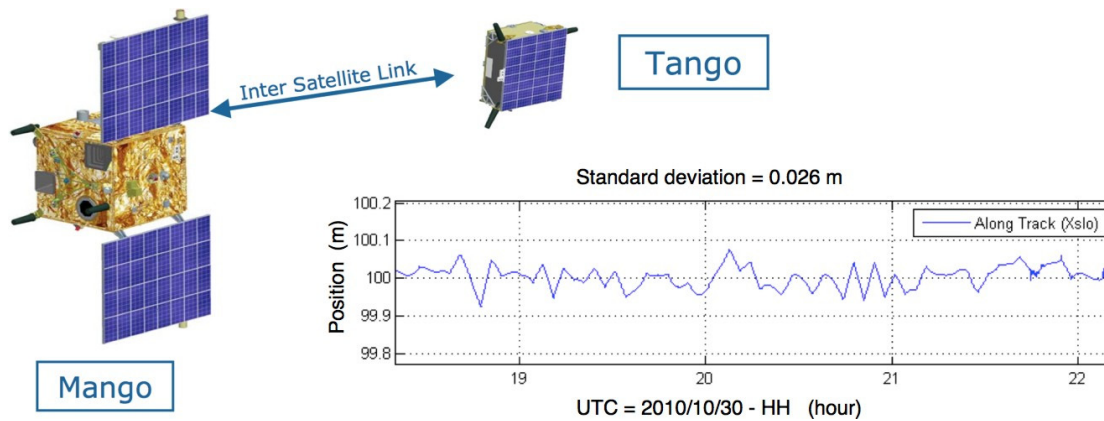


Figure 6: Demonstration of Formation Flying by PRISMA, 30 Oct. 2010. Two spacecraft, Tango and Mango, were successfully maintained at a distance of 100 m during 4 hours. The standard deviations were a few cm, limited by the accuracy of the radio frequency sensors. Laser based sensors will reduce this error to circa 100 μm (ESA's Proba 3 project). © Swedish Space Corporation, CNES and DLR

In the X-array arrangement, the respective destructive outputs of the two short-baseline Bracewell interferometers are combined with opposite phase shifts ($\pm 90^\circ$). This results in an internal “phase chopping” process (Mennesson 2005), which efficiently removes the thermal background and any emission from centro-symmetric sources around the nulled star.

A large effort in Europe has developed extensive expertise on nulling interferometry, both in academic and industrial centers. This has led to the publication of 34 PhD theses and 40 refereed papers. Built on the strong heritage of the Darwin/TPF studies (Cockell et al. 2009), progress has been achieved in various key technological areas, giving additional credit and technical readiness to this instrumental approach in the horizon of an L3 launch in 2034.

A key aspect is the deployment of a space interferometer based on a distributed array involving formation-flying operation. With the successful launch in 2010 of the PRISMA mission (OHB (S), DLR (D), CNES (F), DTU (DK), CDTI (E); Fig. 6) a crucial step has been made with the validation of the “Optical Arm” building block composed of two free-flying units, whose shape (length, orientation, rigidity) is controlled by the GNC/AOCS system. Extending the flight-tested building-block functionality from a distributed 2-S/C instrument to a 5-S/C instrument (i.e. 4 “optical arms” around one A-Unit) mainly relies on the replication of the coordination functionality and does not present additional complexity in terms of procedures according to the PRISMA navigation team. The current positioning accuracy is sub-cm, limited by the metrology system (GPS and RF). The launch of the PROBA-3 mission in 2017 will provide further valuable free-flyer positioning accuracy results (sub-mm). However, it should be noted that the requirement on the S/C positioning for interferometry is only at the sub-cm level, as the additional accuracy for co-phasing the array is provided by nanometer-accuracy servo delay-lines with few cm stroke, as demonstrated by TNO-TPD (NL).

The Planet Detection Testbed (Martin et al. 2010, 2012) has demonstrated the deep nulling needed for the detection and spectroscopy of Earth-mass planets. At $10\mu\text{m}$ with 10% bandwidth, it has achieved nulling of 8×10^{-6} (the flight requirement is 10^{-5}), starlight suppression of 10^{-8} , and planet detection at a planet-to-star contrast of 2×10^{-7} , the Earth-Sun contrast. The *phase chopping* technique (Mennesson 2005) has also been implemented and validated on-sky for the Keck Nuller Interferometer (Colavita 2009).

In parallel, the operation of ground-based interferometers such as the VLTI has permitted to develop a strong European competence in the field of fringe sensing, tracking and stabilization.

7.5. Alternative Concepts

The recent stunning progress in the field of exoplanet detection and characterization has been mostly due to the exploitation of temporal rather than spatial differencing (see Section 4.), with Spitzer, HST, Kepler, and from the ground. It is particularly remarkable that many new “tricks” have been invented and put into practice that only ten years ago were not considered feasible: analyses

of secondary transits (eclipses), out-of-eclipse light curves (phase curves), ground-based detections of molecular bands, and inspired applications of transit timing variations, to name just a few. CHEOPS, TESS, JWST, and hopefully EChO and/or PLATO will build on this legacy and offer many opportunities for the development of new strategies, perhaps even including studies of habitable super-Earths around bright M dwarfs. While it currently appears that the thorough characterization of habitable planets will likely require spatial rejection of the starlight to reach the required contrast, innovative temporal-differencing concepts might become serious alternatives in the coming years, on time for implementation as the L3 mission.

8. Sketch of Possible Implementation Plan

Considering that the scope of the present solicitation encompasses two decades from adoption of the science theme to envisaged launch date, care has to be taken to plan for technology development, and to maintain sufficient flexibility for identifying the mission concept most suitable for attaining the science goals, within the budgetary and technological constraints. For this decision, the proper yardstick clearly is not technology ready for a flight project today, but technology that can be brought to sufficient maturity within one decade. The existence of several viable concepts using very different observing strategies and therefore very different basic technologies should thus be viewed as a strength rather than a weakness, as several years are still available for technology development, risk mitigation, and to match the cost of each concept to the L mission envelope.

While considerably detailed preliminary studies have already been carried out on some of the concepts discussed above, much less has been done on others. In particular, careful system-level industry studies of coronagraphic missions have not yet been carried out in Europe. Moreover, even the more exhaustive studies involving industry were done ~10 years ago, and under the assumption of a much shorter time line available for technology development than currently envisaged. It is thus necessary to take a fresh and uniform look at the system level, to identify the key enabling technologies and the path to advance their maturity, before a sound decision can be made about the concept selection.

Assuming that the exoplanet theme is selected in 2014 for a 2034 L3 mission, a tentative timeline leading to the mission definition could therefore appear as follows:

2014-2015: Initial system-level assessment study for each concept family, identifying key enabling technologies which need maturity enhancement

2016-2020: Intensive R&D program to boost critical key technologies to a TRL level > 5; industry studies including cost estimates where needed

2019: Call for proposals for L3 mission

2020: Selection of mission concept for implementation

2020-2024: Further technology advancement to flight readiness, main trades and system analysis for the selected mission

2026-2034: Phase B

2034: L3 Launch

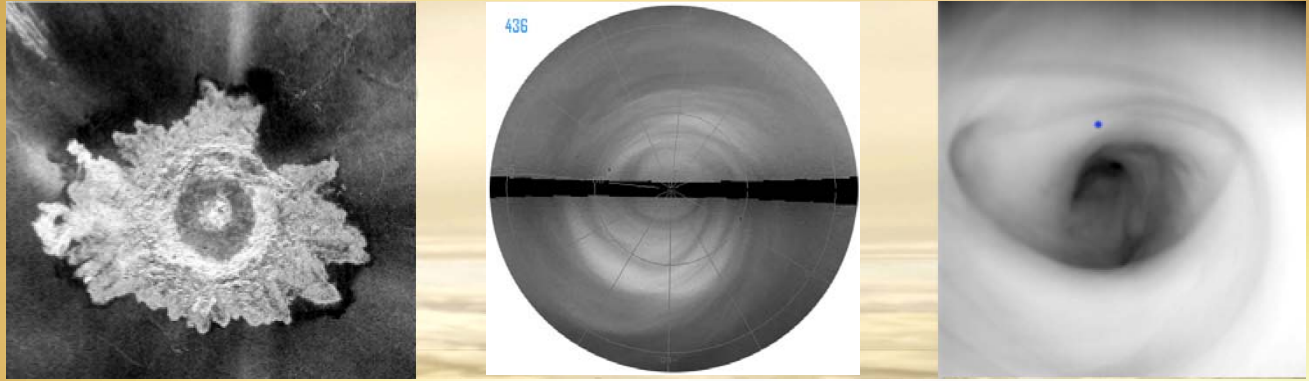
Preparatory science programs will proceed in parallel with these technical activities. As explained in Section 5., programs to search for target planets in the solar neighborhood and to characterize exozodiacal dust disks will proceed largely independently of the L3 mission anyway. However, the adoption of exoplanet exploration as a Large Mission by ESA will provide an added incentive for scientists and funding agencies to intensify their research efforts in these areas.

With the discovery of exoplanets, fiction and dream have become science. Observational exoplanetology has developed with extraordinary rapidity, and will continue to do so, attracting the efforts of our brightest minds, producing results which change the way we perceive the universe and ourselves. It is now realistic to address the great questions about habitable worlds other than Earth. Those questions will never be answered without instruments such as described here. A strong program of exoplanet missions is an essential component of the path ahead.

9. References

- Adams, E.R., et al., 2008, ApJ 673, 1160
Angel, J.R.P. & Woolf, N.J., 1997, ApJ 475, 373
Batalha, N.M., et al., 2013, ApJS 204, 24
Baudoz, P., et al., 2012, SPIE Conf. Series, 84468C
Bean, J.L., et al., 2010, Nature 468, 669
Bean, J.L., et al., 2011, ApJ 743, 92
Berta, Z.K., et al., 2012, ApJ 747, 35
Beaulieu, J.P., et al., 2008, ApJ 677, 1343
Beaulieu, J.P., et al., 2011, ApJ 731, 16
Beuzit, J.-L., et al., 2008, SPIE Conf. Series, 701418
Boccaletti, A., et al., 2012, Exp. Astron. 34, 355
Bonnetfoy, M., et al., 2011, A&A , 528, L15
Bonnetfoy, M., et al., 2013, arXiv:1302.1160
Bracewell, R.N., 1978, Nature 274, 780
Brogi, M., et al., 2012, Nature 486, 502
Burrows, A., et al., 2005, ApJ 625, L135
Charbonneau, D., et al., 2009, Nature 462, 891
Cockell, C., et al., 2009, Astrobiology 9, 1
Colavita, M.M., et al., 2009, PASP 121, 1120
Croll, B., et al., 2011, ApJ 736, 78
Crossfield, I.J.M., et al., 2011, ApJ 736, 132
Des Marais, D.J., et al., 2002, Astrobiology 2, 153
Désert, J.M., et al., 2009, ApJ 699, 478
Ehrenreich, D., et al., 2007, ApJL 668, L179
Fortney, J.J., et al., 2006, ApJ 652, 746
Fressin, F., et al., 2013, ApJ 766, 81
Fujii, Y. et al., 2010, ApJ 715, 866
Galicher R., et al., 2010, A&A 509, A31
Gibson, N.P., et al., 2011, MNRAS 411, 2199
Gillon, M., et al., 2007, A&A 471, L51
Grasset, O., et al., 2013, Plan. & Space Sci. 78, 1
Grenfell, J. et al., 2007, Plan. & Space Sci. 55, 661
Guyon, O., et al., 2006, ApJS 167, 81
Guyon, O., et al., 2012, SPIE Conf. Series, 84424V
Guyon, O., et al., 2013, ApJ 767, 11
Hansen, J.E. & Hovenier, J.W., 1974, J. Atmos. Sci. 31, 1137
Heller, R., 2012, A&A 545, L8
Janson, M., et al., 2010, ApJL 710, L35
Hedelt, P., et al., 2013, A&A 553, A9
Kalas, P., et al., 2008, Science 322, 1345
Kalas, P., et al., 2013, ApJ submitted, arXiv:1305.2222
Knutson, H., et al., 2011, ApJ 735, 27
Konopacky, Q.M., et al., 2013, Science 339, 1398
Lagrange, A.M., et al., 2009, A&A 493, L21

Lagrange, A.M., et al., 2010, *Science*, 329, 57
 Lagrange, A.M., et al., 2012, *A&A*, 542, A40
 Léger, A., et al., 2009, *A&A* 506, 287
 Line, M., et al., 2011, *ApJ* 738, 32
 Macintosh, B.A., et al., 2008, *SPIE Conf. Series*, 701518
 Maire, A.L., et al., 2012, *A&A* 541, 83
 Marois, C., et al., 2008, *Science* 322, 1348
 Marois, C., et al., 2010, *Nature* 468, 1080
 Martin, S., et al., 2010, *A&A* 520, A96
 Martin, S., et al., 2012, *Appl. Opt.* 51, 3907
 Mawet, D., et al., 2009, *Optics Express* 17, 1902
 Mawet, D., et al., 2011, *SPIE Conf. Series*, 81511D
 Mayor, M. & Queloz, D., 1995, *Nature* 378, 355
 Medvedev, A.S., et al., 2013, *Icarus* 225, 228
 Mennesson, B. et al., 2005, *Icarus* 178, 570
 Mouillet, D., et al., 1997, *MNRAS* 292, 896
 Nutzman, P., et al., 2009, *ApJ* 692, 229
 Oppenheimer, B.R., et al., 2013, *ApJ* 768, 24
 Plaut, J.J., et al., 2007, *Science* 316, 92
 Quirrenbach, A. (Ed.), 2005, *ESA WPP-245*, arXiv:astro-ph/0502254
 Rauer, H. et al., 2011, *A&A* 529, A8
 Richardson, L.J., et al., 2006, *ApJ* 649, 1046
 Robinson, T., et al., 2010, *ApJL* 721, L67
 Schneider J., et al., 2009, *Exp. Astron.* 23, 357
 Seager, S., et al., 2005, *ApJ* 632, 1122
 Sing, D.K., et al., 2009, 505, 891
 Snellen, I.A.G., et al., 2008, *A&A* 487, 357
 Snellen, I.A.G., et al., 2013, *ApJ* 764, 182
 Sohl, F., et al., 2012, *Proceedings IAUS* 293, arXiv:1211.3331
 Solomon, S.C., et al., 2005, *Science* 307, 1214
 Spitzer, L., 1960, *AJ* 65, 242
 Stevenson, K., et al., 2010, *Nature* 464, 1161
 Swain, M.R., et al., 2008, *Nature* 452, 329
 Swain, M.R., et al., 2009, *ApJ* 704, 1616
 Tinetti, G., et al., 2007, *Nature* 448, 169
 Trauger, J., & Traub, W., 2007, *Nature* 446, 771
 Trauger, J., et al., 2010, *SPIE Conf. Series*, 773128
 Trauger, J., et al., 2011, *SPIE Conf. Series*, 81510G
 Vanderbei, R.J., et al., 2007, *ApJ* 665, 794
 Williams, D., & Kasting, J., 1996, *LPI* 27, 1437
 Wyatt, M.C., 2003, *ApJ* 598, 1321



Venus: A Natural Planetary Laboratory

Case for ESA's Cosmic Vision L2/L3 Mission to Venus
to Learn about Other Terrestrial ExoPlanets

Spokesperson:

Sanjay S. Limaye

University of Wisconsin

Space Science and Engineering Center

1225 West Dayton Street, Madison, Wisconsin 53706, USA

+1 608 262 9541 (Work)

+1608 262 5974 (FAX)

Sanjay.Limaye@ssec.wisc.edu



Cover: Of all the planets, Venus exhibits the largest vortices relative to its size, one in each hemisphere, centered over respective poles. The southern hemisphere of Venus as seen from Venus Express by the Venus Monitoring Camera in the ultraviolet filter a composite of images from two orbits (center). VIRTIS view of the vortex core in near infrared (right) and crater imaged with radar by Magellan (left)

Authors

#	Name	Institution	E-mail
1	Sanjay Limaye¹	U. of Wisconsin, Madison, WI, USA	Sanjay.Limaye@ssec.wisc.edu
2	James A. Cutts	Jet Propulsion Laboratory, Pasadena, California, USA	james.a.cutts@jpl.nasa.gov
3	Kristen Griffen	Northrop Grumman Aerospace Corp. El Segundo, CA, USA	Kristen.Griffin@ngc.com
4	Ron Polidan	Northrop Grumman Aerospace Corp. El Segundo, CA, USA	ron.polidan@ngc.com
5	Alexander Rodin	Moscow Institute of Physics and Technology, Moscow, Russia	Alexander.rodin@phystech.edu
6	Suzanne A. Smrekar	Jet Propulsion Laboratory, Pasadena, California, USA	suzanne.e.smrekar@jpl.nasa.gov
7	Manuela Sornig	Rheinisches Institut für Umweltforschung an der Universität zu Köln, Germany	sornig@ph1.uni-koeln.de
8	Ann Carine Vandaele	Belgian Institute for Space Aeronomy, Brussels, Belgium	ac.vandaele@aeronomie.be
9	Thomas Widemann	Paris Observatory, Paris, France	thomas.widemann@obspm.fr

Endorsed By:

#	Authors	Institution	E-mail
1	Ludmila Zasova	IKI, Moscow, Russia	zasova@iki.rssi.ru
2	Eric Chassefiere	IDES - Interactions et Dynamique des Environnements de Surface, Orsay, France	eric.chassefiere@u-psud.fr
3	Valerie Wilquet	Belgian Institute for Space Aeronomy, Brussels, Belgium	valerie.wilquet@aeronomie.be
4	Alessandra Migliorini	IAPS-INAF, Rome, Italy	alessandra.migliorini@iasf-roma.inaf.it
5	Veronique Dehant	Royal Observatory of Belgium	veronique.dehant@oma.be
6	Giuseppe Piccioni	INAF-IAPS, Rome, Italy	giuseppe.piccioni@iaps.inaf.it
7	Oleg Korabev	IKI, Moscow, Russia	korab@iki.rssi.ru
8	Maria Luisa Moriconi	IAPS-INAF, Rome, Italy	m.moriconi@isac.cnr.it
9	Romolo Politi	IAPS-INAF, Rome, Italy	romolo.politi@iaps.inaf.it
10	Arnaud Mahieux	Belgian Institute for Space Aeronomy, Brussels, Belgium	arnaud.mahieux@aeronomie.be
11	Ernesto Palomba	IAPS-INAF, Rome, Italy	Ernesto.palomba@iaps.inaf.it
12	Andrea Longobardo	IAPS-INAF, Rome, Italy	andrea.longobardo@iaps.inaf.it

¹ on behalf of the COSPAR International Venus Exploration Working Group

1. INTRODUCTION

1.1 - Why Venus?

We present the case for an ESA mission to lead an internationally coordinated effort to increase the current understanding of Venus, a natural laboratory for understanding terrestrial planets not only in our solar system, but those around other stars. Venus, which has served as a useful target since the beginning of the age of space exploration, is still proving to be very mysterious, even more so after more than half a solar cycle of monitoring by Venus Express mission - spinning slowly and backwards, shrouded in a perpetual thick cloud cover obscuring its surface and its history, Venus invites more comprehensive exploration. As close to a simple, ideal atmosphere one can conceive for meteorological studies – devoid of seasons, oceans, albedo contrasts, hydrologic cycle and vegetation and evolved life, the superrotation of Venus is still a puzzle.

Venus is accessible with shortest cruise to any planet which makes its exploration very cost effective. With improving observational capabilities from large Earth-based telescopes, we're on the verge of learning more about exoplanets in greater detail. By exploring the mysteries of our neighbor Venus, we will be better equipped to understand the diversity of the exoplanets and perhaps learn whether life exists elsewhere. A Cosmic Vision mission to Venus will help solve the riddle that is Venus and further our understanding of rocky exoplanets.

In recent years, spurred by the success of ESA's Venus Express mission, an International Venus Exploration Working Group (IVEWG) was formed during the COSPAR assembly in July 2012 following discussions among the international participants at the recent meetings of the Venus Exploration Analysis Group (www.lpi.usra.edu/vexag). The IVEWG has discussed Venus exploration issues and the present paper is illustrative of these discussions. We anticipate that the Cosmic Vision mission to Venus will lead to an internationally coordinated effort to understand the planet which also serves as proxy for many rocky exoplanets discovered or to be discovered.

1.2 - Major Scientific Questions for Understanding Venus and the Exoplanets

A common thread for Venus and Mars is that the atmospheres on both planets appear to have undergone catastrophic change—Mars may have lost almost all of its atmosphere, while Venus may have driven off much of the water in a runaway greenhouse and perhaps increased its atmosphere. While the atmospheres of Mars and Venus are thus linked by dramatic change, understanding Venus' current and past climate is more critical to understanding our own. A major part of understanding how Venus evolved as a terrestrial planet is its thermal evolution. Venus seems to have outgassed fewer volatiles from its interior than Earth, despite their similar heat budgets. How is the interior evolution linked to the atmospheric evolution on a planet that is active but lacks plate tectonics? However, Venus exploration has largely been a single agency effort, with some international collaboration, and it has been difficult to support a dedicated Venus exploration program within NASA, something that only the Soviet Union carried out until the VeGa missions.

With the emphasis placed in scientific research on learning about the origin of life on Earth and elsewhere, Venus provides an accessible, crucial laboratory for hypothesis testing that is attainable with modest resources as compared to planets in the far reaches of the solar system.

Three key science objectives provide the basis for exploring Venus:

- *Origin and evolution of planets*
- *Past habitability of Venus, present climate*
- *Sun's influence on Venus and a role model for rocky exoplanets*

These objectives require a mix of observing platforms described below.

1.3 - Observations and Observing Platforms

- *Comprehensive measurements of the global atmospheric circulation, clouds, trace species and thermal structure*
 - *Multiple orbiters in polar, inclined and equatorial orbits with remote sensing (imaging spectrometers) and radio science (radiometers) instruments*
 - *Long lived, multi-level floating platforms, descent probes and landed platforms*
- *Understand planetary thermal and tectonic evolution and sources of volatiles*
 - *High resolution SAR imaging and topography, IR spectroscopy from orbit*
 - *imaging and spectroscopy from floating and landed platforms*
- *Sun's influence on planets as proxy for exoplanetary climates*
 - *Long-term monitoring of incoming dust, solar wind from orbit*
 - *Deposition of solar energy in the cloud layer and changes in the cloud cover from orbit and multilevel floating platforms*
- *High resolution map of the global gravity field*
 - *A single orbiter (e.g. GOCE) or two orbiters (e.g. GRAIL and GRACE)*

1.4 - International Collaborations and Coordination of Missions to Venus

The key science questions about Venus and the observations which can be made from a mission under ESA's Cosmic Vision are described below. It is also very pertinent to point out that the selection of such a mission will also generate significant interest among other space agencies to undertake and coordinate efforts to explore Venus. It is being slowly recognized that a major comprehensive effort to explore Venus is not feasible under projections of budgets for exploration in the near future within a period less than the productive life of a typical scientist from conception, development, implementation and analysis of the data to answer the science questions. Hence, coordinated international efforts are essential and such international collaborations have been shown to be productive from the Venus Express experience. It has enabled community groups such as VEXAG and the COSPAR International Venus Exploration Working Group, which have facilitated a grassroots effort to communicate with space-faring agencies.

Venus is also appealing because of rapid return on the investment. Cruise to Venus can be accomplished in about six months (compared to more than eight years for JUICE to reach Jupiter, and more than eighteen years from selection to arriving at target). The large investment of a planetary mission cannot pay off until the spacecraft reaches the planet—and in the case of Venus, the short cruise stage enables greater continuity of workforce expertise and makes a more immediate impact. This has an important significance for the highly trained exploration work force every where in terms of the huge investment in training, the need for their availability in the future and of course their careers.



Figure 1. Why Venus would be a perfect target for L-class mission, the most effective approaches and the most important scientific questions which should be addressed.

2. THE SCIENCE CASE FOR EXPLORING VENUS

The international Venus community has been meeting periodically at VEXAG forums to develop a strategy for Venus exploration. The Goals, Objectives and Investigations deliberated by the international community are freely available (www.lpi.usra.edu/vexag). The essence of these deliberations was the basis of the Venus flagship mission commissioned by NASA in 2009 and again in 2011 for the US academies decadal survey of planetary science.

2.1 - Origin and Evolution Venus and Relevance to Exoplanets

2.1.1 - Evolution of the atmosphere and interior

Previous measurements of the lower atmosphere (below clouds) D/H ratio on Venus as compared to Earth have been interpreted as indicating presence of liquid water on the surface of Venus early in its history (Donahue et al., 1982; de Bergh et al., 1991; Hunten, 1992), thus implying a more hospitable environment for life. Furthermore, the D/H ratio from measurements of Venus Express (VEX) is 120 times higher than for Earth below the clouds and clouds and >200 times higher than Earth's D/H above the cloud tops (Fedorova et al., 2008). This suggests that the Venus' atmosphere is still being lost, and thus still evolving.

A major observational missing link in our understanding of Venus' climate evolution is the distribution of noble gases and their stable isotopes in Venus' atmosphere. The concentrations of heavy noble gases (Kr, Xe) and their isotopes are mostly unknown, and our knowledge of light noble gases and stable isotopes is incomplete and inaccurate. Chassefiere et al. (2012) list the following questions that noble gas and isotope measurements would help to answer:

- Did Venus suffer a similar strong early atmospheric loss as documented for Earth and Mars with radiogenic Xe isotopes?
- What is the He residence time in the atmosphere, and what fraction of atmospheric ^4He is non-radiogenic?
- Is there a genetic relation between the abundance patterns of the three large terrestrial planets? What are the implications of differences and similarities of the three patterns? Did all three planets start with comparable noble gas inventories?
- What can Venus tell about Earth and Mars? Is “gas-rich” or “gas-poor” accretion more likely? What are the acquisition and loss processes? Do isotopically solar-like noble gases in Venus strengthen the inference that solar nebula gases were also acquired by Earth and Mars? What is the fraction of primordial ^3He escaping at present from Venus?

Such observations would help in answering the following fundamental questions:

- For how long has Venus been in its current extreme climate state?
 - When and how did it diverge from a (possible) early Earth-like state?
 - Has Venus been a potentially habitable planet at some time during its early history?
 - Did a “cool early Venus” stage occur between the end of accretion and the late heavy bombardment, as suspected for Earth (Valley et al., 2002)
 - What are the implications of the Venus/Earth comparison for the nature and evolution of habitable terrestrial planets throughout the universe?
-

2.1.2 - Evolution of Venus' interior, surface and rotation state and links to past climate

Measurements of Ar isotopes in the atmosphere of Venus indicate that it has outgassed roughly 25% of its interior volatiles. In comparison, Earth has outgassed approximately 50% and Mars only 3%. Why has Venus retained so much more of its volatile budget? How and when were the 25% released? Understanding how Venus' climate has reached its current state can not be understood without also understanding how the geologic evolution of Venus and Earth diverged. The surface of Venus is comparatively young – 300-1000 m.y. How did Venus wipe out or cover up its past history (large impact basins? Plate tectonics?) so rapidly? What are the implications for release of volatiles into the atmosphere and associated climate change? Spectra from the Venus Express Mission's Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) provide surface brightness data, from which emissivity can be derived. (Helbert et al., 2008; Mueller et al., 2008). High emissivity anomalies at sites previously identified as sites of likely upwelling mantle plume on the basis of their geology, topography, and gravity, were interpreted to indicate relatively recent, unweathered basalts (Smrekar et al., 2010).

High-resolution measurements of Venus' topography field should reveal the signatures of past impacts as has been seen on Mars.. High resolution imaging and topography could also provide a stratigraphy for Venus, and thus provide clues to its resurfacing history. IR spectroscopy could show evidence of large-scale recent volcanism over the entire planet, and if very lucky, reveal active volcanism.

The key to the current orbital state of Venus--its slow rotation rate and the direction of its spin vector nearly opposite to its orbital vector--is how and when its direction changed. It is generally believed that Venus started out with a higher rotation rate, with its spin axis aligned with the solar spin axis. Subsequent events, such as either collisions during planetary accretion or subsequent impacts, atmospheric drag, or core-mantle friction, may have caused it to evolve to its current state of a thick, hot atmosphere. Early impacts could have resulted in loss of its atmosphere. It is thus likely that the climate of Venus has changed dramatically in its early history as the spin period evolved from a faster one to a slower one.

One of the significant results from Venus Express is that the spin period of Venus may not be constant (Mueller et al., 2012), based on comparison with results from ground-based measurements and Magellan. Even the short-term variations of the spin rate are important indicators, as has been learned by monitoring the Earth. The International Time Union has a network of observations that provide daily changes of the rotation rate. Sudden spin rate changes have been detected from tsunamis, earthquakes, and changes in surface or near-surface mass distributions. Near-surface mass distribution changes (e.g., due to volcanism) are possible on Venus and hence it is important to monitor the changes from ground-based large radio telescopes equipped with powerful transmitters, and from future Venus orbiters.

Past changes in the spin period of Venus are more challenging to estimate, but there is one observational strategy that can provide some clues. There are two different theories about what caused Venus to change its spin. One involving interaction between solar gravity and thermal tides (Correia and Laskar, 2001, Correia et al., 2003; Correia and Laskar, 2003) and

another involving multiple impacts (Alemi and Stevenson, 2006), which were postulated to explain the current absence of a Venus moon. Caudal (2010) has investigated whether Venus' core is trapped in a 5:1 spin-orbit resonance by hypothesizing that the Venus core consists of a solid core inside a liquid core undergoing $0.31^\circ/\text{year}$ differential rotation with the mantle. This is much larger than the difference in the rotation rate estimated from Venus Express VIRTIS data ($0.154^\circ/\text{year}$).

Venus is unusual in that it has no moons. As a possible explanation, an early moon may have disintegrated either due to tidal forces or impacts, with the debris impacting Venus. High-resolution gravity data would enable a search for such debris. Subsequent in-situ measurements in several places on the planet should then provide some confirmation of such an event that could have led to changes in the spin state of Venus.

2.2 Past Habitability of Venus and Present Climate

A recent report, "Towards Understanding the Climate of Venus" (Bengtsson et al., 2013) reviews the scientific questions that still face us in realistically simulating the observed atmospheric circulation of Venus. This only emphasizes the significance of Venus exploration if we are to make meaningful progress in understanding exoplanets towards the ultimate goal of answering the question, "are we alone?"

2.2.1 - How does Venus' atmosphere work and what are the missing pieces of key information?

As we begin to discover terrestrial exoplanets orbiting other stars in our galaxy, some of them will likely be Venus-like, and learning how they reach this evolutionary state will be absolutely crucial for our understanding of the origin and longevity of habitable conditions on Earth-like planets. Pioneer Venus informed us about the past presence of water on Venus (Hunten, 1992). Its subsequent loss tells us that the history of water on Venus is even more significant for improving our capability to understand future Earth climates as the rising surface temperatures lead to increasing water vapor in the atmosphere. This in turn raises the saturation vapor pressure, the same process that is believed to have raised the surface temperature on Venus and led to the loss of its surface water (Sagan, 1960). In the last few decades the discovery of life in extreme environments on Earth's seafloor (and subsurface?) has revitalized the concept of a habitable zone. As we look for life elsewhere, it is also important to remember that the Venus clouds present a potentially habitable environment for certain bacteria (Sagan, 1971; Schulze –Makuch and Irwin, 2002; Schulze-Makuch et al., 2004). Bacteria can survive at high altitudes in Earth's atmosphere and are found to be dominantly of terrestrial origins, but bacteria are also found in cosmic dust samples collected at high altitudes (Yang et al., 2009; Juanes-Vallejo, 2011).

One of the most important unknowns for understanding the Venus atmosphere is the *identities of UV absorbers* in the upper atmosphere. Knowledge about the properties of known (e.g., gaseous SO_2) and unknown UV absorbers in the upper atmosphere are necessary to understand the impact on the cloud structure and the enigmatic dynamical behavior of Venus. This also involves the *characterization the sulfur cycle* within the Venus clouds and the cloud haze. Information about the formation processes and characteristics of the sulfuric acid aerosols (H_2SO_4) cloud particles (e.g., particle sizes and local abundances) are necessary to determine the role of cloud heating in powering

superrotation. Several relevant gaseous and liquid/solid aerosol components such as SO₂, H₂O, OCS, CO are likely involved. A better knowledge about the UV absorbers and the sulfur cycle will illuminate the coupled dynamical/chemical processes involved in the cloud formation processes of a sulfur-driven, CO₂-based terrestrial planet. Very little is known about the *structure of the local dynamical meteorology in the cloud layer*.

Venus presents a perfect opportunity for comparative planetology studies which now include the exoplanets. Grinspoon et al. (2009) draw attention to, and summarize, the important role that planetary exploration, and research with a comparative planetology focus, have played and should continue to play in our understanding of climate, and climate change, on Earth. In particular, they note that:

“Venus is Earth’s closest planetary neighbor and a near twin in terms of overall properties such as mass and size. Their bulk densities and inventories of carbon and nitrogen are similar, suggesting similar primordial volatile inventories. Mars, Earth’s next nearest neighbor, has surface conditions most closely resembling Earth’s and a wide range of meteorological and geological phenomena that are recognizable as variations on familiar terrestrial themes. Current understanding of planetary formation, volatile accretion, isotopic signatures, and the well-preserved ancient geological record of Mars all suggest that these triplet planets started out with more closely comparable surface environments, geological processes, and atmospheric compositions. Yet, despite their close proximity and similar origins, these three planets have evolved into very different states. Rotation rates, magnetic fields, surface temperatures and pressures, atmospheric inventories of radiatively active gasses, total water inventories, polar deposits, and global patterns of geological activity are among the properties that differ dramatically.

An understanding of the evolutionary histories and current states of the Venus and Mars climates is directly relevant for studies of the past, present and future climates of Earth [1]. As extreme examples of very different climate on otherwise similar and nearby planets, Venus and Mars provide opportunities to improve and validate our knowledge of planetary climate data and modeling. For example, Venus can provide a test bed for an extreme case of global warming where nonlinear effects have evidently played an important and irreversible role in climate evolution. In addition to providing an instructive and fruitful challenge for understanding the terrestrial climate, Venus also serves as a model for the long term fate of Earth’s climate, under the future influence of a warming sun. Mars has also experienced irreversible climate change from a more biologically clement surface environment, as well as a climate history where “Milankovich cycles on steroids” have resulted in a history of extreme climate variations from quasi-periodic changes in obliquity. [2] [3] [4] [5] [6].

This synergism between Venus, Earth, and Mars goes both ways: Our understanding of Venus and Mars would benefit greatly from use of the best Earth observations and models along with engagement and expertise of the larger community of Earth scientists. Although such efforts are to some degree hampered at present by limitations in the data available for Venus and Mars, much deeper understanding of these very different global climate systems should be possible, given the techniques developed to understand climate change on Earth. At the same time, these extreme cases can help to validate the crucial ability of terrestrial models to correctly predict climate on Earth forced by variations from the current atmospheric composition, increase the ability of Earth modelers to work with unforeseen climate feedbacks, and expose potential weaknesses or limitations in our current generation of Earth climate

models. One of the most vexing problems of current terrestrial climate studies is separating anthropogenic from natural signals. Our neighboring terrestrial planets provide examples, devoid of human interference, that can help us with the important work of untangling these signals.”

VEXAG goals, objectives and investigations reflect the recent advances in our knowledge of Venus from the Venus Express results.

2.2.2 - Superrotation of the Atmosphere

The bulk of the atmosphere rotates faster than the underlying planet and the driving source is still not known. The key explanations for the superrotation rely on transport of angular momentum meridionally by the mean and eddy circulations. However, this transport of angular momentum has not been observable from missions to date due to limitations of coverage and accuracy. Connections to wave activities are assumed (Lebonnois et al, 2012; Japanese group) and observations, such as with VEX (Picciali et al, 2012; Tellmann et al, 2012) indeed show extensive local and temporal changing wave activities. A consistent explanation, however, is still missing.

As described in Limaye et al. (2009), the following observations are needed to solve the riddle of the superrotation of the Venus atmosphere

- Global longitudinal and latitudinal structure of the zonal and meridional flow at a known level over at least one Venus day
- Vertical structure of the horizontal flow, which is critical to understanding Hadley cell structure
- Amplitudes and phases of solar thermal tidal winds at any level
- Reliable estimates of zonal and time-averaged latitudinal profiles of the meridional transports of absolute angular momentum, heat and trace species at any level
- Exchange of angular momentum between the atmosphere and the solid planet over time
- Middle atmosphere circulation (70-140 km) at high spatial and temporal resolution
- Structure of planetary and small scale waves and their role in the atmospheric circulation
- Structure of waves in the thermosphere such as the nine day oscillation observed in the thermosphere (Forbes and Konopliv, 2007)

2.2.3 - The Ultraviolet Absorber and Venus Clouds

The white paper submitted to the US Academies Planetary Science Decadal Survey (Limaye et al., 2011) addressed this need as follows:

“Besides the identity of the ultraviolet (UV) absorber, the next major question about the clouds is the high spatial inhomogeneity and rapid temporal variability. Why isn’t the UV absorber well mixed? The answer may lie in sulfur chemistry, and high spatial and temporal resolution observations of temperature and of characteristics of cloud/aerosols and chemical species to understand the latitudinal and temporal behavior of the UV absorbers.

The nature of the UV absorber is still a mystery and the possible existence of large and/or crystalline cloud particles has not yet been confirmed (Esposito et al., 1997). It is important to monitor the allotropes of sulfur, which have been suggested as the possible UV absorber. Monitoring of the UV contrasts through imaging from Venus Express has shown a dynamic behavior that has been suggestive of convective connections in equatorial latitudes (Titov et al., 2009). At polar latitudes a correlation between dark streaks and warmer (deeper) brightness temperatures at infrared wavelengths has been discovered from concurrent observations from VMC and VIRTIS instruments on Venus Express (Ignatiev et al., 2009), but there are puzzling exceptions as well. It has not been established whether these high-latitude contrasts are due to increased absorber, an artifact of the scattering, or both. Data from Venus Express suggest that OCS is converted to CO in the lower atmosphere around 30 km by heterogeneous chemistry. The process may produce polysulfur, a source of cloud top region UV absorber (Yung et al. 2009).

The lower cloud region is also poorly understood. Although it is likely to be composed primarily of sulfuric acid droplets, X-ray fluorescence data from the Soviet VeGa descent probes (Andreychikov, et al., 1987) found significant and spatially variable quantities of Cl, P, and Fe in the lower cloud particles, although it is not clear in what chemical form. It has not been possible to study this using remote sounding, so future in situ missions are required to investigate the complex cloud chemistry of the lower cloud deck.

Additionally, mechanisms responsible for the formation and dissipation of Venus's clouds have not been verified. For example, while a sulfur cycle involving a variety of sulfur-bearing molecules and water has been implicated, the correlation of these molecules and their daughter sulfuric cloud particles has not been measured. The nature of these photochemically produced aerosols can have significant effects on the makeup of the Venus clouds. For example, different compounds may have differing solubilities in sulfuric acid, which in the middle and lower clouds will lead to condensation of vapor onto these particles at different supersaturations. In the upper cloud, these different compounds might induce the supercooled sulfuric acid to freeze, changing its optical properties. Also, dynamical and chemical variations responsible for the spatial inhomogeneity of clouds have not been identified. In situ measurements by balloon-borne instrumentation floating for several days in the clouds should provide fundamental new information critical to clarify understanding of cloud meteorology."

2.2.4 - Cloud Structure, Thermal Structure and Radiative Balance

The need for measurements was also addressed by Limaye et al. (2011) as follows:

Venus Express has provided some new information about the Venus cloud structure. Early results indicate spatial and temporal variability in cloud properties and suggest more convective activity in equatorial latitudes, consistent with the solar heating (Satoh et al., 2009; McGouldrick et al., 2009). A combined analysis of VMC UV images and VIRTIS data indicates that the UV cloud tops are at the ~ 70 km level in low latitudes and ~ 67 km in high latitudes (Ignatiev et al., 2009). Near-infrared observations in spectral windows into the Venus atmosphere indicate significant variability in Venus' middle and lower cloud decks. Although strong downdrafts have been implicated as the source of holes in the clouds, causes for these downdrafts remain elusive, and the coupling between the microphysics, radiative balance and atmospheric dynamics is incompletely known. Furthermore, clouds play an important radiative role in both the UV/visible and infrared spectral regions. Thus, variations of cloud cover in latitude or in local solar time can both drive or be driven by the dynamics.

One of the important measurements needed is the near-surface vertical temperature structure at latitudes from equator to the poles, which impacts convective activity. The VeGa 2 lander is the only probe that has been able to measure the temperature in the lowest 12 km of the atmosphere down to the surface; it indicated superadiabatic lapse rates. Confirmation of this superadiabatic layer is critical to a better understanding of the lower atmosphere and its impact on the deep atmosphere circulation via convection and maintenance of the thermal structure (Crisp and Titov, 1997). Of particular interest is the lower atmosphere thermal structure in polar latitudes.”

Indications of atmospheric circulation variability have also been detected for the upper atmosphere (Altieri et al, 2012; Nakagawa et al., (2013) , Sornig et al., 2013), where the superrotation plays a decreasing role and a flow from the hot subsolar point to the anti-solar position arises. Despite of increasing modeling efforts, a globally consistent description of the *dynamical structure of the upper atmosphere* is still missing and additional observations are needed with the final aim to include the *interaction between the cloud layer and the upper and lower atmosphere*.

Non-hydrostatic GCM simulations (Mingalev et al., 2012, Rodin et al. 2013) suggest that thermal regime at the upper cloud boundary in the polar regions drives significant change in the global circulation pattern between superrotation and subsolar-antisolar circulation. Simulations also imply that the polar vortices (Piccioni et al. 2007) are strongly connected with the global circulation. Thus, the detailed characterization of the wind field at different scales, from local gravity waves to global patterns, is necessary to determine mechanisms maintaining the current regime of the Venus atmospheric circulation. An important aspect of Venus climate system interpretation in terms of comparative planetology and possible reference model for exoplanet observations is that, unlike other superrotation atmospheres such as Titan's, the transition from superrotation to the solar-antisolar circulation occurs in fairly narrow altitude range. Consequently the atmospheres between 70 and 110 km should be sounded more accurately by several independent methods to reach complete and reliable characterization.

2.2.5 - Present habitability

With the pervasive emphasis on the large greenhouse effect on Venus leading to its high surface temperature, it is often forgotten that Venus has a habitable zone in the cloud region of Venus. Indeed, Venus may be the only terrestrial planet that experiences the coldest and the warmest temperatures in its atmosphere 100K or lower near 125 km (Mahieux et al., 2012). Sagan (1961) suggested the notion that bacteria may be present in the clouds of Venus and the speculative idea was followed up by Morowitz and Sagan (1967) although the true nature of the clouds was not known then. More Recently Schulze-Makuch and Irwin (2002) examined whether bacteria can survive in the cloud layer of Venus and Scuulze-Makuch et al. (2004) look into how the putative life could survive in the Venus clouds using a sulfur based UV adaptive strategy.

This is an important goal for investigation especially given the growing interest in searching for biomarkers in exoplanet observations and the quest for life elsewhere. Venus

provides an easy, accessible target to test and validate the ideas for possibility of life under extreme conditions.

2.3 - Sun's Influence on Venus and a role model for rocky exoplanets

While the connection between the sun and climate may be obvious considering that the climate is controlled by the balance between the energy absorbed from the Sun and that emitted to space, the response of the climate to the solar variability is complicated and not fully understood (Lean and Rind, 1996). The NASA Living With a Star Sun-Climate Task Group (J. Eddy, Chair) noted in its report (Eddy, 2003) that “at this time we simply do not know whether longer-term climatically-significant variations in solar irradiance exist or don't exist. Nor do we know the magnitude of these conceivable changes.” Much of the difficulty is due to the different time scales characteristic of the climate markers and the solar irradiance. Other difficulties arise in terms of the spectral variability of the irradiance over time and the total solar irradiance. It is in this instance that Venus serves as a near-perfect natural laboratory – a planet with uniform cloud cover that contains heterogeneous ultraviolet absorber(s) responsible for controlling the climate. Therefore, one should expect variability in the Venus cloud cover in response to the solar output. Measurements to monitor such changes from orbit are feasible and may be simpler to some degree than for Earth.

3. NEEDED MEASUREMENTS ON VENUS

Recently, two comprehensive missions have been studied for Venus that provide some basis for potential missions to Venus suitable for being chosen for the Cosmic Vision, especially in the context of international collaboration. The Venus Design Reference Mission was commissioned by NASA with an international Science and Technology Definition Team (www.lpi.usra.edu/vexag/venusSTD/). The final report has been published (vfm.jpl.nasa.gov/). The Venus Climate Mission, a scaled back version of this mission, was recommended by the US National Academies Planetary Science Decadal Survey (2013-2022) as one of the large missions to be considered for launch by NASA within the decade². Since the completion of that study, new science questions and examination of long-standing questions and developing technology for future missions has led to some additions and modifications to the proposed implementation strategy regarding platforms and measurements.

These include obtaining very high resolution image and topography of the surface of Venus, a focus on learning the identity of the unknown ultraviolet absorber that is responsible for a dominant fraction of the absorbed solar energy in the cloud layer, and the need to obtain long term detailed global maps of the three-dimensional atmospheric circulation as is feasible for Earth, but with practical constraints, and improved information about geodesy and gravity field anomalies. The cloud cover, surface imaging, topography and gravity measurements suggest a need for two or more spacecraft in a ~ 400 km circular or near circular, polar orbit. (Table 1).

² (www.lpi.usra.edu/vexag/reports/VCM_report_FINAL.pdf)



Figure 2. (Left) Components of the Venus Flagship Design Reference Mission studied by the international team of Venus Scientists – an orbiter with InSAR, two balloons and two landers. Russia's Venera-D mission under study has similar components but with a sub-orbiter for bi-static radar observations of Venus. (Right). The balloon for the proposed VALOR Discovery Mission being readied for deployment test.

Orbiter	2 Balloons	2 Landers	
Lifetime (4 years)	(1 month)	Descent Phase (1–1.5 hour)	Landed Phase (5 hours)
InSAR — Interferometric Synthetic Aperture Radar	ASI — Atmospheric Science Instrument (pressure, temperature, wind speed,)	ASI	Microscopic imager
Vis–NIR Imaging Spectrometer	GC/MS — Gas Chromatograph / Mass Spectrometer	Vis–NIR Cameras with spot spectrometry	XRD / XRF
Neutral Ion Mass Spectrometer	Nephelometer	GC / MS	Heat Flux Plate
Sub–mm Sounder	Vis-NIR camera	Magnetometer	Passive Gamma Ray Detector
Magnetometer	Magnetometer	Net Flux Radiometer	Sample acquisition, transfer, and preparation
Langmuir Probe	Radio tracking	Nephelometer	Drill to ~10 cm
Radio Subsystem (USO — Ultra Stable Oscillator)			Microwave Corner reflector

Table 1 - Venus Flagship Design Reference Mission Baseline Science Payload.

For ESA Cosmic Vision L2/L3 opportunities, any missions that are capable of addressing the science goals and measurement objectives are desirable. But more importantly, a coordinated international effort is also needed to maximize the science return which will require overtures and dialogue with other interested space agencies in Russia, India, Japan, US and other countries interested in participating in Venus exploration.

Much of the needed science can be addressed through the coordinated combination of small, medium and large missions. Some such missions are described below as illustrative examples. ***A key capability that is needed at Venus is an internationally coordinated reliable communications capability to relay data from long lived floating platform, landed probes and platforms which has been a main hindrance for many of the previous proposed missions. No new technology is needed for relay communications at Venus since the technology developed for Mars. However, it is necessary to adhere to standard communications protocols. The Interagency Operability Working Group (IOWG), which coordinates communication planning across space agencies, is already aware of the need to apply the needed standards to future Venus missions.*** Such a capability is available at Mars on many orbiters and has proved to be extremely valuable and very much needed at Venus.



Figure 3. Artists's renditions of the US Decadal Survey recommendation of the Venus Climate Mission (left) and Russia's Venera-D mission (right). These mission demonstrate the international approach in Venus exploration and for coordination of the international efforts. Such missions can be implemented with ESA's contribution under the L2/L3 opportunities.

3.1 - High Resolution Radar Mapping of the Surface and Altimetry

An improvement over the global Magellan data sets for SAR imagery (resolution ~ 75 m) and altimeter data (resolution: 100 m vertical accuracy ~ 10 km footprint) is desired to understand the geologic evolution of the planet, including how it has resurfaced, and whether or not there is evidence for a past tectonic environment. Similarly the resolution for the gravity field at Venus is non-uniform; at least 1/3 of the planet has a resolution too low to allow for an elastic thickness estimate. Estimating elastic thickness provides a window into the planets thermal state. A mission similar to the EnVision mission (Ghail et al., 2011) proposed to the previous Cosmic Vision opportunity is a candidate. There were four separate proposals to map the surface of Venus to the last Discovery call for proposals. Numerous countries, including India, Italy, Germany, and the US have advanced SAR capabilities.

3.2. - Atmospheric Composition, Trace Species and Atmospheric Circulation

Venus Express has demonstrated the capabilities of a modern suite of instruments, even instruments that were not originally designed for Venus. Improved and new instruments can provide even more pertinent observations of the Venus atmosphere. Atmospheric measurements needed require orbiters, floating platforms, descent probes and landed platforms. The instruments needed on each platform are noted below as illustrative examples.

3.2.1 - Orbiters

Proposed instruments for two orbiting spacecraft in circular orbiters in orthogonal inclined orbits (e.g. 45° and 135°) include:

- A Laser Altimeter (for cloud top topography) similar to the MOLA instrument and capable of detecting cloud top altitude
- A Doppler LIDAR for vertical profiles of line of sight wind (Singh et al., 2010)
- An improved imaging spectrometer (300 - 3400 nm)
- An infrared imaging radiometer (10-15 μ)
- USO with orbiter to orbiter radio science capability for mutual occultation probings of the Venus atmosphere
- An infrared heterodyne spectrometer (1-12 μ m)

One of the methods not used previously in spacecraft studies of planetary atmospheres, that has already demonstrated its high efficiency in ground-based observations, is high resolution infrared heterodyne spectroscopy (Kostiuk et al., Sornig et al.). With unprecedented spectral resolution up to $\sim 10^8$ this method gives an opportunity for direct wind measurements that have never been done before. Extensive use of fiber-optical and photonic technologies, including integrated waveguide chips, allows building a compact, lightweight heterodyne spectrometer for the orbiter. It is planned to implement ultra-high resolution measurements in solar occultation mode at 1.3-2 μ m and in nadir mode at 5-12 μ m. This instrument will be used for vertical profiling of minor constituents, precise measurements of isotopic ratios, and comprehensive characterization of 3D temperature and wind field above the cloud layer.

3.2.2 Floating Platforms

The VeGa balloons have already validated the value of floating platforms in the atmosphere of Venus at an altitude (~ 54 km) where Earth-like temperatures exist, even in the presence of sulfuric clouds. Lower and higher altitudes require different designs, but recent developments in use of Unmanned Aerial Vehicles in Earth's atmosphere and future deployment on Venus appears feasible. The UAVs offer advantages in for cloud top region and payload capacity.

UAVs at 60 – 70 km altitude to measure:

- Ambient wind flow and turbulence
-

- Characteristics of uv absorber(s) including nature – gaseous and/or particulate, composition, number density and shape and particle size distribution if particulate
- Heat deposition and next flux of radiation
- High-resolution surface imaging
- Distribution of trace species

One possible platform for in situ atmospheric exploration is a semi-buoyant, propelled aerial vehicle (Figure 4). Such a vehicle could have a long lifetime in the Venus atmosphere (months to years), limited only by the gradual loss of buoyant gas through the envelope and/or corrosive effects of the atmosphere. A Venus UAV would be maneuverable via (not real time) control from Earth and could provide mobility in latitude, longitude, and altitude. During propelled flight, the combination of lift and buoyancy provides altitude control between 55 and 70 km; in passive flight during the Venusian night, the vehicle floats at its altitude of 100% buoyancy (55 km). This 3-D maneuverability creates a science platform capable of surveying large areas and/or focusing on specific regions of interest, to make the following crucial science measurements:

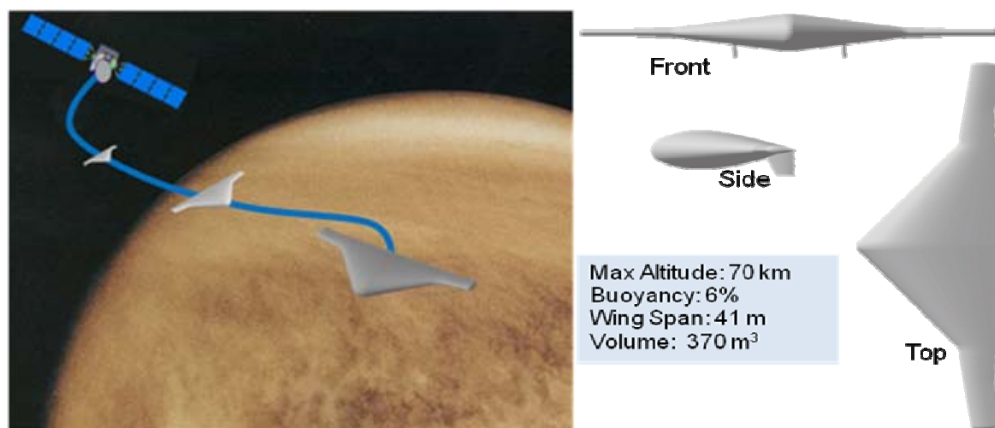


Figure 4. - Left: UAV released from orbiting spacecraft and inflating exo-atmospherically for benign entry. Right: Vehicle views and dimensions.

The payload accommodation capabilities to accomplish these measurements are comparable to those baselined in the Design Reference Mission balloon and can be significantly extended with minimal loss of performance.

From a technical perspective, a Venus UAV makes a good addition to a more comprehensive mission. The accompanying orbiter serves as the data and communications relay to Earth, as well as the delivery vehicle. The UAV is deployed from the orbiter and inflated in space, with a large surface area that then enables a benign entry. As a result, data collection is possible during descent to the cruising altitude, and loads on the science instrumentation are minimal. The elimination of the need for an aeroshell maximizes the mass available to the instrumentation.

In flight, the risks to the vehicle are low. The flight plan and power cycling are simple as the vehicle traverses day-to-night side once every 6 Earth days. If an event causes an entry into safe mode, the vehicle will drop to its altitude of 100% buoyancy (55 km) and passively float indefinitely until recovered. The Venus UAV concept is in the development

process, with systems architecture engineering efforts and demo vehicle fabrication and testing in relevant environments planned for this year and next.

Balloons at 50 – 60 km altitude

Ambient flow and turbulence, UV absorber characteristics (nature - gaseous/particulate, composition, number density and particle size distribution if particulate), heat deposition or net flux, surface imaging at NIR, trace species.

3.2.3 - Descent Probes and Landed Platforms

- Lower atmosphere chemistry and structure
- Surface images, composition, mineralogy, meteorological measurements near surface, incident solar (dayside), quake activity.

3.3 High Resolution Gravity Field and Geodesy

There are two approaches to detailed gravity field mapping: one using a single orbiter (PVO, Magellan for Venus and GOCE for Earth) and another using two orbiters (GRACE for Earth and GRAIL for the moon). The single-orbiter approach has been used for Venus already and has yielded the gravity field to degree 180 spherical harmonics. Due to the incapacitated S-band transmitting capability on Venus Express, the ability to retrieve the detailed gravity field has been impaired. GRAIL mission (Zuber et al., 2013) has provided some exciting results for the moon, and a similar high resolution gravity field may provide some clues about Venus.

GRACE, the DLR-NASA joint mission and the GRAIL mission to the Moon a NASA Discovery Program mission provide a two-spacecraft approach to mapping the detailed gravity field. The orbiters could also serve as communication orbiters and may also be able to support other instruments.

4. COORDINATION OF INTERNATIONAL VENUS EXPLORATION MISSIONS

The lessons from recent ESA and NASA planetary exploration programs have been very instructive. In particular, the exploration of Venus has been stymied, except for two low cost missions in the last decade – ESA's Venus Express, and JAXA's Akatsuki (Venus Climate Orbiter) mission which is still awaiting a second chance to orbit Venus and succeed in returning useful observations. The numerous Discovery and New Frontiers proposals for Venus as well as NASA's flagship mission study and the recommendations of the US Academies decadal survey of planetary science have demonstrated the scope and breadth of Venus science and of the magnitude of the effort required to answer the science questions about the planet. The evolving budget projections suggest that an international coordinated effort, supported by our group, is required for exploring Venus in the coming decades. A complex international mission, Venera-D is being studied by Russia and is now anticipated to be launched in 2023 – a time frame quite compatible with the Cosmic Vision. It is anticipated that India may also undertake a small mission to Venus between 2020-2022.

REFERENCES

- Alemi, A., and D.J. Stevenson, 2006. Why Venus has no moon. American Astronomical Society, DPS meeting #38, #07.03; Bulletin of the American Astronomical Society, Vol. 38, p.491.
- Altieri, F., et al., 2013, "VIRTIS/VEX O2 (a1g) nightglow profiles affected by gravity waves action; modeling and results", EGU General Assembly 2013; Copernicus
- Baines, et al., 2011. [Venus Atmospheric Explorer New Frontiers Mission Concept](http://www.lpi.usra.edu/vexag/resources/decadal/KevinHBaines.pdf), a White Paper for the Planetary Sciences Decadal Survey, <http://www.lpi.usra.edu/vexag/resources/decadal/KevinHBaines.pdf>
- Bengtsson, L. et al., 2013. Towards Understanding the Climate of Venus, ISSI Scientific Report Series, Springer, 193 pp.
- Caudal, G.V., 2010. Hypothesis of a spin-orbit resonance between the Earth and Venus's core, J. Geophys. Res., 115, E07011, doi:10.1029/2009JE003496.
- Chassefière, et al., 2012. The evolution of Venus: Present state of knowledge and future exploration, Plan. & Space Science, Vol. 63-34, 15-23, <http://dx.doi.org/10.1016/j.pss.2011.04.007>.
- Cottareau, L. et al., 2011. The various contributions in Venus rotation rate and LOD. Astron. Astrophys. 531. doi:10.1051/0004-6361/201116606.
- Correia, A.C.M., and J Laskar, 2001. Final four rotation states of Venus, Nature 411 (6839), 767-770.
- Correia, A., J Laskar, O.N. de Surgy, 2003. [Long-term evolution of the spin of Venus: I. theory](#), Icarus 163 (1), 1-23
- Correia, A., and J. Laskar, 2003. Long-term evolution of the spin of Venus: II. Numerical simulations, Icarus 163 (1), 24-45.
- Crisp, D., et al., 2002. "Divergent Evolution Among Earth-like Planets: The Case for Venus Exploration," The Future of Solar System Exploration, 2003–2013, Community Contributions to the National Research Council (NRC) Solar System Exploration Survey (Mark Sykes, Ed.), ASP Conference Series, vol. 272, pp. 5–34.
- [Davies, J. A.](#) et al, 2007. A Coronal Mass Ejection at Venus observed with STEREO HI and Venus Express, American Geophysical Union, Fall Meeting 2007, abstract #SH52B-06
- De Bergh et al., 1991. Deuterium on Venus - observations from Earth, Science, 251 (1991), pp. 547–549.
- Donahue, T.M., et al., 1982. Venus was wet: a measurement of the ratio of deuterium to hydrogen, Science, 216, pp. 630–633.
- Fedorova, A., et al., 2008. HDO and H₂O vertical distributions and isotopic ratio in the Venus mesosphere by Solar Occultation at Infrared spectrometer on board Venus Express, J. Geophys. Res., 113, E00B22, doi:10.1029/2008JE003146
- D.M. Hunten, 1992. [Venus: Lessons for Earth](#) Original Research Article *Advances in Space Research, Volume 12, Issue 9, September 1992, Pages 35-41*
- Garvin, J.B., et al. 2009. Venus: Constraining Crustal Evolution from Orbit Via High-Resolution Geophysical and Geological Reconnaissance. *Submitted to the Solar System Decadal Survey Inner Planets Panel Submitted to the Solar System Decadal Survey Inner Planets Panel*,
- Ghail, R.C., et al., 2012. [Experimental Astronomy](#), April 2012, Volume 33, [Issue 2-3](#), pp 337-363.
- Grinspoon, et al., 2009. Comparative Planetary Climate Studies, A White Paper for the Planetary Sciences Decadal Survey. <http://www.lpi.usra.edu/vexag/resources/decadal/ComparativeClimateWhitePaperV4.pdf>
- Helbert Jet al., 2008. Surface brightness variations seen by VIRTIS on Venus Express and implications for

- the evolution of the Lada Terra region Venus, *Geophys. Res. Let.*, 35:11201, doi: 10.1029/2008GL033609.
- Juanes-Vallejo, C., 2011. A Surface Enhanced Raman Scattering (SERS) Payload for Life Detection On-Board a Venusian Balloon Platform, Presented at the 2011 VEXAG International Venus Workshop, Chantilly, Virginia. Available at: http://venus.wisc.edu/workshop/vexag/2011/presentations/CJV_SERS-Life-Detection.pdf.
- Lebonnois, S., et al., 2013. Models of Venus atmosphere. In *Towards Understanding the Climate of Venus*, Bengtsson et al., Eds. ISSI Scientific Report Series, doi: , Springer, 193 pp.
- Limaye et al., 2009. VEXAG Goals, Objectives, and Investigations, a white paper submitted to the US National Academies Decadal Survey of Planetary Science. <http://www.lpi.usra.edu/vexag/resources/decadal/SuzanneESmrekar.pdf>
- Limaye, S.S., et al., 2009. Venus atmosphere: major questions and required observations. A white paper submitted to the Planetary Science Decadal Survey Inner Planets Panel. Available at: <http://www.lpi.usra.edu/vexag/resources/decadal/SanjaySLimaye.pdf>
- Morowitz, H., and C. Sagan, 1967. Life in the clouds of Venus?, *Nature*, 215, 1259-1260.
- Mueller N., J. Helbert, G. L. Hashimoto, C. C. C. Tsang, S. Erard, G. Piccioni, and P. Drossart. (2008) Venus surface thermal emission at 1 μm in VIRTIS imaging observations: Evidence for variation of crust and mantle differentiation conditions. *JGR Planet*, doi: 10.1029/2008JE003118.
- Nakagawa H., et al., 2013. Comparison of general circulation model atmospheric wave simulations with wind observations of venusian mesosphere, *Icarus* 2013, in press.
- Picialli A., et al., 2012. Dynamical properties of the Venus mesosphere from the radio occultation experiment VeRa onboard Venus Express, *Icarus*, 217, 669-681.
- Sagan, C., *Science*, 133, 849 (1961).
- Saunders, R.S., et al., 1992, "Magellan mission summary," *J. Geophys. Res.*, v.97, no. E8, p. 13,067-13,090
- Schulze-Makuch, D. and L. N. Irwin, 2002. Reassessing the Possibility of Life on Venus: Proposal for an Astrobiology Mission, *Astrobiology*. June 2002, 2(2): 197-202, doi:10.1089/15311070260192264.
- Schulze-Makuch, D., D.H. Grinspoon, O. Abbas, 2004. A sulfur-based UV adaption strategy for putative phototrophic life in the Venus atmosphere. *Astrobiology*, 4, 11-18.
- Smrekar, S.E., 1994. Evidence for active hotspots on Venus from analysis of Magellan gravity data, *Icarus*, 112, 2-26.
- Smrekar, S.E., et al., 2010. Recent Hotspot Volcanism on Venus from VIRTIS emissivity data, *Science*, 328, p. 605-608.
- Sornig et al., 2013. Venus' upper atmospheric dynamical structure from ground-based observations shortly before and after Venus' inferior conjunction 2009, *Icarus*, in press.
- Taylor, F., and D. Grinspoon (2009), Climate evolution of Venus, *J. Geophys. Res.*, 114, E00B40, doi:10.1029/2008JE003316.
- [Williams, C.](#) et al. 2009. Radar Reflectivity Changes of Rock Surfaces due to Thermally-Activated Charge Carriers: Contributing to the Understanding of the Magellan Venus Radar Reflectivity Data, American Geophysical Union, Fall Meeting 2009, abstract #P33A-1277.
- Valley, J., et al., 2002. A cool early Earth. *Geology* 30: 351-354.
- Zuber et al., 2013. Gravity Field of the Moon from the Gravity Recovery and Interior Laboratory (GRAIL) Mission, *Science*, 8 February 2013, Vol. 339 no. 6120 pp. 668-671.
-

Space-Time Structure Explorer

Sub-microarcsecond astrometry for the 2030s

Response to the call for White Papers for the definition of
the L2 and L3 missions in the ESA Science Programme

24.05.2013

Spokesperson: Anthony Brown

Contact details:

Dr. Anthony G.A. Brown

Chair, Gaia DPAC Executive

Leiden Observatory, Leiden University

Niels Bohrweg 2

2333 CA Leiden

The Netherlands

e-mail: brown@strw.leidenuniv.nl

phone: +31-71-5275884

Contributors to this white paper

Members of the Gaia Science Team

C. Jordi	University of Barcelona - ICC/IEEC
L. Lindegren	Lund University
S.A. Klioner	Technische Universität Dresden
M.S. Randich	Osservatorio Astrofisico di Arcetri
C. Soubiran	Laboratoire d'Astrophysique de Bordeaux
N.A. Walton	Cambridge University

Members of the Gaia Data Processing and Analysis Consortium Executive

C. Babusiaux	Observatoire de Paris
U. Bastian	Zentrum für Astronomie der Universität Heidelberg
C.A.L. Bailer Jones	Max Planck Institut für Astronomie Heidelberg
S. Els	Zentrum für Astronomie der Universität Heidelberg
L. Eyer	Observatoire de Genève
X. Luri	University of Barcelona - ICC/IEEC
D. Pourbaix	Université Libre de Bruxelles
P. Sartoretti	Observatoire de Paris
A. Vallenari	INAF - Osservatorio Astronomico di Padova
F. van Leeuwen	Cambridge University

This white paper also benefited from contributions and comments by C. Turon and A. Helmi.

Motivation for a brief white paper on space astrometry

The call for white papers for the definition of the L2 and L3 missions in ESA's science programme asks for the development of science themes for these future large missions. With the launch of Gaia to occur later in 2013, it is appropriate to assess the full scientific potential of Gaia in the light of its actual in orbit performance, before embarking on a detailed road mapping exercise for the next leap in astrometry.

We anticipate the production of the final science data products from Gaia in 2021/22 and thus a range of new questions to emerge around 2025. This implies the need for the next leap in observational capability in the 2030+ time frame – consistent with an L3 astrometry mission around 2034.

Thus, in this short white paper we outline a few broad discovery themes that could be addressed by such a mission and stress the need to keep the ESA and European expertise on space astrometry up to date in order to make sure that in due time suitable proposals for the next space astrometry mission can be put forward.

Options for future space astrometry missions

Gaia will provide a jump of two orders of magnitude in accuracy and four orders of magnitude in number of stars with respect to Hipparcos. Needless to say that the impact of Gaia on the mapping and understanding of the Milky Way, stellar astrophysics, solar system science, extragalactic astrophysics, and fundamental physics, may even go much farther than what we now expect. In addition Gaia will surely uncover surprises which may point to entirely new directions in which to take astrophysics. Thus once the Gaia results are known the most interesting directions to take for future space astrometric missions will be much clearer.

These future directions will also have to be considered in the context of the state of astronomy and astrophysics when the large digital sky surveys (Pan-Starrs, LSST, spectroscopic surveys) and the Euclid mission have delivered their results. Data from the new astronomical mega-facilities ALMA, E-ELT, and SKA will further influence the global context. In addition developments in the area of exoplanet science should be taken into account. Nevertheless it is possible to identify a number of options for future astrometry missions that would cover capabilities not offered by Gaia. We list three options here, starting from the most ambitious.

2.1 Nano-arcsecond astrometry

We believe that the long term goal for space astrometry should be to make the next big breakthrough beyond Gaia and aim for the sub-micro-arcsecond (sub- μ as) or even nano-arcsec (nas) regime, preferably using techniques that enable the 'global astrometry' delivered by Hipparcos and Gaia. The figure illustrates the power of sub- μ as astrometry by translating the astrometric accuracies to accuracies achieved in direct distance and transverse motion measurements. We briefly list some of the science cases that would be enabled by sub- μ as to nas astrometry:

Census of terrestrial planets around nearby stars This case was already developed briefly by M.A.C. Perryman for the 'Cosmic Vision: Space Science for Europe 2015–2025' report. He showed that a 10 nano-arcsec precision mission will be able to survey hundreds of thousands of stars out to 100 pc for the presence of earth-sized planets. This science case has been developed further in the context of the NEAT proposal, which aims at the 50 nano-arcsec level for small-field astrometry. Clearly the interest of this science case should be weighed against future developments in the field of exoplanets.

Geometric or real-time cosmology It was already noted in the Gaia Concept and Technology Study report that astrometric measurement accuracies much better than a micro-arcsec would allow the direct determination of the transverse motions of external galaxies and quasars, and thus measure their kinematic properties independently

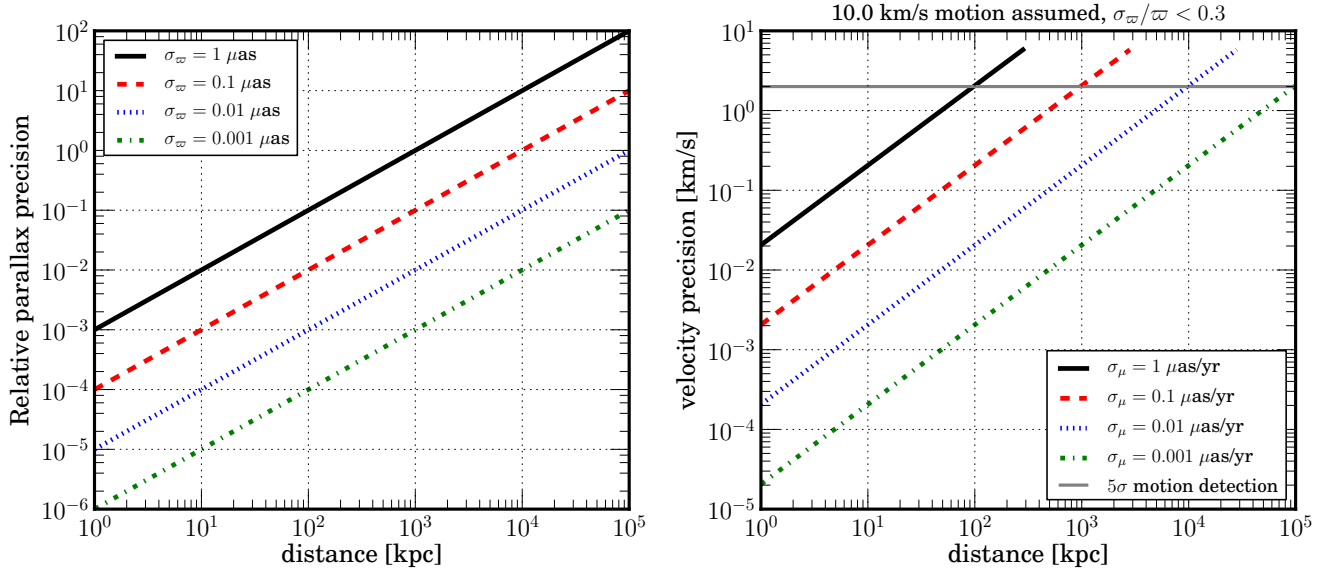


Figure 1: Illustration of the reach of a sub- μas astrometric mission. The left panel shows the relative parallax error achieved for individual sources as a function of distance for four levels of parallax accuracy. The right panel shows the transverse motion errors achieved for the same levels of accuracy on the proper motion measurements for sources moving at 10 km/s. For the calculations of the transverse motion errors the relative parallax error was assumed to be smaller than 30% and the ratio of parallax to proper motion measurement errors was assumed to be 2.

of a dynamical model of the universe. In addition several papers have been written in which the concept of ‘real-time’ cosmology is explored (Quercellini et al., 2012, 2009), in which very high accuracy astrometry permits the measurement of changes over time in the angular separation between sources at cosmic distances, providing a powerful consistency test of the assumed metric and independent constraints on cosmic anisotropy.

Fundamental physics A global astrometric mission at the μas level would allow much more stringent tests of General Relativity through light bending and would also enable the measurement of the energy density of stochastic gravitational wave background. In the narrow field regime one could for example weigh many neutron stars in binary systems in a model-independent way. In addition, reaching 1 μas or better for faint sources would allow the observation of pulsar-white dwarf systems and lead to tests of GR in the strong field regime.

Dark energy At the nano-arcsec level the quasar secular parallax shift due to the Solar system motion with respect to the CMB will be detectable and would lead to geometric constraints on dark energy (Ding & Croft, 2009).

Dark matter astrometry As laid out in Majewski et al. (2009) and Shaya et al. (2009) sub-micro-arcsec to nano-arcsec astrometry on a global scale (and to faint magnitudes) will allow the precise mapping of dark matter from the outer reaches of the Milky Way to beyond the local group.

Internal motions throughout the Local Group Gaia will not be able to directly measure internal motions of the nearby dwarf galaxies. Going to the sub- μas or μas regime would enable resolving 10 km/s motions out to 1–100 Mpc, thus opening up the tantalising possibility of measuring internal motions and thus astrometrically resolving the dynamics of Galaxies even beyond the Local Group and out to the nearest galaxy cluster.

Standard candles throughout the Local Group Sub- μas astrometry will enable the direct distance measurement (avoiding the problems caused by (patchy) extinction or metallicity effects) of various stellar standard candles all throughout the Local Group, and even up to the closest galaxy clusters for the brightest candles, i.e. in very different environments (galaxy type, metallicity) from the Milky Way.

More science cases have been developed in the context of the SIM mission (Unwin et al., 2008) and of course include applications to stellar physics. The latter will obviously benefit from the next level in parallax accuracy, bringing very rare and typically distant stellar types into the range of per cent level or better distance accuracy.

2.2 Global infrared astrometry at sub micro-arcsecond accuracy

An obvious complement to Gaia would be an infrared mission able to peer through the dust in the plane of the Milky Way. It should achieve at least the same astrometric accuracy as Gaia. This would enable a much better astrometric mapping of the inner disk, spiral arms, the bar, and the bulge of our galaxy. In addition one would have access to low-mass stars, brown dwarfs, and free floating planets over a large volume around the Sun.

We note that the Japanese community has plans for a near-infrared astrometric mission, JASMINE, which will be preceded by the nano-JASMINE and Small-JASMINE missions. JASMINE would only cover the inner disk and bulge and thus not be an all sky mission.

The interest and detailed requirements for an infrared astrometric mission will be much clearer once the Gaia results have been analyzed, especially through the combination of Gaia measurements with data from surveys probing the interstellar medium (such as Planck).

2.3 Repeat of Gaia

Although at first glance this option looks rather unambitious, it is nonetheless of significant interest, in particular in the area of extending the baseline of proper motion measurements. This will allow for more precise but also more accurate proper motions, especially for unresolved binary stars. In addition one would get better measurements of the motions of globular clusters and nearby dwarf galaxies. The repeat of Gaia would permit some of the real time cosmology experiments by making use of the secular parallax effect over the extended time baseline. Finally, a ‘Gaia-2’ in the 2030s would also improve the parallaxes from Gaia.

This option would possibly fit a lower cost envelope (a medium class mission perhaps) and one should of course consider to enhance the design, for example by going to fainter magnitudes at the same accuracies, and by optimizing the photometric and spectroscopic instrument designs. Note that at Gaia or better precision levels simultaneous photometry and spectroscopy are a must for the correct interpretation of the astrometric measurements.

2.4 Staying competitive

Gaia was first proposed 20 years ago and will deliver its final results 10 years from now. In the intervening 30 years observational astrophysics will have evolved a lot and many of the science cases for Gaia will have been addressed already to some extent. For example the RAVE and SDSS surveys have delivered many new results in the area of the Milky Way structure and formation history and the same will hold for the Gaia-ESO survey, Pan-Starrs and LSST.

However Gaia provides unique capabilities which will keep its results at the forefront of astrophysics over the coming decades:

- Global astrometry: absolute parallaxes and proper motions to unprecedented accuracies, simply not achievable from the ground.
- All sky, homogeneous, multi-epoch photometry and spectroscopy.
- Spectroscopy for numbers of objects out of reach of ground based efforts.
- Mapping of the full sky at HST-like angular resolution to 20th magnitude.

In fact Gaia would have remained competitive even if other astrometric missions such as SIM, DIVA and JASMINE would have flown already. The message is that any ‘astrometric’ science case and corresponding mission concept should remain competitive over the time scale from now to 2040–2060. Therefore we strongly believe that the most attractive option (as judged at this point in time) for a next large space astrometry mission is one that aims for the sub- μ as to nano-arcsec regime over the full sky.

Challenges on the road to (global) sub-microarcsec astrometry

Although we do not present a mission concept here we do outline some of the challenges that need to be solved in order to achieve global sub- μ as astrometry in the future.

3.1 Engineering

If we assume that the sub- μ as astrometry mission will be done in the optical then basic considerations (e.g. Lindegren, 2005) show that astrometric precision scales as:

$$\sigma \propto \frac{\lambda}{B\sqrt{N}},$$

where λ refers to the effective wavelength of the measurements, N to the number of photons collected and B to the aperture size of the mirrors or the baseline of an interferometric system. In the optical the collection efficiency of photons by modern detection systems is almost 100%, meaning that gains in accuracy can only be achieved through increasing B .

This quickly leads to the argument for an interferometric mission with collecting areas of a few m^2 and baselines of 100–1000 meter (as argued by Perryman in his Cosmic Vision proposal, see also Lindegren, 2007). Thus precision formation flying will have to be developed and a concept to do global astrometry with such a configuration.

An important challenge is that the (thermo-)mechanical stability of the entire spacecraft down to the level required for a smooth motion of the telescope is far from trivial. For stability at measurement levels below 1 μ as internal movements well below a micron over the entire body of the spacecraft will become critical. Experience with Gaia so far has shown that this is well beyond (3 orders of magnitude) anything engineers can estimate or predict in a large structure. Gaia will teach us a lot in this respect.

Similar considerations hold for the requirements on attitude control and knowledge of the barycentric velocity of the spacecraft.

Alternatively one would develop technology that could break the scaling relation above in such a way as to enable sub- μ as astrometry at much lower cost (in terms of engineering).

Further developments that should be pursued:

- Detectors that allow for photon collection in a way not sensitive to radiation damage effects. Already for Gaia dealing with these effects will be very challenging. It is not likely that ‘software solutions’ can be made to work for the sub- μ as regime.
- If the infrared option is pursued the challenge of developing IR-sensitive detectors that could be operated in Time-Delayed Integration (TDI) mode would have to be tackled in addition to issues of cooling the payload sufficiently without disturbing the astrometric measurements.

3.2 Data processing

Data processing to achieve sub- μ as precisions will be complicated by (among others) the following:

- Relativistic modelling of astrometric measurements will have to be pushed to the nano-arcsec level in order to correctly interpret the raw data. It is believed that 100 nas precision in relativistic modelling can be achieved already. Going further would require research and improvements in our knowledge of the solar system (asteroid masses for example).

- System calibration at this level will have to be much better than (by an order of magnitude) than the astrometric accuracy aimed for and will thus be extremely challenging. The design of the instruments and mission concept will have to incorporate the data processing demands from the start. That is, any future ‘concept and technology study’ as well as subsequent development phases should treat engineering and data processing on an equal footing.
- Approaching μ as levels it is not obvious that simple models of the time dependence of source coordinates will be sufficient. In addition research into sources of astrometric jitter (such as star spots or microlensing in crowded regions) and their effect on the interpretation of image locations in the data stream is required.
- Another conceptual problem when approaching the μ as regime is that for parallax measurements small (< 1 AU) sources which are sufficiently bright (hot) are required in large numbers (for global astrometry). It is not clear that at cosmological distances this requirement is fulfilled (Lindgren, 2007).

For the first two elements above a lot will be learned from the data processing for the Gaia mission.

4

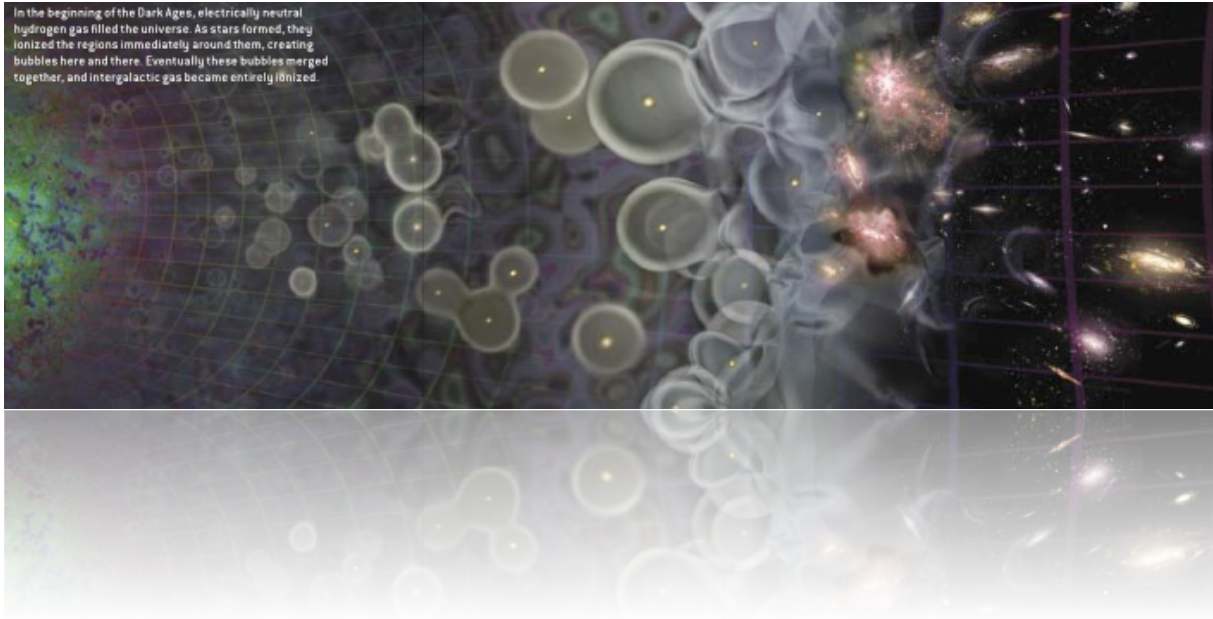
Conclusions

Gaia will soon provide a detailed view of our Milky Way enabled through astrometry at the micro arcsec level. European leadership in the delivery of this advanced astrometric mission will ensure that the European astronomical community is at the forefront of a range of research furthering our understanding of a wide domain of key astrophysical topics. However, Gaia is likely to throw up new questions which will demand the next level of data in order to fully understand the nature of Dark Matter, and how it fundamentally drives the formation and evolution of galaxies, large and small. These questions will only be answered through detailed precision mapping of not only our Milky Way, but neighbouring galaxies in our local group. This in turn will demand the leap to sub- μ as or even nano-arcsecond astrometry.

Over the coming years the Gaia science community will embark on an intense series of workshops to develop the key science themes which will scope the requirements for a future ‘Space-Time Structure Explorer’ mission. This will culminate in a detailed white paper which will be published to coincide with the first releases of Gaia data. This will allow for a simultaneous exposure of not only the early science yield of Gaia but also set our European led programme in motion to ensure continued momentum in the development of the required new technologies and eventual mission to deliver the next leap in our understanding of the fundamental interplay of matter (dark and light) and energy shaping our cosmos.

References

- Ding, F., Croft, R.A.C., 2009, MNRAS, 397, 1739, [ADS Link](#)
- Lindgren, L., 2005, In: Turon, C., O’Flaherty, K.S., Perryman, M.A.C. (eds.) The Three-Dimensional Universe with Gaia, vol. 576 of ESA Special Publication, 29, [ADS Link](#)
- Lindgren, L., 2007, Highlights of Astronomy, 14, 481, [ADS Link](#)
- Majewski, S.R., Bullock, J., Burkert, A., et al., 2009, ArXiv e-prints, [ADS Link](#)
- Quercellini, C., Quartin, M., Amendola, L., 2009, Physical Review Letters, 102, 151302, [ADS Link](#)
- Quercellini, C., Amendola, L., Balbi, A., Cabella, P., Quartin, M., 2012, Physics Reports, 521, 95, [ADS Link](#)
- Shaya, E., Olling, R., Ricotti, M., et al., 2009, In: astro2010: The Astronomy and Astrophysics Decadal Survey, vol. 2010 of ArXiv Astrophysics e-prints, 274, [ADS Link](#)
- Unwin, S.C., Shao, M., Tanner, A.M., et al., 2008, PASP, 120, 38, [ADS Link](#)



A White paper for a low-frequency radio interferometer mission to explore the cosmological Dark Ages for the L2, L3 ESA Cosmic Vision program

24/05/2013

***DEX Spokesperson:
Dr. Marc Klein Wolt***

Department of Astrophysics - Research Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen

***M.KleinWolt@astro.ru.nl
Heijendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
T: +31 (0) 644130582***

Authors:

Marc Klein Wolt (RU-Nijmegen, NL)
Garrelt Mellema (Stockholm University)
Leon Koopmans (RUG-Groningen, NL)
Heino Falcke (RU-Nijmegen, NL)
Willem Baan (ASTRON, IFR, NL)
Albert-Jan Boonstra (ASTRON, NL)
Leonid Gurvits (Jive, TUD-Delft, NL)

Mark Bantum (Univ. Twente, ASTRON,
TUD, NL)
Chris Verhoeven (TUD- Delft, NL)
Joseph Lazio (JPL, USA)
Jack Burns (U. Colorado, USA),
Jan Bergman (IFR, Sweden)
Philippe Zarka (LESIA, OBSPM, CNRS)

Supporters of the DEX white paper:

Ian Crawford (Birkbeck College London)
Michael Garret (ASTRON, NL)
Jan-Erik Wahlund (IFR, Sweden)
Jean-Louis Bougeret (LESIA, OBSPM)
Amin Aminaei (RU, NL)
Hamid Pourshaghghi (RU, NL)
James Carpenter (ESA/ESTEC, NL)
Kees van 't Klooster (ESA/ESTEC, NL)
Olaf Scholten (KVI, NL)
Stephane Corbel (CEA-AIM, Paris)
Huub Röttgering (Leiden Observatory, NL)
Jason Hessels (ASTRON, NL)
Ger de Bruyn (ASTRON, NL)

Dark Ages eXplorer - DEX

Summary

The Dark Ages eXplorer, DEX, is a low-frequency radio interferometer in space or on the moon, that will probe deep into the early universe and provide an unprecedented detailed view of the evolution of the first structures in the universe. DEX is sensitive in the 1-80 MHz regime which allows a complete view of the Dark Ages and Cosmic Dawn period, essentially covering the $z=17-80$ redshift regime. DEX will explore the Dark Ages and the Cosmic Dawn and observe the global neutral hydrogen (21 cm) emission and its variations on arcmin scales in order to constrain cosmological models on the evolution of the early universe, the onset of the epoch of reionization and basically constrain the models and predictions that will follow from the Planck mission. These issues form the holy grail of cosmology and the Dark Ages is the treasure-trove for Dark Matter and Early universe physicists. In addition, DEX will open up the last virtually unexplored frequency regime below 30 MHz, extending the view of LOFAR and SKA to the ultra-long wavelength regime that is not accessible from Earth, and among a wealth of science cases will provide high resolution low-frequency sky maps, constrain models on the jet power in radio galaxies, observe auroral emission from the large planets in our solar system and possibly find Jupiter-like exoplanets. DEX requires a large collecting area in the order of 10 km^2 (10^5 individual elements) and a location (preferably the lunar far-side) that provides shielding from man-made radio interference (RFI), absence of ionospheric distortions, and high temperature and antenna gain stability. The realization of large collecting area interferometers has been achieved on earth already (e.g. LOFAR and in the near future SKA) and the technique behind it is well developed (TRL 4-5). The realization in space or on the moon is a very challenging task. However, we are convinced that given the technology developments, especially in the areas of nano satellites, RF technology and low-power and high-performance computing, the construction of a large low-frequency radio array in space or on the moon is feasible in the 2020-2030 timeframe.

Introduction & Motivation

The Earth's turbulent ionosphere gives rise to "radio seeing", making ground-based radio imaging of the sky difficult at frequencies below ~ 100 MHz. At even longer wavelengths below about 10-30 MHz one encounters the ionospheric cut-off where radio waves are reflected, permitting long-distance short wave transmission around the Earth, but making observations of the sky very difficult, which are finally prohibited by the ionosphere below ~ 10 MHz. Observing just above the cut-off, i.e., between $\sim 10-50$ MHz requires especially favorable geomagnetic and ionospheric conditions to obtain any decent images. The range below the cutoff is only readily observable from space. Hence, the dominant "low-frequency/long-wavelength" regime for which ground-based telescopes are being designed is at frequencies above 30 MHz and wavelengths below 10 m. Even at this range, there are windows of strong interferences (e.g. FM frequencies, 80-100 MHz), which cannot be used for radio astronomy.

The best resolution achieved so far in the range below ~ 30 MHz is on the scale of a few degrees, but more typically of order tens of degrees. This compares rather unfavorably to the milli-arcsecond resolution that can be routinely obtained in very long baseline interferometry (VLBI) at higher radio frequencies. Hence, the low-frequency Universe is the worst charted part of the radio spectrum, and perhaps even of the entire electromagnetic spectrum. By today, only two kinds of maps of the sky have been made at frequencies below 30 MHz. The first are maps made from earth-based observatories such as the maps obtained by Cane & Whitham (1977), Ellis & Mendillo (1987) and Cane & Erickson (2001) of a part of the southern sky near the Galactic center and those made by the UTR-2 radio telescope (Braude et al., 2002). These have angular resolutions ranging from ~ 1 to 30 degrees. The second kind are maps obtained by the

RAE-2 satellite (Novaco & Brown, 1978) with angular resolution of 30 degrees or worse. None of these maps show individual sources other than diffusive synchrotron emission of the Galaxy, nor do they cover the entire sky.

To improve this situation and to overcome these limitations, space-based low-frequency telescopes are required for all observations below the ionospheric cutoff (Weiler, 1987; Weiler et al., 1988; Kassim & Weiler, 1990). This is also true for a significant part of the seeing-affected frequency range above the cutoff frequency (see Figure 1) where high-resolution and high-dynamic range observations are required, such as imaging of redshifted 21-cm emission of neutral hydrogen in the very early Universe (Carilli et al., 2007).

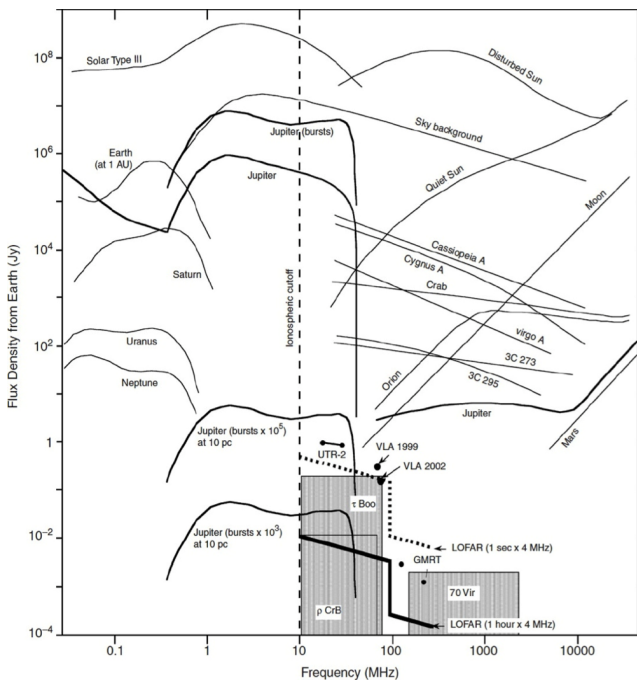


Figure 1: Spectra of astronomical radio sources detected from the Earth's vicinity (1 Jansky = $1 \text{ Jy} = 10^{-26} \text{ Wm}^{-2}\text{Hz}^{-1}$). Galactic, extragalactic and solar spectra are adapted from (Kraus, 1986). Planetary spectra, corresponding to auroral radio emissions, are adapted from (Zarka, 1992). Jupiter's spectrum, which includes auroral and Io-induced decameter emissions, is from Zarka et al. (2004). Its average flux density is about 10^6 Jy , while peak flux densities reach or exceed 10^7 Jy during short-lived bursts. If all planetary emissions were normalized to the same observer distance of 1 AU, Jupiter's spectrum should be upscaled by x20, Saturn's by x100, Uranus' by x400, and Neptune's by x900, so that all are grouped within 2–3 orders of magnitude of each other. Jupiter's peak spectrum is reproduced with two different scalings to illustrate the possible radio spectrum of hot Jupiters. Shaded boxes are predictions from (Farrell et al., 1999). Sensitivities of UTR-2, VLA, GMRT and future LOFAR observations are indicated. The Earth's ionospheric cutoff is indicated at 10 MHz. (from Zarka, 2007)

So far, there have only been two space missions whose primary purpose was low-frequency radio astronomy: the Radio Astronomy Explorers (RAE) 1 and 2. It was a surprising finding of RAE-1 that the Earth itself is a strong emitter of low-frequency bursts, the so-called Auroral Kilometric Radiation (AKR) generated by solar-wind interactions with the Earth's magnetic field. This emission is so strong that RAE-2 was placed in a lunar instead of the originally planned terrestrial orbit to provide shielding from this natural terrestrial interference. The structure of the Galaxy's emission was only seen by RAE-2 at very low spatial resolution (with beam sizes between $37^\circ \times 61^\circ$ and $220^\circ \times 160^\circ$) and fairly low signal-to-noise ratio. Due to the very limited angular resolution and the large power of AKR, it proved impossible to image any discrete sources directly.

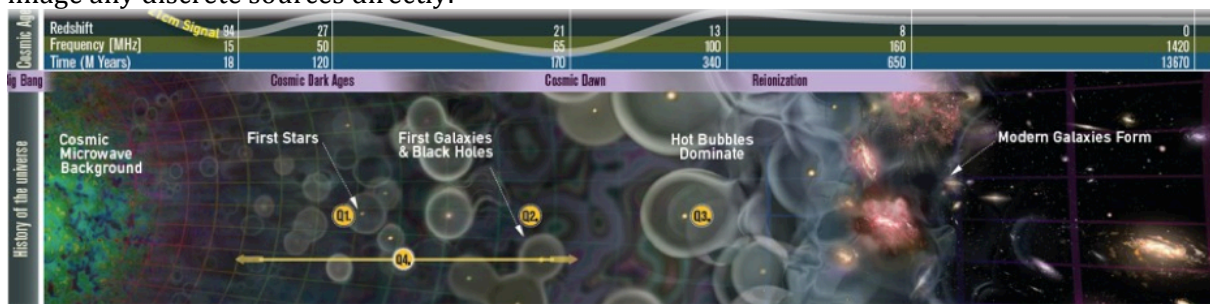


Figure 2: Formation history of the Universe. The Cosmic Dark Ages is the period with no stars and the only source of light is originating from the most abundant element, hydrogen. This H-I emission (21 cm line emission) is the only source of light that provides clues about the distribution of matter and dark matter in the early universe (Burns et al. 2012)

A fundamental question of current cosmological research is on the nature of structure formation in the Universe (for a review, see Ciardi and Ferrara, 2005): how is the observed distribution of visible matter created from the initial conditions just after the big bang, when matter and radiation were distributed

extremely smoothly, with density variations of just one part in 100,000? The Cosmic Microwave Background (CMB) radiation was emitted at $z \approx 1200$, about 400,000 years after the Big Bang, when the Universe had cooled off sufficiently for electrons and protons to recombine into neutral hydrogen atoms (Epoch of Recombination, EoR), allowing the background radiation to move freely without being scattered by the electrons and protons. At the same time, however, the Universe became opaque to visible light, because neutral hydrogen atoms absorb visible and infrared photons and re-emit them in random directions. Moreover, there were no sources of light in the Universe yet: the hydrogen and helium that were created in the big bang first had to cool in order to be able to clump together and form stars and galaxies. Hence, this era is called the "cosmic dark ages" in the redshift range $z=30-1200$ (Rees, 1999), see Figure 2.

Things only changed after the first stars, galaxies and active black holes had formed and emitted enough UV and X-ray photons to reionize the neutral hydrogen, allowing all radiation to pass freely. The time when this happened is called the Epoch of Reionization (EoR) and is believed to have occurred around $z \approx 11$, about 400 million years after the Big Bang, though it is at present not known whether the reionization happened more or less instantaneously, similar to a global phase transition, or was more or less spread out in time, depending on local conditions.

Throughout all these epochs hydrogen played a major role, emitting or absorbing the well-known 21-cm (1.4 GHz) line due to the spin-flip of the electron. This emission is redshifted by a factor 10 - 1000 due to the cosmological expansion and ends up in the frequency range from 140 - 1.4 MHz. When the hydrogen spin temperature is not coupled perfectly to the radiation temperature of the cosmic background radiation¹, but changed by other couplings with the surrounding matter and radiation, it can be seen against the cosmic background radiation in absorption or emission, depending on whether the spin temperature is lower or higher than the background radiation temperature. In this way, the cosmological 21-cm emission carries information about the evolution of the Universe.

Placing a radio antenna array at a large distance in space (e.g. Sun-Earth L2) or on the moon (e.g. in an eternally dark crater at the south (or north) pole) would provide perfect shielding from man-made radio interference (RFI), absence of ionospheric distortions, and high temperature and antenna gain stability that allows for sensitive and high resolution radio astronomy observations addressing a wealth of science cases – essentially opening up the last virtually unexplored frequency regime. In particular, such a low-frequency array would allow for the first time to probe the deep into the dark ages and observe the global neutral hydrogen (21 cm) emission and its variations in order to constrain cosmological models on the evolution of the early universe, the onset of the epoch of reionization and basically constrain the models that will come out of the Planck mission. These issues form the holy grail of cosmology and the Dark Ages is the treasure-trove for Dark Matter and Early universe physicists. This is a very challenging task that requires a large collecting area in space (or on the moon), very stable conditions and long integration times. However, we are convinced that given the technology developments, especially in the areas of nano satellites, RF technology and low-power and high-performance computing, the construction of a large low-frequency radio array in space or on the moon is feasible in the 2020-2030 timeframe.

DEX is a radio interferometer in space or on the moon, with a large collecting area in the order of 10 km², covering the low frequency regime between 1-80 MHz which is (partly) not accessible from earth and allows a complete view of the Dark Ages period essentially covering the $z=17-80$ redshift regime. The realization of large collecting area interferometers has been achieved on earth already (e.g. LOFAR and in the near future SKA) and the technique behind it is well developed (TRL 4-5). For the space application DEX will rely on the expected rapid technology developments in the domain of nano-satellites, swarm-technologies, low-power & high-performance processing, and (optical) communication.

¹ Which is not a *microwave* background at those redshifts.

Science questions addressed by DEX

Primary science case: Dark Ages

After the epoch of recombination (redshift 1200) the Universe remained in neutral state until the large scale formation of galaxies released a sufficient number of ionizing photons into the intergalactic medium to ionize it. This process is thought to have completed by redshift 6 after which the only areas of neutral hydrogen in the Universe were to be found inside galaxies. Before the completion of the reionization, large regions of the Universe consisted of neutral hydrogen that, under favorable circumstances, produces an observable 21cm (1.4 GHz) signal that is observable today in the 1.2-200 MHz range due to the redshift. The whole period between $z \sim 1200$ and $z \sim 6$ can be divided into three different periods based on the nature of the physical processes at work and the 21cm signal generated (see also Figure 2):

1. Dark Ages (DA) - In this period which stretches from $z \sim 1200$ to ~ 30 the Universe was nearly completely neutral and no sources of light had formed. The evolution of matter was relatively simple as the only process at work was gravity. Due to the coupling of radiation and matter only the lower redshift part of this period is expected to generate 21cm absorption against the cosmic microwave background ($z < 200$). Models show that this signal is strongest around $z \sim 50$ (~ 28 MHz) and becomes increasingly weaker as $z \sim 30$ (~ 46 MHz) is approached.
2. Cosmic Dawn (CD) - In this period which stretches from $z \sim 30$ to ~ 15 the first dark matter halos collapsed and the baryons which followed in the collapse formed the first stars. The ultraviolet radiation from these stars changed the quantum state of the cold neutral hydrogen so that it once more produced an observable absorption against the CMB. Once a sufficient number of x-ray sources had heated the Inter-Galactic Medium (IGM) above the CMB temperature the signal changed from absorption to emission.
3. Epoch of Reionization(EoR) - From about $z \sim 15$ to 6 the increasing number of ionizing photons escaping from galaxies started to substantially reduce the amount of neutral hydrogen in the IGM leaving it in a fully ionized state by the end of it. The still neutral areas of the heated IGM produced a 21cm emission signal (redshifted to ~ 90 -200 MHz).

The first generation of low frequency radio telescopes on Earth such as LOFAR, MWA, PAPER, are targeting the EoR period (> 100 MHz) and are hoping to detect a signal within the next 5 years. The low frequency part of the Square Kilometer Array (SKA_Low) will target both the EoR and the CD periods (frequency range 50 – 300 MHz) and is expected to become operational around 2020. As explained above, Earth-based observations of even lower frequencies are challenging if not impossible and require space missions. ***Aiming for a frequency range of 1 – 80 MHz, DEX will target both the observable part of the DA period as well as early CD period.*** The science cases for these two periods are different and below we address them separately.

Dark Ages Science

During the DA gravitational collapse was mostly in the linear regime. Initially matter and radiation were sufficiently coupled to make the gas temperature equal to the CMB temperature, dropping as $(1+z)$. Only around $z \sim 200$ had the matter density dropped enough to allow the gas temperature to evolve independently, falling off adiabatically as $(1+z)^2$. The density was however still high enough to collisionally couple the quantum state of neutral hydrogen (expressed by the spin temperature T_s) to the gas temperature, causing the neutral hydrogen to produce an absorption signal against the CMB. As the density dropped even more, collisions became less and less effective and T_s tended towards T_{CMB} , causing the 21cm signal to disappear. Observations of the global 21cm absorption from the DA will test whether this simple scenario is correct. Deviations from the expected signal will be sensitive to a range of exotic heating processes such as decaying or annihilating dark matter, evaporating black holes, or cosmic strings. The physics behind the standard prediction is sufficiently simple that deviations will provide strong constraints on additional heating processes. The maximum absorption predicted lies around -50 mK.

Observations of the 21cm fluctuations, characterized by the power spectrum will provide unique cosmological data. The power spectrum will be dominated by the density fluctuations, just as for the CMB power spectrum. However, unlike the CMB which comes from a unique epoch, the DA 21cm signal

originates from a range of epochs and therefore contains $\sim 10^6$ more information than the CMB (Loeb & Zaldariagga, 2004). In addition, the DA 21cm signal can possibly probe the matter power spectrum at scales which are damped in the CMB (Loeb & Zaldariagga, 2004). The typical fluctuation amplitudes are predicted to lie in the range 0.1 to 1 mK.

Cosmic Dawn Science

Towards the end of the DA the 21cm absorption weakens because the collisional coupling to the gas temperature disappears. However, the appearance of the first generations of stars helps the 21 cm signal to reappear, due to the Wouthuysen-Field effect: absorption and re-emission of Ly-alpha photons by neutral hydrogen pushes the spin temperature back to the gas temperature. Initially the signal will be again in absorption as the gas temperature will still be below the CMB temperature. However, as X-ray producing sources develop (early X-ray binaries, supernova remnants and accreting black holes), this radiation will heat the IGM above the CMB temperature and push the signal into emission. The detailed timing of these processes is as yet unknown but models indicate that an absorption signal will be seen from about $z \sim 30$ to 15 (46-89 MHz) after which the 21cm signal will be in emission (Pritchard & Loeb 2010). Observations of the global 21cm signal from the CD will establish when the first stars formed and when X-ray heating pushed the signal from absorption to emission, thus establishing among other things the rise of the first accreting black holes. The strength of the absorption signal could be as low as -200 mK, the emission signal has a maximum around 30 mK.

Observations of the 21cm power spectrum of fluctuations will probe the spatial variations in the above processes. Measurements of these variations can establish the distribution of the first stars and first X-ray sources. Substantial variations in the global star formation at this age have been predicted due to supersonic bulk flows in the neutral hydrogen on scales of a few cMpc with large scale variations on scales of ~ 100 cMpc (Tselikhovich & Hirata 2010).

The spin temperature variations complicate the extraction of pure cosmological information from the power spectra. However, due to the presence of redshift space distortions it may still be possible to extract the cosmological matter power spectrum from the 21cm signal (Barkana & Loeb 2006). Models indicate that typical fluctuation amplitudes during the CD range from 1 to 10 mK.

Dark Ages signatures

The Dark Ages signature can typically be divided in three ways (e.g. Jester & Falcke, 2009):

- the Global Dark Ages signal – which is essentially the redshifted 21-cm line absorption feature;
- Tomography of the 21 cm line – essentially imaging the Dark Ages period and the distribution of the Hydrogen and matter in the early universe by observing the 21 cm line at different frequencies and hence different times, in this way forming a movie of the evolution of the early universe;
- Power Spectral analysis of the 21 cm line – performing high-resolution analysis of the spatial variations in distribution of the matter in the Dark ages period.

The Global Signal is expected to peak around 30-40 MHz and is weak, $\sim 10^6$ below the foreground signal. However, Jester & Falcke (2009) show that even with one dipole antenna (under RFI-low and stable temperature and gain conditions) the signal can be detected at a 5σ level within integration times in the order of one year, see Figure 3. With DEX the detection of the Global Dark Ages signal is achievable in the order of days, assuming that a full scale DEX array consists of 10^5 individual elements.

As explained by Jester & Falcke (2009), observing the two-dimensional structure of the neutral and reionized hydrogen gas at different frequencies corresponding to different emission redshifts, will provide a tomographic movie of the Dark Ages showing the evolution of the pristine structures in the early universe. This would require DEX to cover the 1-80 MHz regime with arcminute-scale spatial resolution and the ability to detect cosmological milli-Kelvin brightness fluctuations. In Figure 4 (Left Panel) we show the number of dipole antennas required for a 3σ detection of a 1mK variation in one year of integration time and spatial resolutions of $1'$ or $10'$ for different values of the emission redshift. To be able to reach high spatial resolution ($1'$) for instance at $z=20$ (66 MHz) in the order of 10^7 dipoles are

required and hence a collection area of $\sim 50 \text{ km}^2$ (see Jester & Falcke, 2009). However, for a spatial resolution of $10'$, in the order of 10^5 dipoles are required, but more importantly this corresponds to “only” 0.5 km^2 of collecting area. For $z=50$, where the 21 cm absorption feature is peaking, a collecting area of $\sim 250 \text{ km}^2$ ($10'$, one year integration) is required, so at higher values of the redshift the detection of hydrogen variations becomes increasingly difficult but would benefit from longer integration times and higher gain antennas.

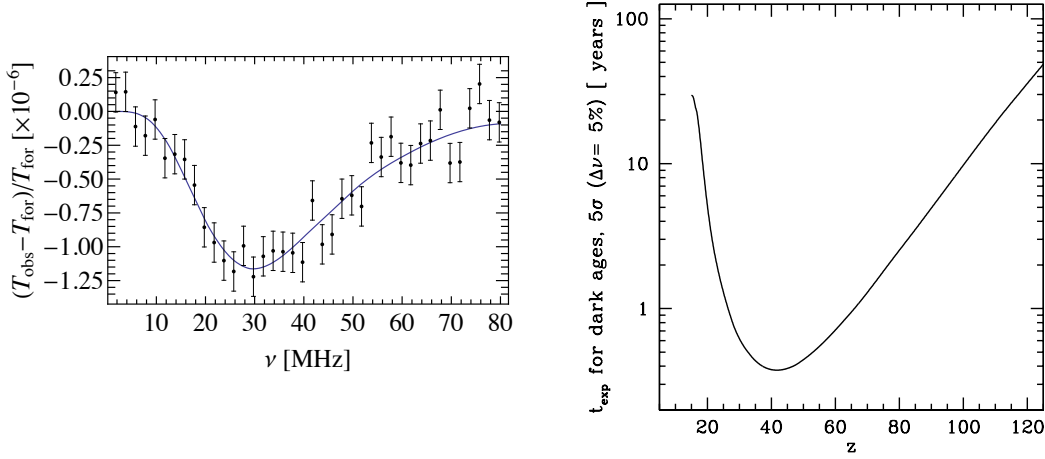


Figure 3: Left Panel: simulation of the 21 cm line after one year of integration for a single sky-limited dipole. Right Panel: Integration times for the 5σ detection of the 21 cm line for different redshifts. Adding N antennas decreases the integration time with \sqrt{N} . Images adapted from Jester & Falcke (2009).

As pointed out by Loeb & Zaldarriaga (2004) the power spectrum of the redshifted 21 cm line provides cosmological information of higher angular scales of $1'$ or less and independent samples of the cosmological parameters in the $z=30$ -50 range, compared to the information carried in the CMB. In the right panel of Figure 4 we show the number of dipole antennas necessary to achieve a $5\text{-}\sigma$ detection of fluctuations in the 21 cm line power spectrum at the mK level for $2'$ angular resolution for one year of integration, as a function of redshift (see Loeb & Zaldarriaga (2004), Jester & Falcke (2009)). As the signal strength is again in the mK level and the noise is determined by the galactic background noise, the sensitivity and hence the collecting area is again the limiting factor here. For a detection at $z=30$ and $z=50$ the number of individual elements is in the order of $10^{5.5}$ and 10^6 , which correspond to 3.5 km^2 and 30 km^2 , respectively. Again, lower collecting area's are required for longer integration times or for lower resolutions. For instance, for $10'$ angular resolution the collecting areas for $z=30$ and $z=50$ correspond to 0.03 km^2 and 0.28 km^2 , respectively (Jester & Falcke, 2009).

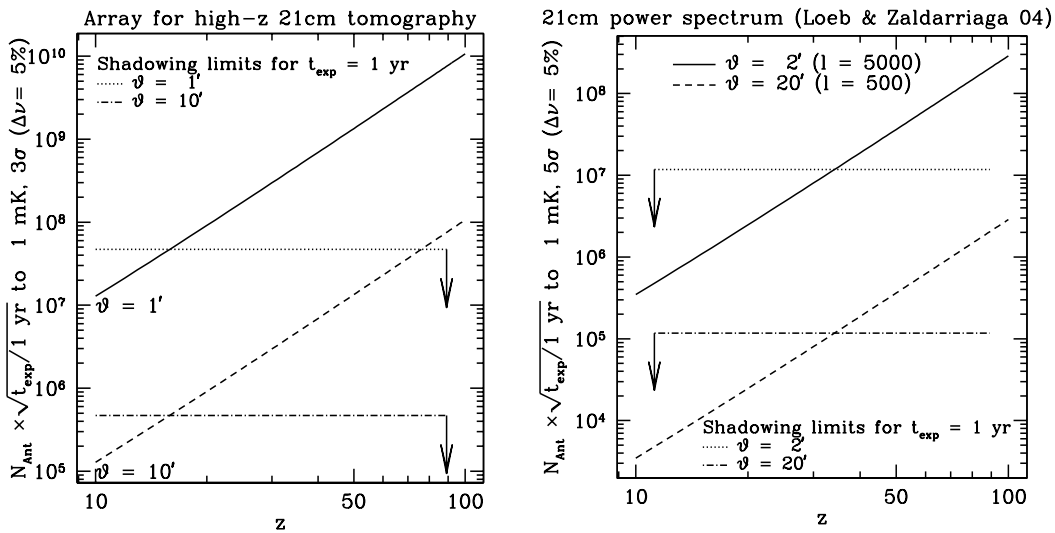


Figure 4: Left Panel: Array size for 21 cm tomography. Right Panel: Array size for Dark Ages power spectrum. Images adapted from Jester & Falcke.

In Figure 5 we show simulated power spectra of the 21 cm line from Pritchard & Loeb (2008); it shows the brightness temperature fluctuations variance integrated over the 3D power-spectrum $P(k)$ in 1 dex shells (hence the $k^2 dk \sim k^3$ factor). It reflects the Dark Matter power-spectrum (as the gas follows the Dark Matter at these redshifts), assuming that the spin-temperature follows the temperature of the adiabatically cooling H-I gas. The dashed line corresponds to the DEX array for $z=80$ (17 MHz; lower panel) and $z=40$ (34 MHz, top panel). Here we have assumed that we are sky-noise background limited, i.e. the dashed lines in Figure 5 represent the DEX array sensitivity. In this particular example we have taken the DEX array to consist of 10^5 individual dipole antennas, which corresponds to a collecting area of 10 km^2 , with an antenna distribution that is constant in a core of 1 km radius and then falls off as $1/r$ until 6 km where it becomes zero. The integration time was set to almost one year and the bandwidth was set to 10 MHz. The resulting field-of-view is all-sky. The blue area in the top panel of Figure 5 corresponds to the effective sensitivity of the DEX array: for $z=40$ it is only limited for the detection of variations on larger scale, by the decrease of strength of the variations ($^{10}\log(k^3 P(k))$ is close to 0) at $^{10}\log(k)=-0.8$. For $z=80$ the DEX array is more limited for detecting variations on all scales as the sensitivity is reduced by about a factor 100, see lower panel Figure 5. From this we conclude that the analysis of the 21 cm power spectra is only meaningful for $z < 80$, hence the lower frequency limit required for the DEX array is $\sim 17 \text{ MHz}$. In addition, Figure 5 shows that for $z=40$ arcmin scale variations can be achieved for an array with a 10 km^2 collecting area. Again, the collecting area can be reduced either by increasing the integration time, antenna gain (i.e. changing the type of antenna) or the required angular resolution.

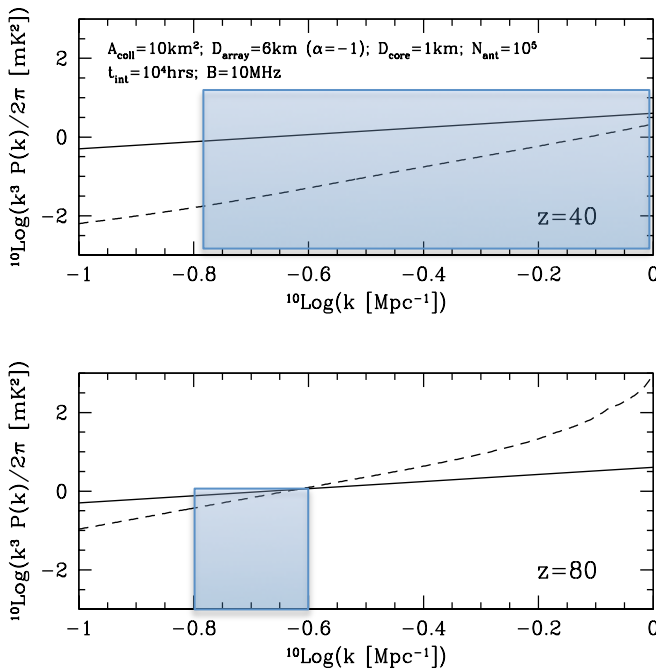


Figure 5: Simulated power spectra of the 21 cm line from Pritchard & Loeb (2008) (solid line) and the expected sensitivity of the DEX array (dashed line) for $z=40$ (top panel) and $z=80$ (bottom panel). The blue area corresponds to the effective sensitivity of DEX and hence the range of power spectral variations that can be achieved: for larger values of $^{10}\log(k)$ (to the left on the vertical axis) the corresponding size scales that one can observe is increasing from Mpc scales for $^{10}\log(k)=0$ to 10 Mpc for -1 .

Secondary science cases

Independently from the cosmological breakthroughs that undoubtedly will be achieved in the Dark Ages science, DEX will open up the virtually unexplored low frequency domain below 30 MHz. Like with preceding science instruments that have gained excess to a previously unexplored frequency domain, it is not exactly known which discoveries DEX will provide, but it is unquestionable that they will be

made. The low frequency regime corresponds to physical processes that occur at low energies (100 neV – 10 peV) and are associated with relatively large physical scale lengths (several meters to hundreds of km). Hence, it will be in these regimes that DEX will provide new insights and possible scientific breakthroughs. In the following we highlight a number of science topics for which DEX is expected to provide new discoveries.

Non-thermal Planetary Radio Emissions

Planetary magnetospheric emissions

The Earth and the four giant planets in the solar system have magnetospheres, where electrons are accelerated to keV-MeV energies by various processes resulting in intense non-thermal low frequency radio emissions in the auroral regions near and above the magnetic poles. Radio emission is produced at the local electron cyclotron frequency by a resonant mechanism that transfers a fraction of the energy of

the electrons to electromagnetic (radio) waves. This Cyclotron Maser instability (CMI) is a most efficient LF radio generation mechanism and operates at all “radio-planets” (Zarka 1998). The spectral characteristics of all auroral radio emissions predicted for the Lagrange point L2 of the Sun-Earth system are depicted in Figure 6. Jupiter with its 14 Gauss surface magnetic field emits up to 40 MHz, while the other planets only emit below about 1.5 MHz. Most of these emissions occur below the cutoff frequency of the Earth ionosphere. The broadband CMI emissions have a complex (anisotropic) morphology in the time-frequency domain and are tied to the local magnetic field (Fig. 3.5; Zarka et al. 2004).

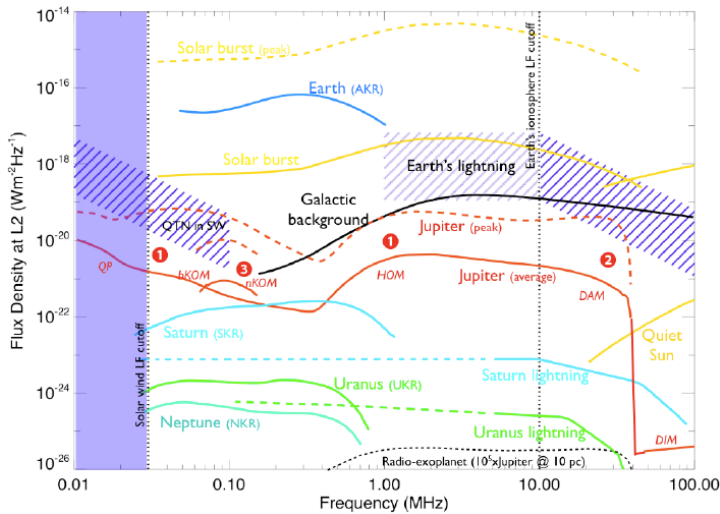


Figure 6: Solar, planetary, and predicted exoplanetary radio emission spectra at the Lagrange point L2 (similar for near-Earth orbit with scaled terrestrial emissions). Numbered features refer to Jovian spectral components depicted in Figure 7: 1) auroral, 2) Io-Jupiter, and 3) Io's plasma torus (Zarka et al. 2008).

Jupiter's radio emissions are intense (MJy), 100% circularly or elliptically polarized, and point-like at a given frequency. Jovian activity is quasi-permanent in the hectometer-kilometer (3-0.3 MHz) range, and reasonably predictable at decameter wavelengths (30 MHz) (Zarka et al., 2004). For all the radio planets, quasi-continuous observations by

DEX will allow study of the time variability of the radio emissions, from short pulses to planetary rotation periods, solar wind and satellite modulation. Accurate planetary rotation periods (and phases) determine the atmospheric wind speed and allow merging longitude-dependent data. Modulations due to natural satellites and solar wind strength show magnetospheric dynamics, solar wind - magnetosphere coupling, and electrodynamic coupling of the magnetosphere with satellites. The radio planets are indirect monitors of the solar wind from 1 to 30 AU.

Jupiter serves as an ideal calibration source and a perfect target for cross-calibration or correlation with ground-based observations. DEX imaging mode observations may be used to resolve Jovian emission spread over several radii (Jovian diameter ~40") and science objectives include (Zarka 2004): coarse fast imaging of (moving) sources in the magnetosphere, beaming of the emission and new constraints on the radio generation process, interaction of satellites and the magnetosphere, plasma torus probing (Faraday rotation), and detecting Io's volcanic outbursts.

Based on Figure 6 DEX system can detect the Jupiter and Saturn magnetospheric radio emissions, while The Uranus and Neptune emissions will be detectable when they are in a high state.

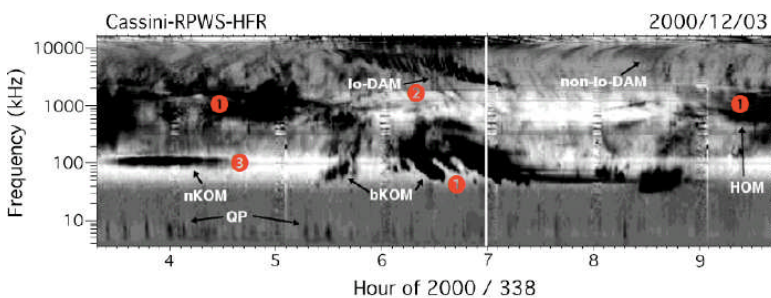


Figure 7: Jovian emissions observed by the RPWS experiment onboard Cassini between 3.5 kHz and 16.1 MHz. Jovian spectral components numbered as in Figure 6 (Zarka et al. 2004).

Exoplanets

Coherent planetary emissions such as Jupiter's are nearly as intense as the

radio bursts from the solar corona. Hot Jupiter exo-planets (orbiting at a few stellar radii from their parent star) are predicted to have radio emission up to 10^5 - 10^6 times the flux of Jupiter if they are strongly magnetized or orbiting a strongly magnetic star or are bombarded by numerous Coronal Mass Ejections from their parent star (Zarka 2007; Greissmeier et al. 2007; Chian et al. 2010). Conservative emission levels are indicated in Figure 6. Hot Jupiter exoplanetary radio emission will be detectable in DEX imaging surveys using long integrations (≥ 12 h) if the emission covers a large bandwidth over

timescales of hours. A premier detection of extrasolar magnetospheric radio emission would confirm the predictions of hot Jupiter sources and would initiate further study of their properties.

Solar Physics and Space weather

Our Sun is a very strong radio source: superimposed on the thermal emissions of the quiet sun are the intense radio bursts that are associated with solar flares and coronal mass ejections (CMEs), clouds of ionized plasma ejected into interplanetary space. Three main types of radio bursts are observed from the Sun particularly in its active state, both related to flares and CMEs. Type II bursts have a frequency drift with time at rates consistent with the speed of the shock through the solar corona and interplanetary medium ($\sim 1000\text{--}2000$ km/s). Type III bursts are emitted by mildly relativistic ($\sim 0.1 - 0.3$ c) electron beams propagating through the corona and interplanetary space that excite plasma waves at the local plasma frequency. Their frequency drift rate is much higher than that of Type II bursts. Type IV bursts are emitted by energetic electrons in the coronal magnetic field structure (such as coronal loops). Both the Type II and III bursts can be imaged by DEX in the 1-30 MHz range.

The density model of the heliosphere (Mann et al. 1999) directly relates the radio source location (in solar radii) to the emission frequency: higher frequency radio emission originates closer to the surface of the sun, while lower frequency emission originates further out. By providing dynamic spectra and detailed imaging of the solar radio emissions, DEX will allow monitoring and modeling of plasma instabilities in the solar corona and wave-particle interactions in the activity regions of the Sun. In addition, DEX offers great opportunities for radio studies of the solar wind and the heliosphere. It will permit observations of solar radio bursts at low frequencies with much higher spatial resolution than possible from any current or planned space mission. It also allows observations much further out from the solar surface than possible from the ground, where the ionosphere confines the field of view to within a few solar radii. DEX will dynamically image the evolution of CME structures as they propagate out into interplanetary space and potentially impacts on the Earth's magnetosphere.

Transient phenomena

Pulsars are among the best-studied cosmic radio sources. They are quite dim below 30 MHz, and are strongly affected by interstellar dispersion and scattering, but DEX is expected to allow detection of roughly 1% of the 2000 currently known pulsars. This will allow for groundbreaking research into low-frequency pulsar emission, which is not understood. One of the main open questions is the intrinsic bandwidth of pulsar emission. Earth-based observations cut out below 16-18 MHz; only a space-based mission such as DEX can illuminate the low-frequency behavior. Month-long TBM mode integrations can produce folded, coherently de-dispersed profiles for the dozen brightest nearby pulsars in the sky, very similar to the highly successful Fermi gamma-ray mission (Figure 8). These observations will also shed light on the expected change in scattering properties of the ISM at low frequencies, which could be explained by a cutoff in the Kolmogorov turbulence spectrum of the ISM.

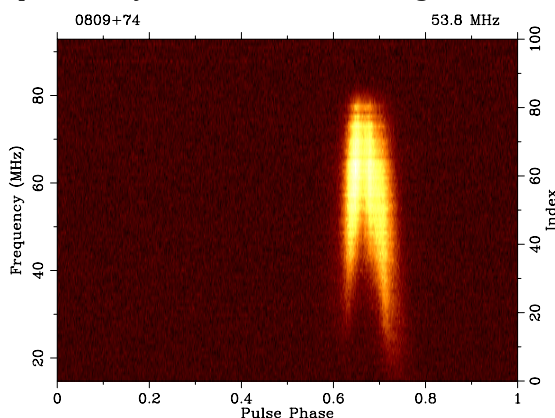


Figure 8: A 1-hr observation of pulsar B0809+74 using a coherent addition of all 24 LOFAR low-band core stations from 15–93 MHz. The data has been de-dispersed and folded using a rotational ephemeris to produce a cumulative pulse profile as a function of frequency. Given that the central observing frequency is 53.8 MHz, the fractional bandwidth is 145%. This wide bandwidth is key to following the drastic evolution of the cumulative profile with frequency. At the bottom of the band there are two distinct pulse components that almost completely merge toward the top of the band (van Haarlem et al. 2013)

The transient radio emission from Gamma-ray bursts (GRBs), supernovae, as well as from accreting black holes, neutrons stars and white dwarfs are all enabled by coherent emission processes. For many of these and similar sources, the emission above 30MHz cannot complete their energy balance. DEX will be the optimal instrument to detect any energy release below 30 MHz, and thus can sample a large discovery

space in RBM or TBM mode observations for fast transients and ASI mode observations for slow transients.

Extra-galactic surveys

high-redshift galaxies and quasars

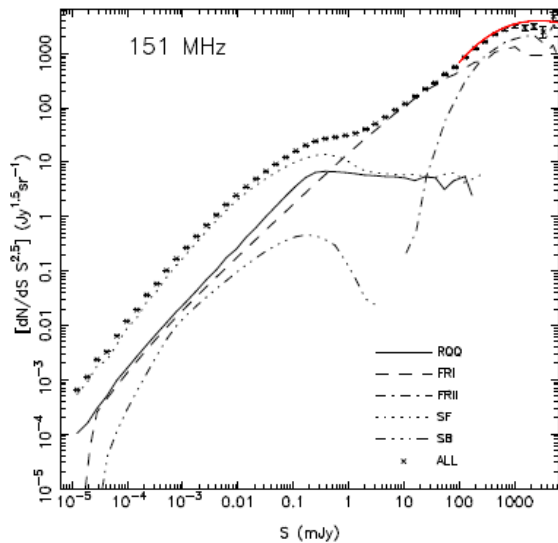
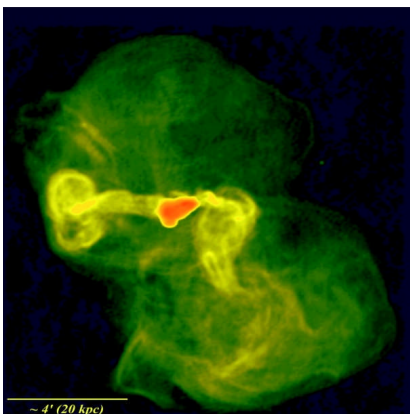


Figure 9: Source counts simulations at 151 MHz for different source populations. The red curve shows the currently known counts (Wilman et al. 2008). In the equivalent detection curves at low frequencies, DEX will detect sources down to 200 mJy at 10 MHz and 10 mJy at 30 MHz.

At low radio frequencies, the brightest sources of the extragalactic sky are the nuclei and synchrotron-emitting radio lobes of giant radio galaxies and quasars such as those found in the 3CRR survey (Laing et al. 1983). Multi-frequency radio observations have shown that sources with ultra steep spectra ($\alpha < -1.3$, where $S \sim \nu^\alpha$), i.e. relatively bright in the DEX range, are usually hosted by high redshift galaxies (Röttgering et al. 1997). A simple extrapolation from source counts at 74 MHz in the VLSS (VLA Low-Frequency Sky Survey; Cohen et al. 2004) suggests that a DEX survey will discover (at 3 σ significance) ~ 2 million sources above the confusion

limit of 210 mJy at 10 MHz and of 40 mJy at 30 MHz for a survey time of one year. These source populations occupy the upper end of the flux density distribution similar to the curve depicted in Figure 9. Most of these sources will have relatively steep spectra characteristic of more extended sources rather than the flatter spectra of AGN sources. The survey results are essential for study of the evolutionary history of the different source populations (as in Figure 9). The source catalog will be used to study the spectral properties of specific groups of sources and relate these to their peculiarities.

The source populations that will be detected in such a low frequency survey consist mainly of standard radio galaxies and quasars but will also include other active galaxies (Figure 9). A special class of sources consists of very compact start-up radio galaxies having an active black hole with jets still confined within the recently activated galaxy. Such start-ups are also seen locally (O’Dea 1998) and display a synchrotron self-absorption spectral turnover in the range 100-1000 MHz (Falcke et al. 2004). Some models suggest these to be numerous at redshift 20, where they peak between 5-50 MHz and will be readily identified by a DEX all-sky survey.



Jet power of radio galaxies

Figure 10: The famous radio Galaxy M87 at a frequency of 74 MHz showing the enormous extent of the lower frequency emission. At higher frequencies only the (orange) central region will be detected (Owen et al. 2000).

The low-frequency emission of radio galaxies and quasars are generated predominantly in the radio lobes. These structures with scales of tens to hundreds of kpc (Begelman et al. 1984) are inflated by relativistic jets symmetrically emerging from the accretion disk around the nuclear black hole of the host galaxy. The lowest-energy electrons in the jets and lobes of these radio galaxies dominate the synchrotron radiation output, which yields the observed steep radio spectra. The shorter cooling time of the higher-energy electrons causes further steepening (spectral ageing) of the radio spectrum in the older parts of a source (Blundell & Rawlings 2001). At the location of the terminal shock front of powerful jets, the “hotspots”, the plasma can become optically thick to its own synchrotron emission due to self-absorption. As a result the hotspots frequently show a spectral turnover at frequencies as low as 10 MHz.

From their full size and the turnover frequency, one can deduce reasonable limits of the magnetic fields and mechanical work done by the jet feeding it. In current models of galaxy and black hole formation and evolution (Croton et al. 2006; Hopkins et al. 2007), this AGN feedback is key to regulating both the growth of the accreting black hole and the star formation activity in its host galaxy, giving rise to the old, red stellar populations of massive elliptical galaxies (e.g. Best et al. 2006). DEX observations will constrain the power of the relativistic jet outflows, their physical models, and their interaction with the ISM.

Fossils and AGN feedback

Observations below 30 MHz also probe the oldest structures in a radio galaxy and they provide constraints on their age and the feedback history of the central engine. Radio galaxies and other AGN sources are normal galaxies going through a temporary phase of more or less strong accretion onto their central black hole (Croton et al. 2006). A fundamental issue for interpreting such activity phases is their duration and recurrence timescales. Only for the so-called “double-double” radio galaxies with two sets of double radio lobes (Schoenmakers et al. 2000), this recurrence is directly visible. From synchrotron ageing arguments for these rare sources, both the duration of an activity phase and the recurrence timescale are likely within an order of magnitude of ten million of years.

The spatial extent of (currently) active radio galaxies is much larger at lower frequencies than the structures observed at higher frequencies (600 MHz and above). The gas injected into the lobes by the jets eventually ‘cools’ and expands into the surrounding medium, creating a low-frequency emission halo as observed in M87 (see Figure 10). However, the emitted power of a radio galaxy decreases as its lobes grow in linear extent. Therefore, only the fairly young radio galaxies (ages $\sim 10^5$ - 10^7 yr) are observable at high redshifts ($z \gtrsim 0.8$), while at low redshifts they would be observable up to ages of 10^8 - 10^9 yr (Blundell et al. 1999). Because of this degeneracy, a DEX all-sky search for halo remnants or “fossils” of switched-off (nearby) radio galaxies, will directly address the lifetime and recurrence issues of nuclear activity in radio galaxies.

The only known radio galaxy relics are cool holes in the X-ray emission of the hot cluster gas (see Fabian et al. 2006). This intra-cluster medium (plasma) (ICM) emits (radio) synchrotron radiation, as well as X-rays through inverse-Compton scattering. Existing (and future) all-sky radio observations miss such sources because their observing frequencies are too high to detect these fossil emissions. The time interval for detection after switch-off roughly increases as the inverse of the observing frequency, which is good for a DEX survey. The mechanical energy input into the ICM resulting from AGN activity is important for understanding galaxy formation and ICM thermodynamics. ICM bubbles blown by central engines and ‘ghost bubbles’ resulting from radio galaxy relics can be detected in X-ray emissions of the nearby Centaurus and Perseus clusters (Fabian et al. 2005, 2006). Such bubbles rise buoyantly in the ICM and are part of the energy transport mechanism in clusters. Analogous to the case of radio galaxies, low-frequency radio observations of clusters will also constrain their energetics, their magnetic fields, their formation history, and the nature of cosmic rays.

The DEX survey observations will discover (hundreds of) thousands of steep-spectrum cluster halo sources during its lifetime, preferentially those of low-mass and high-redshift objects (Cassano et al. 2008). In targeted observations, DEX would detect radio emission from these bubbles and yield stringent constraints on their ages.

Cosmic Rays and High-Energy Neutrino’s

Radio emission observed towards Galactic HII regions at low frequencies predominantly arises from material along the line of sight, i.e., the synchrotron emission from cosmic-ray electrons (Duric 2000). Similarly, supernova remnants (SNRs) accelerate high-energy particles through a first-order Fermi shock acceleration mechanism, which only operates on mildly relativistic electrons. Since the frequency of synchrotron emission scales with the particle energy, low-frequency DEX all-sky imaging of visible HII and SNR targets can be used to trace the 3D energy distribution of the lowest-energy particles.

DEX strawman mission concepts

For the definition of the DEX mission concepts we have to take a number of requirements into account that drive the design. The dominant requirements in that respect are the sensitivity and hence collecting area ($\sim 10\text{km}^2$), temperature and gain stability, and RFI-quiet environment. For this reason we consider here two options: a space-based mission and a lunar-farside mission. Although there are essential differences in these two designs, there are also clear similarities. We will therefore first describe a basic (strawman) design that applies to both the space-based and lunar design, and then highlight some of the mission specific design characteristics.

The basic DEX design is that of a low-frequency radio interferometer. DEX will employ the well-tested interferometry techniques to achieve its science objectives. The sensitivity of a radio interferometer with N antenna elements depends on the total number of baselines (relative distance vectors) between pairs of separated antenna elements and scales as $[N(N-1)]^{-1/2}$. The signals from each pair of elements are cross-correlated and integrated, yielding one (u,v,w) point in the three-dimensional spatial Fourier domain for each sampled baseline and frequency channel. These “snap-shot” correlation data are sent to Earth for offline processing and imaging. For a (to be designed) DEX imaging mode, the bandwidth of the spectral channels is chosen such that the narrow-band condition holds ($\Delta f \leq 1\text{kHz}$), thus allowing all-sky imaging. The actual all-sky image is obtained for example by concatenating many small FOV image patches, each obtained from an inverse Fourier transform of the set of sampled (u,v,w) points. The spatial (u,v,w) points for each baseline yield independent (u,v,w) points for each frequency channel (“bandwidth synthesis”). The motions of the individual antennas in space or the rotation of the moon (with respect to the sky), increases the number of unique baselines during the course of the mission, which gradually fills the (u,v,w) volume and greatly improves the quality of the final sky image. In addition, for each combination of three or more elements in the array, self-calibration techniques may be applied that significantly improve the imaging quality. Hence, many elements, large bandwidths, and long integration time generally improve the imaging performance of the interferometer. A second performance figure of an interferometer is its angular resolution, which scales as λ/D , where λ is the observed wavelength and D is the separation distance between antennas. For DEX, angular resolution in the order of 1-10 arcmin will be required for the frequency range between 1-80 MHz. Note that, although the Dark Ages science is limited to the 17-80 MHz regime, we have chosen DEX to be sensitive in the 1-80 MHz regime to allow for additional science (i.e. auroral emission from planets, see Figure 6).

The DEX science objectives will be addressed using three basic operational modes:

- **Wide Band Spectroscopy mode:** runs continuously in parallel with any of the other modes, and produces one time averaged complex spectrum at full bandwidth every 5 minutes. This mode is particularly designed for study of the highly redshifted 21-cm signal during the Dark Ages over the full 0.1-80 MHz range.
- **All-Sky Imaging mode:** employs the imaging capability of the full antenna array and allows imaging of the whole sky at the spatial resolution afforded by the array. The nominal bandwidth correlated by the system is 10 MHz and the centre frequency of the observations can be selected across the whole 70 MHz frequency range of the system. The complete auto- and cross-correlation matrix data will be produced as 1-10 sec records for further calibration and image processing. Correlation is done in narrow-band frequency channels, thus allowing all-sky imaging. Ground-based processing of the cross-correlation data will allow calibration and imaging of the whole sky, while the auto-correlation data provides the signal integrated over the whole sky.
- **Burst mode:** employs beamforming to phase up the array for a particular target area. Spectral data will be produced at significantly faster sampling time as low as 50 ms for further ground-based processing. This mode will be well suited for sensitive observations of pulsars, transients, and variable planetary emissions.

In Figure 11 we provide the data acquisition system design that will be very similar for both the space- and moon-based option, as they address the same science. The difference between these two options lies in the way the data is processed and transformed, as will be discussed in the next sections.

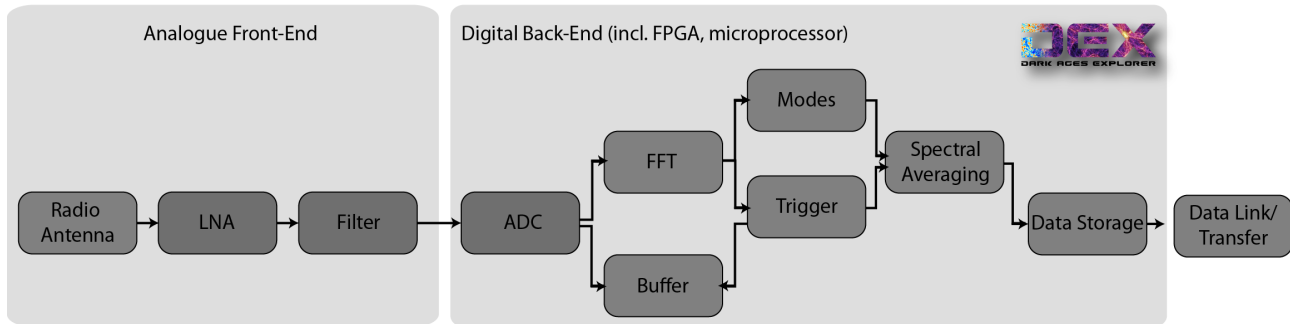


Figure 11: DEX data acquisition design. The modes correspond to the science operational modes as described in the text.

Space-based mission: Swarms of nano-satellites

The main design considerations for an astronomical low-frequency array in space relate to the physical characteristics of the interplanetary and interstellar medium. The configuration of the satellite constellation and the achievable communication and processing bandwidths in relation to the imaging capabilities are also crucial design considerations. The system will consist of a swarm of 10^5 identical satellites (sensors) spread over kilometric distances that will orbit faraway from terrestrial radio frequency interference. The distributed solution should provide redundancy and robustness, as it is insensitive to failure or non-availability of a small fraction of its components.

Using current-day technologies, a space-based low-frequency array would be bulky and, thus, costly. A logical next step would be to miniaturize the electronics and use very small satellites, perhaps even nano satellites with masses between 1-10 kg to build the radio telescope. The approach is to use a swarm of satellites to establish a virtual telescope to perform the astronomical task. In recent studies, such as DARIS (Saks et al., 2010) and FIRST (Bergman et al., 2009) it is shown that with extrapolation of current signal processing and satellite technologies, a low frequency radio telescope in space could be feasible in the coming years. DARIS has already shown that a 9-satellite cluster, with a centralized system can be implemented in moon orbit with today's technology.

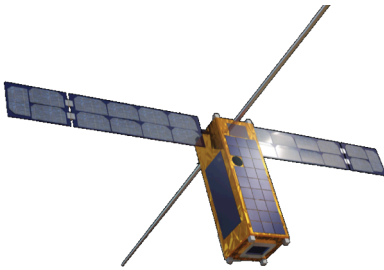


Figure 12: OLFAR antenna and platform concept; i.e. one swarm element.

To enlarge the cluster, the satellites must be smaller. The OLFAR project (Bentum et al., 2009; Rajan et al., 2011; Engelen et al., 2010) aims to develop a detailed system concept and to design and build scalable autonomous satellite flight units to be used as an astronomical instrument for low frequencies. To achieve sufficient spatial resolution, the minimum distances between the satellites must be more than 10 km and due to inter stellar scattering this maximum baseline is limited to 100 km, giving a resolution of 1 arc minute at 10 MHz. The OLFAR 3-dimensional cluster will comprise of 50 - 1000 satellites, each containing a dipole (or tripole) antenna (Figure 12), observing the sky from 0.3-30MHz. The satellites will employ passive formation flying and yet maintain sufficient position stability for a given integration time. In the presence of a stable orbit and thus stable baseline, position estimates can be more precisely known and thus the integration time can be extended up to 1000 seconds and thereby reducing the down-link data rate. For astronomical observations, mechanical dishes will be very expensive in terms of mass, power and operation. Instead OLFAR satellites will use a relatively simple antenna.

Each individual satellite will consist of deployable antennas. The sky signals will be amplified using an integrated ultra-low power direct sampling receiver and digitizer. Using digital filtering, any subband within the LNA passband can be selected. The data will be distributed over the available nodes in space. On-board signal processing will filter the data, invoke RFI mitigation algorithms (if necessary), and finally, correlate the data in a phased array mode. If more satellites are available, they will automatically join the array. The final correlated or beam-formed data will be sent to Earth as part of the telemetry data using a radio link. As the satellites will be far away from Earth, communication to and from Earth will require diversity communication schemes, using all the individual satellites together.

Based on previous designs for space-based, low-frequency, interferometry missions such as DARIS (Daris, 2010), SURO, HEIMDAL, DARE (Burns et al. 2012) and OLFAR (Olfar 2012) we summarize the DEX space-based mission design as follows:

- The individual antennas are omni-directional tripoles (active antennas), with their sensitivity optimized in the 1-80 MHz regime. For the secondary science cases the lower limit of this range can possibly be extended downwards to 0.1 MHz.
- The antennas are mounted on nano-satellites that provide power (solar panels) and basic processing (FFT and averaging algorithms) and communication.
- The array is build up from these individual antennas and together with a mothership that provides data storage, processing and communication with Earth, it acts as a swarm. This means that the individual antennas are interchangeable, and that each antenna should know its position but that the array is not controlled from a central hub. Due to the motion of the individual antennas, the baselines are constantly changing and complete coverage of the (u,v,w) plane is achieved.
- The array should be placed in a RFI-quiet location, for instance at the Sun-Earth L2 point or in an Earth-leading or Earth-trailing orbit. This will minimize the need for RFI-mitigation techniques.
- Each individual antenna should be calibrated carefully and the noise pattern (EMC) should be characterized in detail.
- Data processing can be done at a central mother-ship which has dedicated data processing facilities, more power available (larger solar sails) and large data rate available for data transfer, the individual antenna then have to transfer the data to the mothership and have to perform some on-board processing to reduce the data volume somewhat;
- Alternatively, data processing and transfer can be arranged by the array of antennas, i.e. they act as a swarm. This has the advantage that with the increase of the array the processing power is increased, but also requires each individual element to be provided with significant processing and communication capabilities;
- For the communication between the individual antennas and the mothership, optical communication and nano-photonics can be used;
- Optionally, the individual antenna can be based on inflatable space structures, see the discussion below.

Inflatable space structures

DEX requires a significant collecting area (10 km^2) and while the radio antenna and receiver technologies are well developed (TRL levels of 6 and higher), bringing them into space or to the moon is a costly and technologically challenging endeavor. In order to reduce the weight, the mechanisms required for the deployment and hence the costs of such a mission, we suggest the use of inflatable space structures. Already since mid 20th century inflatable space structures are being used to bring sizable structures in space, and the first two missions launched by NASA in 1960, Echo 1 and 2, were successfully used as communication reflectors for transmission of radio signals (telephone, radio and television signals) from one point to another on Earth. For this reason they were operated in the MHz regime. At the end of the 20th century interest in inflatable space structures was renewed which resulted in the successful deployment of the NASA Inflatable Antenna Experiment (IAE) from space shuttle mission STS-77 in 1996 (Freeland et al., 1997). In addition, a number of mission concepts were proposed, for instance QUASAT which is an ESA/NASA concept for a free-flying VLBI antenna with a spaceborn 15m reflector which operates at 1.6, 5 and 22 GHz. For DEX the inflatable space structure technology can be adapted for the platform or deployment of the antenna system, comparable to the IAE concept, see Figure 13.

As an alternative for the traditional tripole radio antenna design, inflatable balloon structures can be considered. For instance, the 10 km^2 required for DEX can be obtained by forming an array of 300 Echo 2 antennas. Given the current development in the commercial space flight (SpaceX, Bigelow Aerospace, Virgin Galactic, XCOR) the deployment and realization of DEX is becoming feasible in the near future from a technological and financial point-of-view.

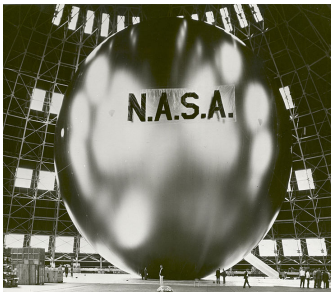


Figure 13: Left Panel: Echo 2 satellite (1960). The balloon was made of metalized PET film, had a diameter of 41.1 m and a mass in the order of 180-200 kg. Right Panel: the NASA IAE launched from the space shuttle in 1996.

Moon-based mission: LOFAR on the moon

The moon, and in particular the lunar far-side, has already for many years been considered to be the most ideal location for low-frequency radio astronomy (e.g. Basart & Burns, 1990). Not only has

the moon no atmosphere or significant ionosphere it can also provide significant attenuation (40 dB or more) of man-made RFI signals and locations with stable temperature and gain conditions. In the past several lunar mission concepts have been suggested, some proposing more “traditional” designs using tripole antennas such as the LRX instrument on the European Lunar Lander, see Klein Wolt et al., 2012, while others propose to use innovative and lightweight designs, such as the ROLSS mission proposed by NASA (see Burns et al. 2009; Lazio et al., 2011). But these concepts differ also in other respects. While LRX consists of one single, omni-directional active antenna (kHz-100 MHz range), the ROLSS concept consists of 500m long arms forming a Y-shape on the lunar surface, with each arm having 16 antennas. The arms are made of thin polyimide and the antenna is operational in the 1-10 MHz regime reaching less than 2° resolution at 10 MHz. In addition, note that both concept are addressing different science. In any case, both the LRX and ROLSS should be considered path-finder missions for future large arrays, such as the Dark Ages Lunar Interferometer (DALI, see Lazio et al. 2007) which is a concept very similar to DEX.

The DEX lunar concept design on the one hand draws from the previous proposed lunar mission concepts, but on the other hand will be based on the experience and expertise gained with Earth-based low-frequency interferometers such as LOFAR and in the future SKA. In particular the experience in the technical realization of a large collecting area radio array in (often) remote locations, calibration of the instruments, RFI mitigation techniques and the handling and processing of large data volumes is essential for the development of DEX. In short, the DEX lunar interferometer design has the following characteristics:

- Interferometer array, consisting of $\sim 10^5$ individual antenna elements together realizing a ~ 10 km² collecting area on the lunar surface, sensitive in the 1-80 MHz frequency regime.
- The individual antenna elements can consist of traditional dipole or tripole antennas (e.g. LRX, see Klein Wolt et al. 2012) or be placed on thin metal sheets that are rolled out on the surface (ROLSS, see Lazio et al. 2011).
- The location should be chosen carefully in order to provide temperature and gain stability as well as attenuation of RFI signals from Earth and Solar activity, preferred is the lunar far-side but South- or North Pole locations using shielding from mountains is also an option. Note that while the lunar far-side is preferred, it does require an additional orbiter for communication to the Earth.
- At a Lunar far-side location the data processing and communication should be done from the orbiter: it provides more power from solar panels and has better communications with Earth available compared to stationary platforms on the lunar far-side.
- Each individual antenna should be calibrated carefully and the noise pattern (EMC) should be characterized in detail.

Long-term focus

The realization of a 10 km² DEX interferometer in space or on the moon is a significant technological challenge that should be approached in a step-wise fashion and depends on future developments in the areas of space transportation, light-weight inflatable space structures, data communication and low-power-high-performance data processing and swarm technologies. Both concepts presented are based on the same science case and have communalities with respect to the data processing and operational modes, but also are equally flexible and scalable. In both cases individual elements can be added to the array which would increase the sensitivity and science output, and in some cases also the total data

processing capabilities. Starting a small-size DEX array, as a path-finder mission, would also immediately provide valuable science output; not only would the DEX path-finder open up the last unexplored frequency domain, but it would allow for the detection of the global Dark Ages signal as well address many of the secondary science cases (planetary radio emission, all sky survey etc).

As mentioned above, a lunar location is much preferred from a scientific point of view but also from an explorational point of view. The realization of a space-based DEX array at for instance Sun-Earth L2 could be considered as a single-shot mission that would be difficult to “share” with other potential users, while a DEX array on the moon would provide a basic infrastructure that would allow other scientific or explorational benefits and could be considered as a long-term investment. Note that, a lunar location would also allow DEX to address a number of additional science that was not mentioned here, for a detailed overview see Klein Wolt et al., 2012.

Finally, scientifically the DEX interferometer is the natural next step. On the one hand it is the successor of the earth-based low frequency interferometers such as LOFAR and SKA, opening up the last unexplored frequency regime for astronomy, which is *not* possible from earth-based antennas. On the other hand DEX's unique and unprecedented detailed view of the evolution and structure formation of the very early universe (Dark Ages and Cosmic Dawn) will help cosmologist to constrain the models and predictions that are coming from the WMAP and PLANCK missions that study the CMB.

TRL levels

Here we provide an estimate of the expected increase in the TRL levels in the 2013-2030 timeframe for some of the key technologies that will be used by DEX.

Technologies	Current TRL	Expected TRL @ L2- L3	Expected critical developments
Radio Antenna (lightweight, foldable, inflatable)	5	6-7	Small, foldable light-weight structures are being designed in the OLFAR and ROLSS project that should fit a nano-satellite.
Radio Receiver (low-power, high processing, 200 MHz receivers)	5-6	6-7	Prototype radio receivers are expected to be tested in rocket flights in the near future, and similar systems will be tested in space environments (e.g. ISS). Further heritage is gained from ground-based low-frequency instruments such as LOFAR, SKA, MWA, LWA and space-based instruments such as LRO
Digital processing system	4-5	6-7	Development of power-saving and smart algorithms to processes large quantities of data with significantly less power are currently ongoing in many Big-Data Science projects (CERN, ITER, LOFAR, SKA)
Optical communication	6-8	7-9	Optical communication and nano-photonics are expected to be employed in space industry (telecom) and has been tested since the 1970s (e.g. SILEX on ESA Artemis)
Swarm Technologies	3-4	5-6	OLFAR, inter-satellite communication, satellite control
Thin-film solar panels	4-5	7-8	Currently thin-film solar panels are considered for future missions with expected launch dates before L2 and L3
Radio Frequency interferometry	7-9	7-9	Based on space-ground interferometers (HALCA and RadioAstron), and there's been some crude time-difference-of-arrival measurements (interferometric-like) using the THEMIS or Cluster spacecraft.
Inflatable space structures	7-8	8-9	NASA Echo1 and 2 missions, NASA inflatable antenna experiment (IAE, 1996)

Conclusions

The Dark Ages eXplorer, DEX, is a low-frequency radio interferometer in space or on the moon, that will probe deep into the early universe and provide an unprecedented detailed view of the evolution of the first structures in the universe. DEX is sensitive in the 1-80 MHz regime which allows a complete view of the Dark Ages and Cosmic Dawn period, essentially covering the $z=17-80$ redshift regime (Figure 14). DEX will explore the Dark Ages and the Cosmic Dawn and observe the global neutral hydrogen (21 cm) emission and its variations on arcmin scales in order to constrain cosmological models on the evolution of the early universe, the onset of the epoch of reionization and basically constrain the models and predictions that will follow from the Planck mission and address the following science questions:

- SQ1 When did the Cosmic Dawn occur and which sources are responsible for the onset of the Cosmic Dawn?
- SQ2 How did the first sources form?
- SQ3 What was the impact and feedback of the first sources on the IGM?
- SQ4 What is the spectrum of the Global Dark Ages signal and how does it change over time / redshift?
- SQ5 What is the tomography of the Global Dark Ages signal?
- SQ6 What is the scale of the density fluctuations in the early universe?
- SQ7 How does the Dark Matter couple to the IGM/gas?
- SQ8 What is the cooling rate of the early universe (Dark Ages)?

These issues form the *holy grail of cosmology* and the Dark Ages is the treasure-trove for Dark Matter and Early universe physicists.

In addition, DEX will open up the last virtually unexplored frequency regime below 30 MHz, extending the view of LOFAR and SKA to the ultra-long wavelength regime that is not accessible from Earth, and among a wealth of science cases will provide high resolution low-frequency sky maps, constrain models on the jet power in radio galaxies, observe auroral emission from the large planets in our solar system and possibly find Jupiter-like exoplanets.

DEX requires a large collecting area in the order of 10 km^2 (10^5 individual elements) and a location (preferably the lunar far-side) that provides shielding from man-made radio interference (RFI), absence of ionospheric distortions, and high temperature and antenna gain stability. The realization of large collecting area interferometers has been achieved on earth already (e.g. LOFAR and in the near future SKA) and the technique behind it is well developed (TRL 4-5). The realization in space or on the moon is a very challenging task. However, we are convinced that given the technology developments, especially in the areas of nano satellites, RF technology and low-power and high-performance computing, the construction of a large low-frequency radio array in space or on the moon is feasible in the 2020-2030 timeframe.

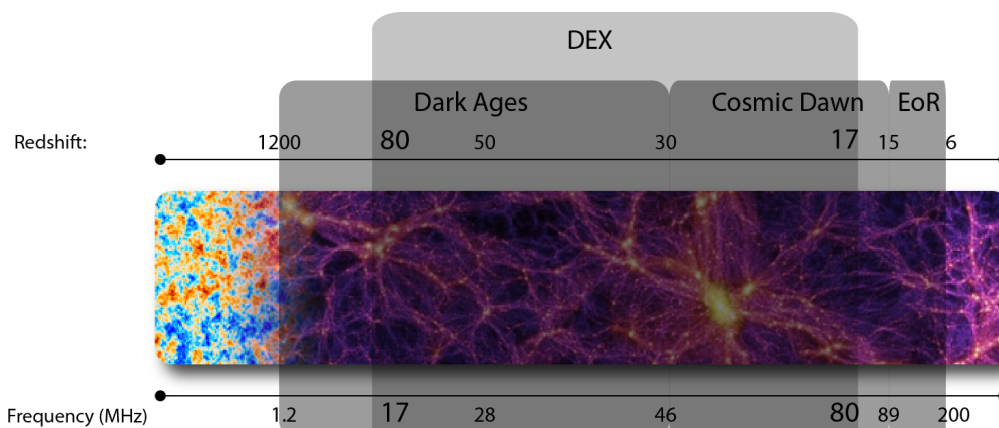


Figure 14: The DEX frequency regime. The lower limit of 17 MHz is only for the Dark Ages science. We have chosen the 1-80 MHz band to allow for additional science.

References

- Barkana & Loeb, 2006, MNRAS, 372, 43
- Begelman, M.C., Blandford, R.D., & Rees, M.J., 1984, Rev. Mod. Phys. 56, 255
- Bentum et al., 2009, *60th International Astronautical Congress, Daejeon, Republic of Korea, 12 – 16 October, 2009*.
- Bergman et al., 2009, ArXiv e-prints
- Best, P.N. et al., 2006, MNRAS, 368, L67
- Blundell, K.M., Rawlings, S., & Willott, C.J., 1999, AJ 117, 677
- Blundell, K.M. & Rawlings, S., 2001, PASP Conf 250, 363
- Braude et al., 2002, IAUS, 199, 490
- Burns et al., 2009, AAS, 213, 451.03
- Burns et al., 2012, AdSpR, 49, 433
- Burns, J.O. et al. 2012, AdSpR 48, 1942
- Cassano et al., 2008, A&A, 480, 687
- Cane & Whitham, 1977, MNRAS, 179, 21-29
- Cane & Erikson, 2001, Radio Science 36, 1765-1768
- Carilli et al., 2007, Proceedings of the workshop "Astrophysics Enabled by the Return to the Moon". Cambridge University Press, in press astro-ph/0702070
- Chian, A.C. et al., 2010, Adv. Space Res., 46, 472
- Ciardi, B., & Ferrara, A. 2005, Space Science Rev., 116, 625
- Cohen, A.S., et al., 2004, ApJS, 150, 417
- Croton, D.J., et al., 2006, MNRAS, 365, 11
- DARIS, 2010, *Very Large Effective Receiving Antenna Aperture in Space*, ESTEC/Contract 22108/08/NL/ST
- Duric, N., 2000, in *Radio Astronomy at long wavelengths*, Stone, R.G., et al. (Eds.), 277
- Ellis & Mendillo, 1987, Australian Journal of Physics 40, 705–708.
- Engelen et al., 2010, *24th Annual Conference on Small Satellites, Utah, USA, August 9-12, 2010*
- Fabian, A.C., Sanders, J.S., Taylor, G.B., & Allen, S.W., 2005, MNRAS 360, L20
- Fabian, A.C., et al., 2006, MNRAS 360, L20
- Falcke, H., K rding, E., & Nagar, N.M., 2004, NewAR 48, 1157
- Farrell et al., 1999, JGR, 14025-14032
- Freeland et al., 1997, IAF-97, 1.3.01
- Griessmeier, J.M., et al. 2007, PSS, 55, 618
- Hopkins, P.F., Richards, G.T., & Hernquist, L., 2007, ApJ, 654, 731
- Jester, S., & Falcke, H., 2009, NewAR, 53, 1
- Kassim & Weiler, 1990, Proceedings of an International Workshop, Crystal City, VA, Jan. 8, 9, 1990.
- Klein Wolt et al., 2012, P&SS, 74, 167
- Kraus, 1986, "Radio Astronomy Receivers", Powell, Ohio: Cygnus-Quasar Books, 1986
- Lazio et al., 2011, Advances in Space Research 48, 1942–1957.
- Lazio J. et al., 2007, AAS, 39, 135
- LOFAR, 2012, The Low Frequency Array, www.lofar.org
- Loeb & Zaldarriaga, 2004, PhRvL, 92u, 1301
- Mann et al., 1999, A&A, 348, 614
- Novaco & Brown, 1978
- O'Dea, C.P., 1998, PASP, 110, 493
- Owen, F., Eilek, J. & Kassim, N., 2000, ApJ 514, 611
- Pritchard & Loeb, 2008? Of 2010?
- Pritchard & Loeb, 2010, Nature 468, 772-773
- Rajan et al., 2011, *IEEE Aerospace Conference, Montana, US, March 5-12 2011*
- Rees, 1999, Vol. 470 of American Institute of Physics Conference Series. pp. 13–+
- R ttgering, H.J.A., et al., 1997, A&A, 326, 505
- Saks et al., 2010, *The 4S Symposium, Madeira, Portugal, 31 May - 4 June 2010*
- Schoenmakers et al., 2000, MNRAS, 315, 371
- Tselikhovich & Hirata 2010
- Van Haarlem et al., 2013, accepted for A&A, arXiv, 1305.3550
- Weiler, 1987, Proceedings of the Workshop, Green Bank, WV, Sept. 30-Oct. 2, 1986
- Weiler et al., 1988, A&A, 195, 372-379
- Wilman et al., 2008, MNRAS, 388, 1335
- Zarka, 1992, AdSpR, 12, 99
- Zarka et al., 2004, Journal of Geophysical Research (Space Physics) 109 (A18), 9
- Zarka, P., 1998, J. Geophys. Res., 103, 20159
- Zarka, P., 2004, Planet. Space Sci., 52, 1455
- Zarka, P., 2007, Planet. Space Sci., 55, 598
- Zarka, P., Cecconi, B. & Kurth, W., 2004, J. Geophys. Res., 109, A09S15
- Zarka, P., Farrell, W., Fischer, G. & Konovalenko, A., 2008, Space Sci. Rev., 137, 257

Solar System Debris Disk - S2D2



Image credit: NASA/JPL-Caltech

Proposal for Science Themes of ESA's L2 and L3 Missions

Ralf Srama

Universität Stuttgart, Institut für Raumfahrtsysteme
Raumfahrtzentrum Baden Württemberg
Pfaffenwaldring 29, 70569 Stuttgart

srama@irs.uni-stuttgart.de

Tel. +49 711 6856 2511, Fax +49 711 685 63596

<http://www.irs.uni-stuttgart.de>

Contributors

Eberhard Grün, Max-Planck-Institut für Kernphysik, Heidelberg, Germany

Alexander Krivov, Astrophysikalisches Institut, Univ. Jena, Germany

Rachel Soja, Institut für Raumfahrtsysteme, Univ. Stuttgart, Germany

Veerle Sterken, Institut für Raumfahrtsysteme, Univ. Stuttgart, Germany

Zoltan Sternovsky, LASP, Univ. of Colorado, Boulder, USA.

Supporters

listed at <http://www.dsi.uni-stuttgart.de/cosmicdust/missions/debrisdisk/index.html>

Summary

Understanding the conditions for planet formation is the primary theme of ESA's Cosmic Vision plan. Planets and left-over small solar system bodies are witnesses and samples of the processing in different regions of the protoplanetary disk. Small bodies fill the whole solar system from the surface of the sun, to the fringes of the solar system, and to the neighboring stellar system. This is covered by the second theme of Cosmic Vision. Associated with the small bodies is a Debris Disk of dust and meteoroids that are constantly generated from the disintegration of their parent bodies due to a wide range of processes.

In recent years the solar system is no longer the only planetary system available to study. An ever-growing number of extra-solar planetary systems is being discovered. Either the planets are seen directly, or the effects of the planet on their central star or on their environment are observed. Planets form in a protoplanetary disk from collapsing interstellar material which is mixed and heated and, finally, condensed and agglomerated into planetesimals that accreted into planets. Material not consumed in planets remains in a wide disk of planetesimals around the central star.

Debris disks have been identified around a significant fraction of main-sequence stars using the mid or far infrared excess in the spectral energy distribution of the stars. Debris disks are optically thin and mostly gas-free disks of 1 μm to 100 μm -sized dust grains. Such short-lived grains are continually replenished through mutual collisions in a ring of unseen km-sized and bigger planetesimals.

Most of the concepts required to explain extra-solar debris disks have been developed from observations of our own solar system debris disk. However, the detailed processes (the combination of planetary scattering and collisional shattering) and the resulting large-scale structure of this disk remain obscure. Indeed, our own solar system has inner and outer debris disks that are directly analogous to those in exoplanetary systems. Moreover, the best current models of the interplanetary dust environment are not in agreement.

Unlike extra-solar debris disks, knowledge of the solar system debris disk comes mostly

from direct observations of the parent bodies. The sizes of these parent bodies range from Near Earth Asteroids with sizes of a few 10 m, to km-sized comet nuclei, to over 1000 km-sized Trans-Neptunian Objects. The inner zodiacal dust cloud has been probed by remote sensing instruments at visible and infrared wavelengths, micro crater counts, in situ dust analyzers, meteor observations and recent sample return missions. Nevertheless, the dynamical and compositional interrelations between dust, interplanetary meteoroids, and their parent objects are still largely unknown. No similar observations exist for the outer debris cloud.

An additional scientifically important population of dust is interstellar dust passing through the solar system. These grains are the present day version of the raw material that was collected in the protoplanetary disk, heated, mixed, and reassembled in planetesimals and planets.

To develop our understanding of exoplanetary systems, we thus need to study our own system. This requires two complementary missions:

- **S2:** for the first time, we will have a 'bird's eye view' of our inner debris disk in the infrared to examine the extent and fine structure of the 'warm' zodiacal cloud, and finally we may observe the 'cold' outer Trans-Neptunian disk.
- **D2:** an in-situ observations and sample return mission to probe the orbital and compositional connection between the dust in the inner interplanetary debris disk and its source bodies, mapping the sky in dust. This mission will provide a direct comparison of the composition of interstellar raw material with the more processed material from comets and asteroids.

The dynamical and compositional interrelations between dust, interplanetary meteoroids and their parent objects are still largely unknown. The outer debris cloud of our own solar system has not been observed in the infrared so far, and exodisks harbor unobserved planets or planetesimals, while their debris disks show clear features. By studying the interaction (compositionally and

dynamically) between dust and these parent bodies, we learn about the exodisks as well as about our own solar system. To date, no compositional and density “map” exists of the debris disk of the Sun and the existing meteoroid models do not provide reliable answers for meteoroid fluxes further away from the Sun as 1 AU.

If we want to understand exoplanetary systems, we must start at “home”. S2D2 will shed light on all these questions by mapping our solar system in dust, using the unique combination of in-situ dust measurements, analyses of returned samples, and a bird’s eye view for infrared observations of our outer “home” debris disk and beyond. This will provide links between interplanetary meteoroids and their parent objects, teaching us about hidden planetesimals in exodisk debris clouds and much more. The wealth of science return of a mission like S2 and D2 is large: S2D2 therefore covers nearly all topics of the first and second cosmic vision themes, providing valuable information for astronomy, exoplanet sciences, solar system formation, planetary sciences, astrobiology and make human interplanetary spaceflight safer by better understanding the meteoroid environment.

Science goals

- **Determine the extent and fine structure of the solar system debris disk**
- **Establish the dynamical and compositional relationships between micrometeoroids and their parent bodies**
- **Characterize similarities and differences of micron to mm-sized meteoroids flux**
- **Determine compositional differences between interstellar, comet, and asteroid dust**
- **Link samples returned and analyzed to meteor streams and parent objects**

1. Introduction

1.1. Extra solar debris disks

Debris disks were first discovered by IRAS in 1984. Since then, a significant fraction of main-sequence stars have been found to harbor such disks. They are usually identified by the mid or far infrared excess in the spectral energy distribution of the stars (Wyatt 2008, Krivov 2010), which is interpreted as originating from cold dust located at tens to hundreds of AU from the central star. Debris disks consist largely of 100 μm to 1 mm-sized dust grains, and are optically thin and mostly gas-free.

Circumstellar dust is short-lived because it strongly interacts with the electromagnetic and corpuscular radiation of the central star. Therefore, the mere existence of these disks is evidence that dust-producing bigger objects are still present in mature stellar systems. Short-lived grains must be continually replenished by mutual collisions in a ring of unseen km-sized and bigger planetesimals. Hence, debris disks are believed to be the aftermath of planet formation, and to have formed within this process. Even after planet formation has long been completed, they continue to evolve collisionally and dynamically, are gravitationally sculptured by planets, and produce dust through ongoing collisional cascades. Therefore, debris disks can serve as indicators of directly invisible small bodies and planets, and are tracers of their formation and evolution. They are an important component of planetary systems.

The dust around almost all Vega-like stars is so cold that it must be orbiting with semi-major axes of 50 AU or more from the central star. Thus, the debris disks are mostly considered analogs of the Sun’s Kuiper Belt. The incidence rate of debris disks, about one-fifth for solar-type stars, is roughly comparable to the frequency of exoplanet detections with current techniques (Eiroa et al. 2013). Several tens of extrasolar systems are known to harbor both planets and debris disks. Spatially resolved images of about 70 debris disks have been obtained (Fig. 1); several of those possibly show the effects of planetary interactions. Gaps, clumps, warps, and sharp edges in the disks may indicate the existence of planets that sculpt these disks and are waiting to be discovered.

Solar System Debris Disk - S2D2

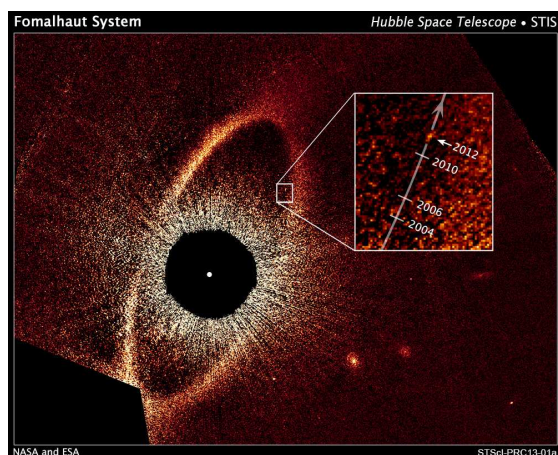


Figure 1. The Debris ring around Fomalhaut is located between about 130 and 160 AU from the star. The inset shows the location of planet Fomalhaut b imaged by Hubble Space Telescope's coronagraph. The interrelation of the planet with the debris disk is still unclear. (NASA, ESA, and P. Kalas, Univ. California, Berkeley)

A number of main sequence stars with far-IR excess also display the signature of warm micron-sized dust at distances of 1 to 10 AU from the star (Absil and Mawet, 2010). These warm inner 'exozodiacal' disks may originate from collisions in an inner planetesimal (asteroid belt analog) ring or from comet-type activity. In younger systems, they may signalize the ongoing terrestrial planet formation.

Debris disks can serve as tracers of planetesimals and planets and shed light on the planetesimal and planet formation processes that operated in these systems in the past. However, the 'visible' dust component in debris disks does not tell us directly about the dust parent bodies and dust production mechanisms. Therefore, many of the basic parameters of the debris disk remain obscure; for example, the bulk of a debris disk's mass is hidden in invisible parent bodies and cannot be directly constrained from analysis of the dust emission. Equally, the exact locations of the planetesimals remain unknown, although one expects that they orbit the star roughly where most of the dust is seen. Many properties of the planetesimals, such as their dynamical excitation, size distribution, mechanical strength and porosity, are completely unknown. For extra-solar systems we know everything from dust disks, but nothing about

their parent bodies. In contrast, in the solar system we know the distribution of the source bodies, TNOs, comets and asteroids, but the link to their dust is still obscure (e.g., Vitense et al. 2012). Therefore, a deeper insight into the solar system debris disk and its interrelation with its parent objects will provide an important step towards the understanding of the formation of our own planetary system and of planetary systems in general.

1.2. Exploration of the solar system debris disk

Our own solar system debris disk must be observed using very different techniques. The major difference is we must observe the dust from a location within the disk. We do have observations that define the disk itself – such as zodiacal light and infrared observations. However, we also have measurements of individual particles, which are the constituent elements of the disk. We must combine these elements to put together an understanding of the properties of the overall debris disk.

Observations of the zodiacal light and the meteor phenomenon provided the first insights into the interplanetary dust cloud. Zodiacal light observations provided a spatial distribution of interplanetary dust, although the size distribution of the dust remained unknown. Triangulation of meteor trails in the atmosphere, especially during meteor showers, demonstrated the genetic relation of some meteor streams to comets. However, the sizes of the particles causing these phenomena remained unknown.

The earliest motivation to study dust in space came from the suspected hazard due to meteoroid impacts onto space vehicles. In 1965 NASA launched three Pegasus satellites (Fig.2) in rapid succession by the Saturn I rocket in support of the Apollo Program. They carried large-area (180 m^2) meteoroid detectors that reliably pinned down the flux of sub-mm meteoroids in near-Earth space and thus opened the door to extended manned exploration of space. These measurements (Naumann, 1966) ended the chaos caused by early unreliable dust measurements in near-Earth space that suggested a fictitious natural dust belt around the Earth. In the mean time an ominous man-made debris belt around the Earth has become reality.



Figure 2. Pegasus 1 satellite with 180 m² micrometeoroid detector (NASA).

In the years following these experiments dust measurements were conducted throughout the solar system by Helios (launch 1974) from 0.3 AU from the sun and by Pioneer 10 (1972) out to 18 AU. While the fluxes of micrometeoroids in narrow mass ranges decreased with increasing heliocentric distance, other properties of these meteoroids, in particular dynamical properties and the relation of these particles to their sources could not be established. Later, more sophisticated dust analyzers on interplanetary and planetary missions (Galileo, Ulysses, Cassini, and New Horizon) traversed the interplanetary dust cloud from South to North, and also in radial directions. These instruments obtained fluxes and limited dynamical information for the recorded meteoroids that helped to constrain models of the interplanetary dust cloud.

Analyses of microcraters on lunar samples returned by the Apollo astronauts finally made it possible to derive the size distribution of interplanetary meteoroids at 1 AU. These allowed the meteoroid flux and spatial density to be calculated at 1AU, using additionally careful crater size to projectile calibrations, determinations of the impact speed distribution from meteor observations, and flux measurements from early in situ instruments like Pegasus the meteoroid flux and spatial density at 1 AU was established. Later measurements of craters on the LDEF satellite and other exposed surfaces generally confirmed the lunar microcrater analyses. With an absence of relevant data outside of 1AU, this meteoroid mass distribution was extrapolated to other regions of the solar system.

Another milestone in the observations of the local debris disk was the first all-sky survey at infrared wavelengths, conducted by

the Infrared Astronomical Satellite (IRAS) in 1983. It mapped the sky at 12, 25, 60 and 100 μm wavelengths, with resolutions ranging from 30 arcseconds at 12 microns wavelength to 2 arcminutes at 100 microns wavelength. Several hundred thousand sources were discovered including stars, galaxies, and solar system objects. Many stars with dust disks were identified through their infrared excess radiation; among them dust disks around Vega, Beta Pictoris, Fomalhaut, and Epsilon Eridani. A surprise was the observed structure in the zodiacal cloud: IRAS discovered a complex system of asteroid bands and many comet trails (Fig. 3). Because the Earth is located in the midst of the bright zodiacal cloud, weaker emissions from a presumed outer disk were not detected by IRAS or any follow-up infrared missions.

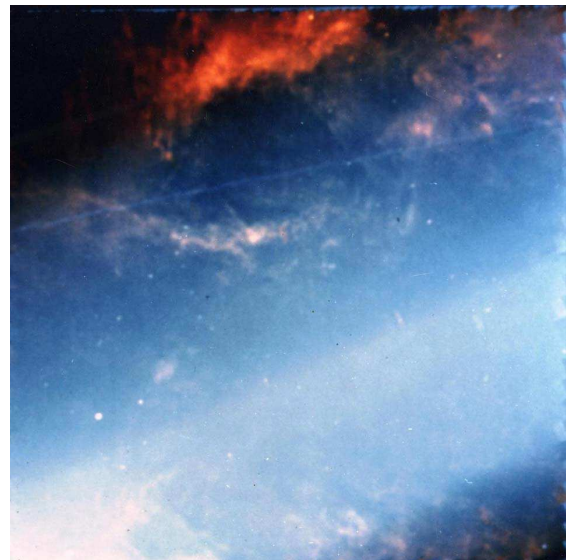


Figure 3. Composite view of a sky section by IRAS (Price, 2009, NASA). False color image with 12 μm intensity as blue and 100 μm as red. The fine streak across the upper part of the image is the dust trail of comet Temple 2 and the diagonal broad light blue band is emission from the zodiacal dust along the ecliptic plane with the asteroidal dust bands. The more complex white and red patches are emissions from interstellar cirrus and cool dust near the galactic plane.

The next break-through came from the Stardust mission that was launched in 2004. This mission returned samples from comet Wild 2 to Earth that revolutionized our understanding of early solar system processes.

Solar System Debris Disk - S2D2

Analysis of returned cometary grains demonstrated that the protoplanetary disk was a highly mixed environment that transported grains from the inner hot regions to the cold regions where comets formed. Earlier analyses of meteorites and interplanetary dust particles collected in the stratosphere always assumed that meteorites containing high-temperature mineral phases must have originated from asteroids that formed in the hot inner regions of the protoplanetary disk.

2. Parent bodies

Our planetary system contains a variety of highly interrelated objects. It extends from the F-corona (at ~ 0.02 AU) to beyond 50 AU from the sun and includes 8 planets, Trans-Neptunian objects (TNOs) and their relatives, Centaurs, and Jupiter family comets (JFCs), asteroids, meteoroids and dust. The eight known planets are arranged in two groups: four terrestrial planets and four giant planets. The main asteroid belt is located between these two groups of planets. It comprises planetesimals that failed to grow to planets because of strong perturbations from nearby Jupiter (e.g., Wetherill, 1980).

The most important dust sources in the solar system are the small solar system bodies. Fig. 4 shows the different populations of small bodies with distance to the Sun and their eccentricity. Understanding these sources allows us to build a picture of the generation of debris disks from the larger objects in planetary systems. In this section we describe the small body populations in our solar system and our present understanding of how they are related to the solar system debris disk. This relates both to an outer solar system dust disk associated with the TNO belt (analogous to the 'cold disks' around other stars), and an inner solar system dust disk interior of Jupiter (likely analogous to the 'warm disks' around other stars).

2.1. Trans-Neptunian Objects and Kuiper Belt Objects

The belt of Trans-Neptunian Objects (TNOs) outside the orbit of Neptune contains planetesimals that did not form planets because the density of the outer solar nebula was too low (e.g., Lissauer, 1987). Objects in this volume of space are normally divided

between the Kuiper belt, the scattered disk, and the distant Oort cloud. The Oort cloud, located near the boundary of the solar system, is believed to be populated by objects scattered out by planetary interactions from the protoplanetary disk.

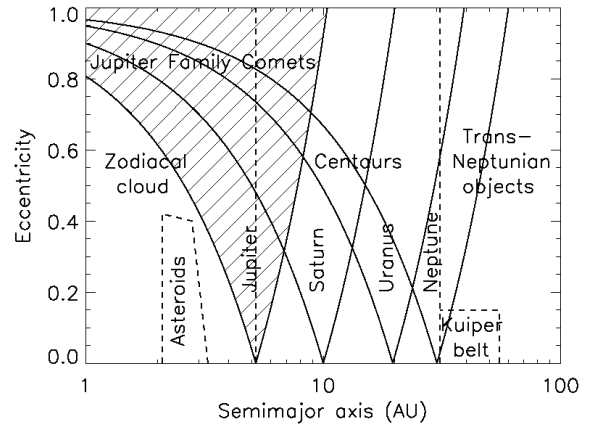


Figure 4. Small solar system bodies that are parents of the solar system debris disk. The triangular shaped regions at the semi-major axes of the giant planets delineate the scattering zones of the respective planet.

Pluto, discovered in 1930, was the first known Kuiper belt object. However, the belt itself was not identified until after the discovery of the next objects after 1992. At present, more than 1600 Kuiper belt objects are known to exist (Fornasier et al. 2013). At such large distances from the sun, these objects are very cold, hence produce black-body radiation at a wavelength of around 60 μm . The scattered disk objects are believed to be the main repository of the short periodic comets. Collisions among the TNOs (Stern, 1996) create fragments down to the size of dust grains: these generate a cold outer debris disk that has yet to be observed. Signatures of dust generated by TNOs have been found in the data from dust instruments on board the Pioneer 10 and 11, and recently the New Horizons spacecraft (Landgraf et al., 2002 and Poppe et al., 2010). While the dust concentrations detected between Jupiter and Saturn were mainly due to the cometary components, the dust outside Saturn's orbit is dominated by grains originating from the TNO belt either from collisional processes or cometary-like activity of TNOs (Altobelli et al., 2013). In order to account for the amount of dust found by Pioneer and New Horizon, a total of about $5 \times 10 \text{ g s}^{-1}$ of dust must be

released by TNOs through collisional fragmentation and erosion by impacts of interstellar dust grains onto the objects' surfaces (Han et al., 2011). This makes the Kuiper disk the brightest extended feature of the Solar System when observed from outside the system. Nevertheless, the solar system's dust disk is orders of magnitude fainter than known debris disks around other stars. However, it is not yet clear why this is the case. Due to current observational limitations, it is not yet known if any other stars possess disks as tenuous as ours (Kuchner and Stark 2010).

TNOs are exposed to intense galactic cosmic radiation. For example, when high-velocity supernova shocks pass through the interstellar medium, the protective heliosphere is compressed allowing high fluxes of cosmic rays reach the Earth (Sonett et al. 1987). This radiation affects the upper meter of the surface layers of TNOs. Estimates of the composition of these objects come from observations of their colors and spectra. They range from very blue to very red. Typical models of the surface include water ice, amorphous carbon, silicates and the reddish tholins. Tholins are substances that are produced by the irradiation of gaseous mixtures of nitrogen or clathrates of water and methane or ethane.

2.2. Centaurs

Centaurs are small icy Solar System bodies with a semi-major axis between those of the giant planets Jupiter and Neptune (Fig. 4). They have unstable orbits that cross or have crossed the orbits of one or more of the giant planets, and have dynamic lifetimes of a few million years. Through a planetary scattering cascade (Quinn et al., 1990) some TNOs become Centaurs, Trojans, and Jupiter Family Comets (JFCs) that finally disintegrate and generate the inner warm Zodiacal debris disk. Because of its unstable nature, astronomers now consider the scattered disc to be the place of origin for most periodic comets observed in the Solar System, with the Centaurs, being the intermediate stage in an object's migration from the disc to the inner Solar System (Horner et al., 2004).

Examples of Centaurs that show cometary activity are 30 to 100 km objects like 95P/Chiron (at 13.7 AU), 39P/Oterma (7.2 AU), and 29P/Schwassmann-Wachmann 1

(5.9 AU). In these cases cometary activity is driven by CO₂ and other gases from even more volatile ices. The Trojans captured in a 1:1 resonance with Jupiter are a group of objects in between Centaurs and asteroids.

2.3. Comets

Comets have a wide range of orbital periods, ranging from a few years to hundreds of thousands of years. Short-period comets originate in the scattered disc. Longer-period comets are thought to originate in the Oort cloud. At the shorter extreme, Encke's Comet has an orbit that does not reach the orbit of Jupiter, and is known as an Encke-type comet. Most short-period comets (those with orbital periods shorter than 20 years and inclinations of 20-30 degrees or less) are called Jupiter Family Comets, JFCs. Those like Halley, with orbital periods of between 20 and 200 years and inclinations extending from zero to more than 90 degrees, are called Halley-type comets. As of 2012, only 64 Halley-type comets have been observed, compared to more than 500 identified Jupiter-family comets. It is useful to classify cometary orbits by the Tisserand parameter, T_J , which describes the interaction with Jupiter: Jupiter family comets have $2 < T_J < 3$, long period and Halley-type comets have $T_J < 2$. For most comets sublimation of water ices drives cometary activity inside 3 AU. But sometimes even JFCs like Hartley 2 show significant activity driven by volatile gases like CO₂. Besides emission of dust carried by cometary gases, comets frequently undergo fragmentation. A recent spectacular event of this type is the splitting 73P/Schwassmann-Wachmann (Fig. 5).

Instruments on the Halley missions (Giotto, VeGa 1 and 2) provided our first analyses of dust grains released from comets. The major goal of these dust measurements was to study the dust composition. Cometary dust particles were found to be mostly composed of by compounds of light elements H, C, N, O and of compounds of rock-forming elements such as Mg, Si, Ca, and Fe, respectively. Isotopic information was obtained only for a few among the most abundant elements in the mass spectra of the grains: C, Mg, Si, S, and Fe (Jessberger et al., 1988). The isotopic ratios - within large uncertainties - are generally

Solar System Debris Disk - S2D2

similar to cosmic abundance. In January 2006 the Stardust sample capsule returned safely to Earth with thousands of particles from comet 81P/Wild 2 for laboratory study (e.g. Brownlee et al., 2006, Flynn et al., 2006, Hörz et al., 2006). The particles collected are chemically heterogeneous; however, the mean elemental composition of particles from comet Wild 2 is consistent with CI meteorite composition. The particles are weakly constructed mixtures of nanometer-scale grains with occasional much larger Fe-Mg silicates, Fe-Ni sulfides, Fe-Ni metal phases. The organics found in comet Wild 2 show a heterogeneous and unequilibrated distribution in abundance and composition. Hydrogen, carbon, nitrogen, and oxygen isotopic compositions are heterogeneous among particle fragments; however, extreme isotopic anomalies are rare, indicating that this comet is not a pristine aggregate of presolar materials. The abundance of high-temperature minerals such as forsterite and enstatite appears to have formed in the hot inner regions of the solar nebula.

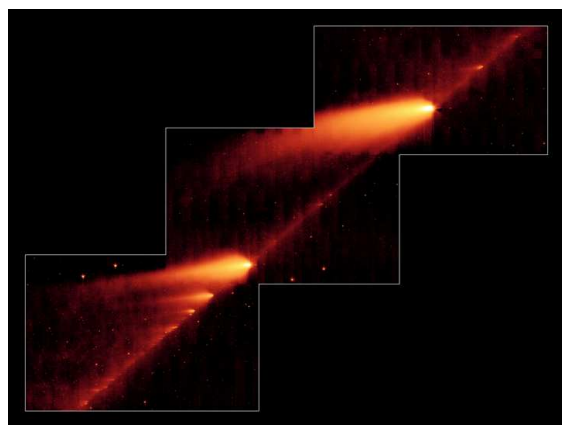


Figure 5. Breakup of 73P/Schwassmann-Wachmann and formation of a major new meteoroid stream (NASA, Spitzer). In 1995, 73P began to disintegrate; currently, the JPL data base identifies 66 individual fragments.

2.4. Asteroids

The large majority of known asteroids orbit in the asteroid belt between the orbits of Mars and Jupiter or co-orbital with Jupiter (the Jupiter Trojans). However, other orbital families exist with significant populations, including the near-Earth asteroids. Members of

asteroid families are fragments of past asteroid collisions; e.g. the Karin asteroid family consisting of at least 90 main-belt asteroids. From orbital analysis it has been concluded that this family was created 5.8 ± 0.2 million years ago (Nesvorný et al., 2002) making it the most recent known asteroid collision. This family may also be the source of one of the interplanetary dust bands discovered by the IRAS satellite (Low et al., 1984).

Individual asteroids are classified by their characteristic spectra, with the majority falling into three main groups: C-type, S-type, and M-type. These were named after and are generally identified with carbon-rich, stony, and metallic compositions, respectively. The Japanese Hayabusa spacecraft picked-up from the surface of asteroid Itokawa some dust samples and returned them to the Earth by 2010. The composition was found to match that of an LL chondrite, i.e. the minerals olivine and pyroxene (Nakamura et al., 2011).

The relation of asteroid (3200) Phaethon to the Geminid meteor shower (in mid-December) demonstrates that asteroids are not just inert rocks that release dust when impacted by other objects. Recently discovered main-belt comets form a distinct class, orbiting in more circular orbits within the asteroid belt. On some of those (e.g. 24 Themis) water ice was found on the surface. These objects produce transient, comet-like comae and tails. However, other mechanisms like rotational instability, electrostatic repulsion, radiation pressure sweeping, dehydration stresses, and thermal fracture are considered as well (Jewitt, 2012).

3. Solar System Debris Disk

3.1. Meteoroids and Dust

The zodiacal meteoroid cloud is the inner part of the solar system debris disk. It extends from the F-corona (at ~ 0.02 AU) to Jupiter's orbit and exhibits global structure such as a central offset, an inclined and warped plane of symmetry, and resonant rings that all result from planetary perturbations (Dermott et al., 2001). At mid-infrared wavelengths (10 to 25 μm) thermal emission from the zodiacal cloud outshines any other diffuse astronomical object when observed from the Earth (Fig. 6).

The zodiacal cloud is a compositionally and dynamically diverse population stemming

from a range of sources. Meteoroids are continually being replenished (via cometary sublimation and fragmentation, asteroid collisions and other production mechanisms), while evolving dynamically (due to radiation forces and planetary perturbations), and are ultimately removed (by inter-particle collisions, planetary accretion and scattering, evaporation, sputtering and ejection from the solar system).

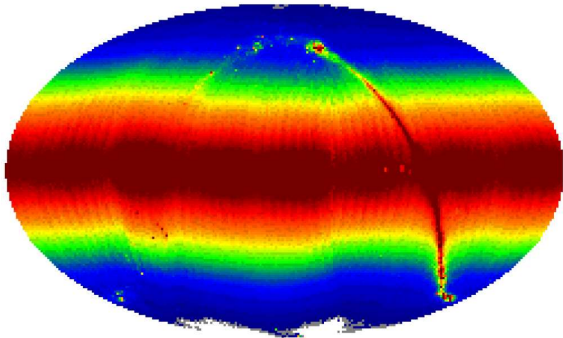


Figure 6. Sky map of zodiacal cloud at 25 microns (COBE DIRBE, Kelsall et al., 1998). Only radiation from the galactic plane is visible in the right part of the image.

Zodiacal dust particles originate in a collisional cascade from bigger interplanetary meteoroids or by direct injection from comets (Grün et al., 1985, Ishimoto 2000, Dikarev et al., 2005). The fate of freshly generated micrometeoroids strongly depends on the size of the particle. Sub-micron sized particles will mostly escape the solar system by the action of radiation pressure. Micron-sized dust is steadily transported by the Poynting-Robertson effect towards the sun, while 100 micron-sized and bigger particles remain on orbits similar to the parent object and rapidly expand into a dust trail: this forms a meteor stream if the Earth crosses its orbit. Cometary trails and meteor streams are a signs of fine structure in the zodiacal meteoroid cloud. On longer timescales (10^5 years), these big particles are randomly scattered through the solar system by the planets' gravitation. Infrared observations show that the majority of short-period comets display trails of big particles at high concentrations (Reach et al., 2007).

It is believed that at the present time comets produce most of the dust at 1 AU and asteroidal collisions contribute little to the

zodiacal cloud (Nesvorny et al., 2010). However, this proportion may have changed with time: at the time of a major collision in the asteroid belt (e.g. at the time of the Karin event ca. 6 Mio years ago) the zodiacal light may have flared up by several orders of magnitude lasting for $\sim 10^5$ years until the optically most active particles have been removed by the Poynting Roberston effect (Durda and Dermott, 1997). Alternatively, when a new major comet enters the inner solar system the zodiacal cloud may be affected for similar time scales. Actually, there are indications that we are presently at the final stages of such an event: comet Encke is currently in a stable 3-year orbit from 0.3 to 4 AU and is still producing significant amounts of dust as evidenced by the associated tail and trail. Actually, the Taurid meteor streams and comet Encke are believed to be remnants of a much larger comet, which has disintegrated over the past 20,000 to 30,000 years (Whipple, 1940; Klačka, 1994). This meteoroid stream is the longest in the inner solar system and it contains a significant amount of big particles causing fireballs in the sky. The stream is rather spread out in space; Earth takes several months to pass through it, causing an extended period of meteor activity, compared with the much smaller periods of activity in other showers.

Spacecraft measurements have identified a significant presumably very fast flow of sub-micron sized dust grains arriving from the solar direction (Berg and Grün, 1973). These beta-meteoroids were probably generated as debris resulting from mutual collisions between larger meteoroids in the inner solar system and were driven out of the solar system by solar radiation pressure (Zook, 1975). Recently, the STEREO wave instrument has recorded a very large number of intense voltage pulses (Meyer-Vernet et al., 2009) that have been interpreted as nano-particles striking the spacecraft at a velocity of the order of magnitude of the solar wind speed. These particles may indicate significant sources of dust inside the Earth orbit.

Most of what we know about the composition of interplanetary dust comes from IDPs collected in the atmosphere. About half of the collected particles have elemental abundances that closely match the bulk abundances of CI or CM carbonaceous

Solar System Debris Disk - S2D2

chondrite meteorites (Rietmeijer et al., 1998). However, the origin of these particles (cometary vs. asteroidal) is inferred by very indirect methods. In-situ analyses of interplanetary dust are sparse because of the small sensitive areas of dust mass analyzers. The Cassini and Stardust missions provided a handful of dust spectra in interplanetary space (Krueger et al., 2004; Hillier et al., 2007). Much more are needed in order to construct a compositional inventory of interplanetary dust and establish the link to its sources: comets, asteroids, Trans-Neptunian objects, and interstellar space.

The spatial and mass distribution of meteoroids everywhere in the solar system is determined by the competition between transport by (1) planetary scattering, (2) the Poynting-Robertson effect and (3) collisional balance (disruption of meteoroids and generation of fragments).

3.2. Models of the Inner Debris Disk

The best quantitative understanding of the meteoroid environment is summarized in the meteoroid environment models of the space agencies.

NASA's Meteoroid Environment Model (MEM) was developed by (Jones, 2004, McNamara et al., 2004) on the basis of sporadic meteor observations from the Canadian Meteor Orbit Radar (CMOR). This is combined with zodiacal light observations from Helios which provide the radial meteoroid density distribution. The mass distribution in the range from 10^{-6} to 10 g was derived from lunar crater statistics (Grün et al., 1985). Since the model heavily uses orbital element distributions from meteor data that must intersect the Earth, it can only be used at best for predictions of fluxes, speeds, and directions from 0.2 to 2.0 AU. This includes the environments of the planets Mercury (0.4 AU) to Mars (1.5 AU). Additionally, the model applies only to an observer moving near the ecliptic plane (McNamara et al., 2004).

ESA's Interplanetary Meteoroid Environment Model (IMEM) is a dynamical evolutionary model (Dikarev et al., 2005). Contrary to all earlier attempts, this model starts from the orbital elements of known sources of interplanetary dust: comets and asteroids. The model assumes that big

meteoroids ($\geq 10^{-5}$ g) stay orbits like their parent objects, while the orbits of smaller meteoroids evolve under planetary gravity and the Poynting-Robertson effect. Thermal radiation measurements by the COBE DIRBE instrument (Kelsall et al., 1998), in situ data from the dust instruments onboard Galileo and Ulysses (Grün et al., 1997) and lunar microcrater distributions (Grün et al., 1985) are used to calibrate the contributions from the known sources. Attempts to include meteor orbits from the Advanced Meteor Orbit Radar AMOR (Galligan and Baggeley, 2004) in the model failed since the AMOR orbital distributions were incompatible with the COBE latitudinal density profile. The derivation of the meteoroid spatial distribution from the actual radar meteor measurements is quite complex and involves numerous assumptions; whereas the derivation of the infrared brightness along a line-of-sight is relatively straight-forward. Therefore, Dikarev et al. (2005) decided not to include meteor data in the IMEM model. A recent similar analysis (Nesvorný et al., 2010) confirms that comets are currently the main contributor to interplanetary dust at 1 AU confirmed that the radar meteor systems underestimate the contributions from slow meteoroids.

Both models describe the cratering flux at 1AU quite well; however, the flux of mm-sized meteoroids differs by a factor two due to the different assumed relative speeds. At other heliocentric distances from Mercury to Mars the predicted fluxes differ by up to 2 orders of magnitude between the two models (Grün et al., 2013). The current knowledge of the interplanetary meteoroid environment as exemplified by these meteoroid models is insufficient to provide reliable assessment of the risk of meteoroid impacts for human travel in interplanetary space (Grün et al., 2013).

The discrepancy between the models results from the lack of data that constrain these models. Most importantly needed are (1) dynamical data of both μm and mm-sized meteoroid populations, (2) the spatial distribution of the different populations and its time variation.

4. *Interstellar dust in the solar system*

Interstellar dust is the major ingredient for protoplanetary disks out of which planetesimals and planets form. From analyses of material we find a surprisingly homogeneous distribution of isotopes everywhere in our solar system. Therefore, by comparing the composition of interstellar dust and its variation with those of cometary, and asteroidal dust we will learn about the mixing processes in various parts of the protoplanetary disk.

Dust grains condense in the expanding and cooling stellar winds from asymptotic giant branch (AGB), post-AGB stars, and also in supernova explosions, which results in a wide range of elemental and isotopic compositions. This so-called ‘stardust’ provides the seeds for ISD grains that grow further in cool interstellar clouds by accretion of atoms and molecules, and by agglomeration. On the other hand, interstellar shocks can efficiently destroy ISD grains by sputtering and high-speed grain-grain collisions (resulting in shattering or vaporization) behind shock fronts. In denser regions, low velocity grain-grain collisions results in coagulation. Ultimately, ISD grains can either be destroyed in newly-forming stars, or become part of a planetary system. The material in ISD grains is repeatedly recycled through the galactic evolution process (Dorschner and Henning, 1995).

The solar system currently passes through a shell of material that is located at the edge of the local bubble (Frisch et al., 1999). It emerged from the interior of this bubble within the past 10^2 - 10^5 years. Since entering the cluster of clouds that appear to have come from the Scorpius-Centaurus Association, the Sun has encountered a new interstellar cloud at least every 10^4 years. It is clear that sampling of dust from our LIC 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 ISM.

In 1992, after its Jupiter flyby, the Ulysses spacecraft unambiguously identified interstellar dust penetrating deep into the Solar System. The motion of the interstellar grains through the Solar System was approximately parallel to the flow of neutral interstellar

hydrogen and helium gas, both traveling at a speed of 26 km/s (Baguhl et al 1995). Ulysses measured the interstellar dust stream at high ecliptic latitudes between 2 and 5 AU. The masses of interstellar grains range from 10^{-18} kg to about 10^{-13} kg with a maximum of the flux at about 10^{-16} kg (Landgraf et al., 2000). The in-situ dust detectors on board Cassini, Galileo and Helios, (Altobelli et al., 2003, 2005, 2006) detected interstellar dust in heliocentric distance range between 0.3 and 3 AU in the ecliptic plane. The interstellar dust stream is strongly filtered by electromagnetic interactions with the solar wind magnetic field and by solar radiation pressure and display a variation with the 22 years solar cycle (Landgraf et al., 1999, Sterken et al., 2012)

En route to the comet Wild 2, Stardust collected and returned interstellar dust particles to Earth for analysis. The CIDA instrument provided the first high mass-resolution spectra of a few tens of presumably interstellar grains. The spectra indicate that the main constituents of interstellar grains are organic with a high oxygen and low nitrogen content. It was suggested that polymers of derivatives of the quinine type are consistent with the impact spectra recorded (Krueger et al., 2004). Three candidate interstellar grains have been identified in the Stardust collections (Westphal et al., 2013) and await their detailed analysis.

Solar System Debris Disk - S2D2

5. Top-Level Scientific Questions and Required Observations

A better understanding of the solar system debris disk requires answers to the following questions:

- **What is the extent and fine structure of the solar system debris disk?**

Infrared observations of the inner and outer debris disks.

- **What are the similarities and differences of the orbits and the fluxes of micron to mm-sized meteoroids in the mass range 10^{15} - 10^{-3} g?**

In situ measurements of fluxes of big and small dust particles and their spatial and temporal variations.

- **What are the orbital and compositional relationships between micrometeoroids and their parent bodies?**

In situ measurements of trajectories and the composition of small dust particles and correlation of the composition of collected particles with their trajectories at the time of collection.

- **What is the composition of interstellar dust and its variation?**

Identification of interstellar particles by their trajectories and measurement of their compositions both in situ and for collected particles.

- **What are the compositional differences between interstellar, cometary, and asteroidal dust and how do they relate to processes in the protoplanetary disk?**

Identification of particles from different sources by their trajectories and comparison of their

compositions both in situ and for collected particles

- **What are the objects and processes that generate nano dust in the inner solar system?**

Determination of the temporal and spatial variation of the nano-particle flux and correlation with solar wind magnetic field conditions and the appearance of sun-grazing comets and other interplanetary phenomena.

6. Strawman Mission

This strawman mission concept to analyze the solar system debris disk comprises of the combination of two missions: **S2**, an infrared survey mission to determine the structure of the inner (zodiacal) and to identify the outer (trans-Neptunian) debris disk, and **D2**, an in situ dust analysis and sample return mission to establish the orbital and compositional relationships between meteoroids and their source bodies.

S2 - A bird's eye view on the solar system debris disk

S2 is a mission to characterize the **Spatial Structure of the Zodiacal Cloud**. Several infrared survey missions have been conducted previously from positions near the Earth (Tab. 1). S2 will be an IRAS-type mission that will characterize the inner (zodiacal) and outer (trans-Neptunian) debris disks from a position outside the densest part of the zodiacal cloud. This mission will reach 40° ecliptic latitude.

Table 1. Mid-infrared survey missions: mission characteristics, telescope diameter, wavelengths (between 10 and 100 μm)

Mission	Launch date	Mass (kg)	Diam. (m)	Wave lengths (μm)
IRAS	1983	1083	0.57	12, 25, 60, 100
COBE DIRBE	1989	2270	0.19	12, 25, 60, 100
AKARI	2006	955	0.68	65, 90
WISE	2009	750	0.4	12, 22

Taking into account a 4 year LHe lifetime (Herschel), Kawakatsu and Kawaguchi (2012) describe a method to achieve a high inclination orbit through Solar Electric Propulsion (SEP) combined with multiple Delta-V Earth Gravity Assists (EGA). It includes the following steps:

1. The spacecraft is injected into the earth synchronous orbit to re-encounter the earth after one year cruise.
2. During the cruise, SEP is used to maximize the spacecraft's v_{inf} to the Earth at the next earth encounter. The thrust does not necessarily increase the inclination by itself. To enhance the efficiency to increase v_{inf} , an elliptic orbit is used for the cruise orbit.
3. By EGA, the direction of v_{inf} is changed to contribute to the inclination increase.
4. By the repetitive use of the steps 2 and 3, the inclination is increased step by step.

The initial mass of the spacecraft was assumed to be 1200 kg, the launcher is capable of injecting the spacecraft into an Earth escape orbit with v_{inf} of 7.3km/s. The orbit eccentricity was constrained to be less than 0.3; that is the perihelion is above 0.7 and the aphelion is below 1.3AU. The specific impulse of the SEP was assumed to be 3800 s, and the maximum assumed thrust was 120mN; these values are about twice the values of the SMART-1 SEP. The SEP thrust is used to decelerate the spacecraft at aphelion and to accelerate the spacecraft at perihelion, which results in a build-up of eccentricity thus increasing the v_{inf} at the next Earth encounter. After 4 years and 4 gravity assists a final orbit inclination of almost 40° is reached. At such latitude the dust density will be reduced by 97% compared to 1 AU in the ecliptic and a view of the outer cold dust disk is possible that is not blurred by the foreground zodiacal emission (cf. Landgraf et al., 2001). This mission will conduct high-resolution partial sky surveys at 12, 25, 60, and 100 micron wavelengths (in order to resolve interplanetary meteoroid streams and comet trails), thereby investigating the structure of the warm inner debris disk and identifying and characterizing the cold outer solar system debris disk. The challenges of this mission are to find optimum strategies for 1. thrust periods, 2. observation periods, and 3. data downlink periods.

D2 - mapping the sky in dust

D2 is an interplanetary mission to characterize the **Dynamical and Compositional Distributions of Meteoroids**. D2 is a mission for the in situ analysis of nano to millimeter-sized meteoroids and for the collection and Earth return of micron-sized interplanetary and interstellar dust particles. This mission will cruise near the ecliptic between the orbits of Venus and Mars and have multiple orbit changes due to flybys of Venus, Earth, and/or Mars. These orbit changes are important in order to vary the relative speed of the spaceprobe with respect to the meteoroid cloud. This way, the recorded fluxes can be used to get statistical information on the orbit distribution of big meteoroids. During its multi-year cruise the spacecraft will pass through several meteoroid streams, e.g. the 73P/SW3 stream in order to record and analyze particles from the respective comet.

D2 will carry a two-sided 200 m² meteoroid detector (Pegasus-type) for mm-sized particles, two Dust Telescopes that will establish the orbital and compositional relationships between micrometeoroids and their source bodies: comets and asteroids. Nano-Dust Analyzers will characterize collisional and other dust sources in the inner solar system. The sample return part of D2 is similar to the SARIM-type mission (Srama et al., 2008, 2012). The capsule contains a mechanism to place the dust collectors individually below trajectory sensor modules and the collectors can be stored repeatedly during unfavorable collection conditions. The capsule provides a clean and sealed environment for the collectors and avoids contamination during non-active collection phases. Seven Active Dust Collectors have a total sensitive area of 1 m². The Active Dust Collector determines the time, trajectory, primary charge, speed and mass of micro-grains entering the aperture and impacting on the collector. This allows a characterization of the origin of individual grains and eases the process to find and locate the impact tracks in the collector surfaces after return to Earth.

The overall profile consists of the launch, in situ meteoroid analyses together with interstellar/interplanetary dust collection campaigns. The mission will include fly-throughs of several meteoroid streams and return of the sample capsule. The mission

Solar System Debris Disk - S2D2

requires moderate relative impact velocities for interplanetary (~ 10 km/s) and interstellar grains (< 20 km/s) and an overall collection phase of approximately three years. The mission will include fly-throughs of several meteoroid streams. A sample return capsule will approach the Earth with an off-targeting approach, in which the final decision to reenter allows to fulfill planetary protection requirements. This off-targeting puts the transfer stage with the in-situ instrumentation in a swing-by trajectory around Earth. The in-situ measurements will continue at least one year after the fly-by exploiting the characteristics of the new orbit.

7. Instrumentation

In the following sections key instruments are listed. They should be supplemented by environment monitors.

7.1. Infrared Telescope

An IRAS type telescope (Tab. 1) with a scanning detector system is capable of the required observations. A camera system similar to the WISE telescope is also a viable alternative. Partial sky surveys at 12, 25, 60 and 100 μm are to be performed from multiple positions within and above the ecliptic plane. The spatial resolution of the IRAS instrument is appropriate for surveys that range from 30 arcseconds at 12 micron wavelength to 2 arcminutes at 100 μm .

Observation mode: At various positions above and below the ecliptic plane the IR telescope will perform slow scans perpendicular to the spacecraft-sun line, thereby imaging a swath along a great circle from 60° to 120° solar elongation.

7.2. Pegasus-type meteoroid detector

A large-area (200 m^2) meteoroid detector (Pegasus-type) is needed in order to record statistically significant numbers of big ($m > 10^{-9}$ kg) particles. The Pegasus meteoroid detector (Naumann et al., 1966) consisted of parallel plate capacitor detectors (Fig. 7) arranged in 208 panels of $0.5 \times 1\text{ m}^2$ area. The deployed wing extended over 29 m and was 4.3 m in height. The instrument weighted 1450 kg total, with more than half of it being structural. Two capacitor detectors were bonded to either side

of a 2.54 cm thick foam structural support core. The penetration depth of micrometeoroid detectors were measured by a double-sided detector consisting of 0.4 mm (total collection area of 171 m^2), 0.2 mm (16 m^2), and 0.04 mm (7.5 m^2) thick aluminum target sheets. The capacitor detector consisted of 12-micron thick Mylar sheets with a metallic (Cu) layer deposited on either side. The capacitor was maintained at a bias of 40 V and a penetrating impact caused a momentary short. The energy stored in the capacitor was discharged through this short and the Cu layers were evaporated, thus restoring the detectors back to operational condition.

A state-of-the-art version of a large-area detector system is envisioned for the D2 mission. Trajectories of big micrometeoroids would be measured with a DTS-type instruments (see below) that would covering $\sim 10\%$ of the sensitive area. The optimum pointing for such detectors is the apex ram/anti-apex direction. Electronics and detection methods now exist to operate these detectors very reliably with very low false detection rates (Auer et al., 2010).

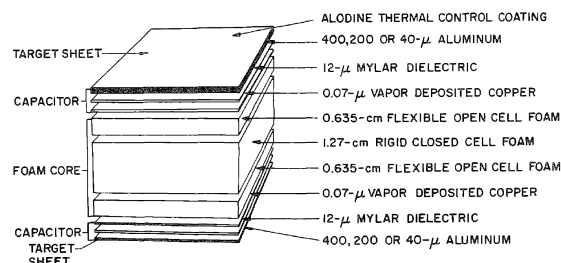


Figure 7. Pegasus meteoroid detector panel (Naumann et al., 1966).

7.3. Dust Trajectory Sensor

Micrometeoroids all acquire electric charge in space. It is possible to use this charge to accurately determine the velocity vector of these dust grains. This detection principle has been demonstrated by the Cassini Comic Dust Analyzer (CDA) (Kempf et al., 2004), where the trajectories are determined using the measurement of induced charge when the charged grains fly through a position-sensitive electrode system (Auer et al., 2008, 2010). The range of detectable particle charges is 10^{-16} to 10^{-13} C and the velocity is measurable up to

100 km/s. The newly developed Dust Trajectory Sensor (DTS) has much improved performance and sensitivity using four sensor planes with 20 wire electrodes in each (Auer et al., 2008; 2010, Xie et al., 2011). A charged dust grain flying through the instrument generates induced charges on the adjacent wires. Each electrode is connected to a separate Charge Sensitive Amplifier (CSA) and signals are digitized and recorded. The operation of a prototype DTS instrument with $40 \times 40 \text{ cm}^2$ cross section has been demonstrated at the Heidelberg and Boulder dust accelerator facilities. The measurements are highly sensitive and accurate as a minimum 8 coincident signals from the closest wire electrodes are used for the analysis. The size of the DTS instrument is scalable to approximately 1 m^2 .

7.4. Active Dust Collector

Novel active dust collectors are a combination of a Dust Trajectory Sensor, DTS, with a dust collector. In this way, not only are the trajectories of collected grains determined (and hence their direction of origin is established), but in addition their impact positions into the collecting material can be determined to sub-millimeter precision.

Intact capture of hypervelocity projectiles in silica aerogel was successfully demonstrated (Tsou et al., 1990). For aerogels having densities about 10 kg/m^3 or less, intact projectiles were lodged at the end of track. This concept has been extremely successful during Stardust's fly-through of the coma of comet Wild2 at a speed of 6.1 km/s: many cometary particles were captured intact in aerogel that had a density of 10 to 50 kg/m^3 (Brownlee et al., 2006).

In an active dust collector all particles impacting the collector must pass through the dust trajectory sensor. To function properly the DTS requires a trigger signal that terminates the cyclic signal processing and starts the data read-out. In the case of a metal collector this will be used as the target of a simple impact ionization detector with a grid in front of it. It has been shown that even impacts into aerogel create impact charges that can provide the necessary trigger signals for DTS (Auer, 1998, 2010, Grün et al., 2012). The sensitivity of the DTS electronics is of the order of 10^{-16} C and thus the trajectory of cosmic dust particles as

small as $0.4 \text{ }\mu\text{m}$ size can be measured and collected.

7.5. Dust Telescope

A Dust Telescope is composed of two parts, a Dust Trajectory Sensor and a Mass Spectrometer subsystem. This sensor is an in-situ instrument designed to detect and analyze individual impacts of sub-micron and micron sized dust grains. First, the dust particles pass the DTS then the grains impact on the plane target of the mass analyzer at the bottom of the instrument (Fig. 9) generating electrons and ions that are analyzed in a time-of-flight spectrometer. Mass spectra are measured at high resolution ($M/\Delta M > 100$). Several versions of a Dust Telescope with sensitive area of 0.1 m^2 have been tested in the lab (Srama et al., 2005, Sternovsky et al., 2007, 2011).

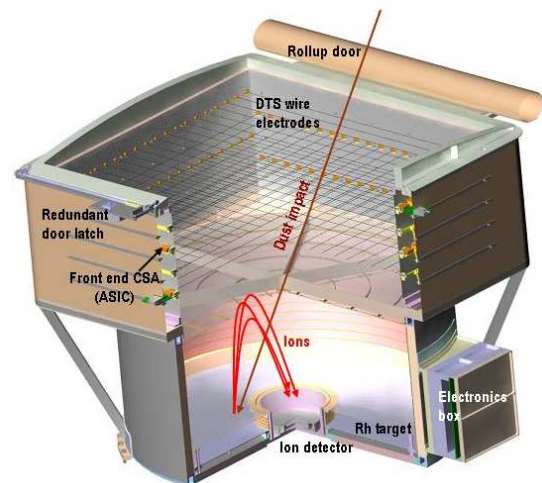


Figure 9. Dust Telescope with the four planes of wires of the Trajectory Sensor (top) and the time-of-flight mass-spectrometer (bottom). A particle impacting on the target generates electrons, ions, and charged cluster-molecules which are focused to the ion detector in the centre. Data acquisition is triggered by the electron signal collected at the target and by the ion signal.

7.6. Nano-Dust Analyzer

Beta-meteoroids and interplanetary nano-dust particles originating from the inner solar system have been detected, but only using simple impact detectors or through collisions with spacecraft. A special type of a dust mass spectrometer is needed to obtain compositional

Solar System Debris Disk - S2D2

information of this important loss mechanism for zodiacal dust. This instrument has to be able to resist extended periods of Sun pointing with high thermal heat input and with interferences from solar UV and solar wind exposure. Methods used in space coronagraphs and solar wind instruments are employed to solve these problems.

8. References

- Absil, O., Mawet, D.: *Astronomy and Astrophysics Review*, 18, 317-382, 2010
- Altobelli, N., Kempf, S., et al., *Journal of Geophysical Research* 108 (2003) DOI 10.1029/2003JA009874.
- Altobelli, N., Kempf, S., et al., *Journal of Geophysical Research* 110 (2005) DOI 10.1029/2004JA010772.
- Altobelli, N., Grün, E., Landgraf, M., *Astronomy and Astrophysics* 448(1) (2006) 243-252.
- Auer, S., E. Grün, S., et al., *Rev Sci Instrum*, vol. 79, no. 8, pp. , 2008.
- Auer, S., G. Lawrence, et al., *Nucl Instrum Meth a*, vol. 622, no. 1, pp. 74-82, 2010.
- Baguhl, M. Grün, E., et al., *Space Science Reviews* 72 (1995b) 471-476.
- Berg, O.E., Grün, E., *Space Research XIII*, 2 (1973) 1047-1055.
- Briggs, R. E., *The Astronomical Journal*, 67, 268-, 1962
- Brownlee, D.E., Tsou, P., et al., *Science* 314(5806) (2006) 1711- 1716.
- Dermott, S. F., Grogan, K., et al., in "Interplanetary Dust", Eds. Grün, E., et al. *Astronomy and Astrophysics Library*. 2001, 804 p., ISBN: 3-540-42067-3. Berlin: Springer, 2001, p. 569, 569-, 2001
- Dikarev, V., Grün, E., et al., *Advances in Space Research* 35:1282-1289, 2005
- Dikarev, V., Grün, E., in "Order and Chaos in Stellar and Planetary Systems", ASP Conference Proceedings, vol. 316. *Astronomical Society of the Pacific*, San Francisco, 2004, (Eds. Byrd, G.G. et al.).
- Divine, N.: *Journal of Geophysical Research*, Volume 98, Issue E9, 17029-17048, 1993.
- Dorschner, J., Henning, T.: *Astron. Astrophs. Rev.* 6, 271-333. 1995
- Drolshagen, G.: *Comparison of Meteoroid Models*, Inter-Agency Space Debris Coordination Committee Action Item 24.1, 2009
- Durda, D. D., Dermott, S. F.: *Icarus*, 130, 140-164, 1997
- Eiroa, C., Marshall, J. P., et al.: *ArXiv e-prints*, arXiv:1305.0155-, 2013
- Flynn, G.J., Bleuet, P., et al.: *Science* 314(5806) (2006) 1731-1735.
- Fornasier, S., Lellouch, E., et al.: *ArXiv e-prints*, arXiv:1305.0449-, 2013
- Frisch, P. C., et al.: *Astrophysical Journal*, 525, 492-516, 1999.
- Galligan, D.P., Baggaley, W.J.: *Mon. Not. R. Astron. Soc.* 353 (September), 422-446, 2004.
- Grün, E., Staubach, P., et al.: *Icarus*, 129, 270-288, 1997
- Grün, E., Zook, H. A., et al.: *Icarus*, Volume 62, Issue 2, 244-272, 1985.
- Grün, E., Sternovsky, Z., et al.: *Planetary and Space Science*, 60, 261-273, 2012
- Grün, E, et al., *Proc. Sixth European Conference on Space Debris*. ESA/ESOC. Darmstadt, Germany. 22-25 April 2013
- Han, D., Poppe, A. R., et al.: *Geophysical Research Letters*, 38, 24102-, 2011
- Hillier, J.K., Green, S.F., et al.: *Icarus* 190 (2007) 643-654.
- Hoerz, F., Bastien, R., et al.: *Science* 314(5806) (2006) 1716-1719.
- Horner, J., Evans, N.W., Bailey, M. E.: *Monthly Notices of the Royal Astronomical Society* 354 (3): 798 (2004). arXiv:astro-ph/0407400.
- Ishimoto, H.: *Astron. Astrophys.*, 362, 1158-1173, 2000.
- Jessberger, E.K., Christoforidis, A., Kissel, J.: *Nature* 332 (1988) 691-695.
- Jewitt, D.: *The Astronomical Journal*, 143, 66-, 2012
- Jones, J.: *Meteoroid Engineering Model - Final Report*, Space Environments and Effects Program SEE/CR-2004-400, NASA Marshall Space Flight Center, 2004
- Jones, J. and Brown, P.: *Monthly Notices of the Royal Astronomical Society*, Volume 265, 524-532, 1993.
- Kawakatsu and Kawaguchi: *Journal of Aerospace Engineering, Sciences and Applications*, Jul. - Sep. (2012), Vol. IV, No 3
- Kelsall, T., Weiland, J.L., et al.: *Astrophys. J.* 508, 44-73, 1998.
- Kempf, S.; Srama, R.; et al.: *Icarus*, 171, Issue 2, 2004, 317-335
- Klacka, J., Pittich, E. M.: *Planetary and Space Science*, 42, 109-112, 1994
- Krivov, A. V.: *Research in Astronomy and Astrophysics*, 10, 383-414, 2010
- Krueger, F.R., Werther, W., et al.: *Rapid Commun. Mass Spectrom.* 18 (2004) 103-111.
- Landgraf, M., Augustsson, K., et al.: *Science* 286 (1999) 2319-2322.

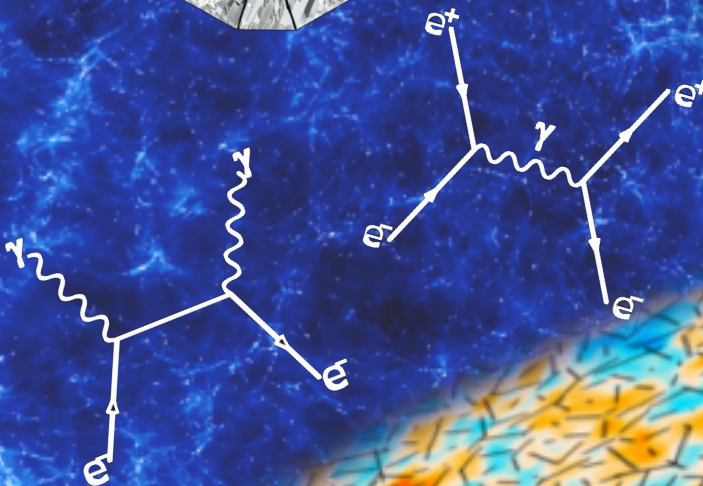
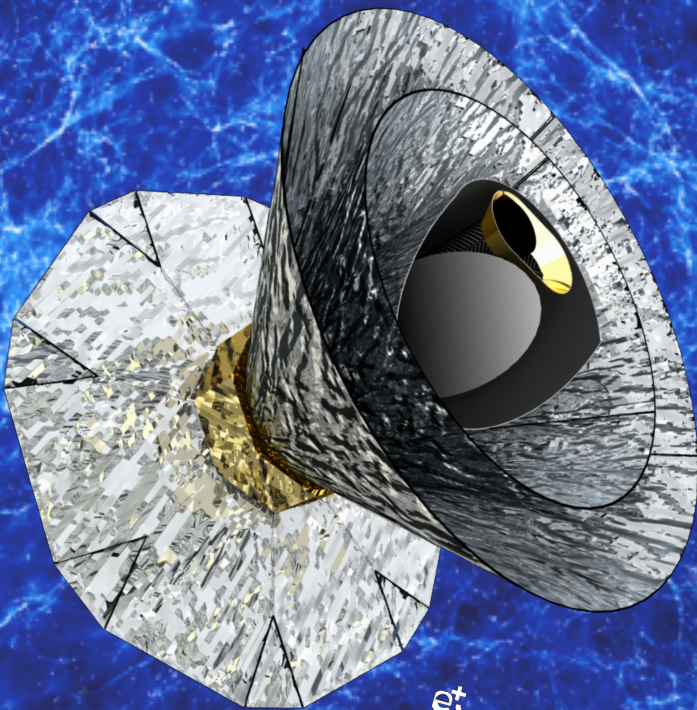
Solar System Debris Disk - S2D2

- Landgraf, M., Baggaley, W.J., et al.: J. Geophys. Res. 105 (2000) 10343-10352.
- Landgraf, M., R. Jehn, et al.: ESA Bulletin, 105, 60-63, 2001.
- Landgraf, M., Liou, J.-C., et al.: The Astronomical Journal, 123, 2857-2861, 2002
- Leinert, C., Ruesser, S. and Buitrago, J.: Astron. Astrophys. 118 (1983), 345-357.
- Lissauer, J. J.: Icarus, 69, 249-265, 1987
- McNamara, H., Jones, J., et al.: Earth, Moon, and Planets, Volume 95, Issue 1-4, 123-139, 2004.
- Meyer-Vernet, N., Maksimovic, M., et al.: Solar Physics, Volume 256, Issue 1-2, pp. 463-474 2009
- Nakamura, T., Noguchi, T., et al.: Science, 333, 1113-, 2011
- Naumann, R. J.: The near Earth Meteoroid Environment. NASA TND 3717 (1966).
- Nesvorny, D., Jenniskens, P., et al.: The Astrophysical Journal, 713, 816-836, 2010
- Nesvorny, D., et al.: Nature 417(6890):720-722 (2002).
- Poppe, A., James, D., et al.: Geophysical Research Letters, 37, 11101-, 2010
- Quinn, T., Tremaine, S., Duncan, M.: The Astrophysical Journal, 355, 667-679, 1990
- Reach, W. T., Kelley, M. S., Sykes, M. V.: Icarus, 191, 298-322, 2007
- Rietmeijer, F.J.M.: Interplanetary dust particles, in: Planetary Materials, J.J. Papike ed., Review in Mineralogy Vol. 36, Mineralogical Soc. of America, Washington, p. 2-01 - 2-95, 1998.
- Sonett, C. P., Morfill, G. E., Jokipii, J. R.: Nature, 330, 458-460, 1987
- Srama, R., Srowig, A. et al.: Earth Moon and Planets, 2005, vol. 95, pp. 211-220.
- Srama, R., et al.: Exp. Astron. (2008). doi:10.1007/s10686-008-9088-7
- Srama, R., et al.: Exp. Astron. (2012). DOI 10.1007/s10686-011-9285-7
- Staubach, P., Grün E., Jehn R.: Advances in Space Research 19:301-308 (1997)
- Sterken, V. J., Altobelli, N., et al.: Astronomy and Astrophysics, 552, A130-, 2013
- Sterken, V. J., Altobelli, N., et al.: Astronomy and Astrophysics, 538, A102-, 2012
- Stern, S. A.: The Astronomical Journal, 112, 1203-, 1996
- Sternovsky, Z., Amyx, K., et al.: Rev Sci Instrum, vol. 78, no. 1, pp. , 2007.
- Sternovsky, Z., Grün, E., et al.: Novel instrument for Dust Astronomy: Dust Telescope, presented at the Aerospace Conference, 2011 IEEE, 2011, pp. 1-8.
- Tsou, P.: Int. J. Impact Eng., 10, 615-627, 1990.
- Vitense, C., Krivov, A. V., et al.: Astronomy and Astrophysics, 540, A30-, 2012
- Westphal, A.J., et al.: Science. in preparation. 2013
- Wetherill, G. W.: Annual Review of Astronomy and Astrophysics, 18, 77-113, 1980
- Whipple, F.L.: Photographic meteor studies. III. The Taurid shower. Proc. Amer. Phil. Soc., 83, 711-745 (1940).
- Wiegert P., Vaubaillon J., Campbell-Brown, M.: Icarus 201 295-310 (2009)
- Wyatt, M. C.: Annual Review of Astronomy and Astrophysics, 46, 339-383, 2008
- Xie, J., Z. Sternovsky, et al.: Rev Sci Instrum, vol. 82, no. 10, p. 105104, 2011.
- Zook, H.A.: Planetary and Space Science 23 (1975) 1391-1397.

Polarized Radiation Imaging and Spectroscopy Mission

PRISM

**Probing cosmic structures and radiation
with the ultimate polarimetric spectro-imaging
of the microwave and far-infrared sky**



Spokesperson: *Paolo de Bernardis*
e-mail: paolo.debernardis@roma1.infn.it — tel: + 39 064 991 4271

Authors and contributors

Philippe André, Carlo Baccigalupi, Domingos Barbosa, James G. Bartlett, Nicola Bartolo, Elia Battistelli, Richard Battye, George J. Bendo, Jean-Philippe Bernard, Marco Bersanelli, Matthieu Béthermin, Pawel Bielewicz, Anna Bonaldi, François Bouchet, François Boulanger, Jan Brand, Martin Bucher, Carlo Burigana, Zhen-Yi Cai, Viviana Casasola, Guillaume Castex, Anthony Challinor, Jens Chluba, Sergio Colafrancesco, Francesco Cuttaia, Giuseppe D'Alessandro, Richard Davis, Miguel A. de Avillez, Paolo De Bernardis, Marco De Petris, Adriano de Rosa, Gianfranco de Zotti, Jacques Delabrouille, Clive Dickinson, Jose Maria Diego, Edith Falgarone, Pedro Ferreira, Katya Ferrirère, Fabio Finelli, Andrew Fletcher, Gary Fuller, Silvia Galli, Ken Ganga, Adnan Ghribi, Joaquin Gonzalez-Nuevo, Keith Grainge, Alessandro Gruppuso, Alex Hall, Carlos Hernandez-Monteagudo, Mark Jackson, Andrew H. Jaffe, Rishi Khatri, Luca Lamagna, Massimiliano Lattanzi, Paddy Leahy, Michele Liguori, Elisabetta Liuzzo, Marcos Lopez-Caniego, Juan Macias-Perez, Bruno Maffei, Davide Maino, Silvia Masi, Marcella Massardi, Sabino Matarrese, Alessandro Melchiorri, Jean-Baptiste Melin, Aniello Mennella, Arturo Mignano, Marc-Antoine Miville-Deschênes, Federico Nati, Paolo Natoli, Mattia Negrello, Fabio Noviello, Francesco Paci, Rosita Paladino, Daniela Paoletti, Francesca Perrotta, Francesco Piacentini, Michel Piat, Lucio Piccirillo, Giampaolo Pisano, Gianluca Polenta, Sara Ricciardi, Matthieu Roman, Jose-Alberto Rubino-Martin, Maria Salatino, Alessandro Schillaci, Paul Sheldard, Joseph Silk, Radek Stompor, Rashid Sunyaev, Andrea Tartari, Luca Terenzi, Luigi Toffolatti, Maurizio Tomasi, Tiziana Trombetti, Marco Tucci, Bartjan Van Tent, Licia Verde, Ben Wandelt, Stafford Withington.

Coordination group

The preparation of this science case submitted to ESA has been coordinated by:

James G. Bartlett, François Bouchet, François Boulanger, Martin Bucher, Anthony Challinor, Jens Chluba, Paolo de Bernardis (spokesperson), Gianfranco de Zotti, Jacques Delabrouille (coordinator), Pedro Ferreira, Bruno Maffei.

Supporters

Through the highly successful ESA Planck and Herschel missions, Europe has acquired considerable scientific and technical expertise in the fundamental and strategic field of microwave and far-infrared observations. A new generation of young European astronomers has been trained in this area. The science themes outlined in this proposal will allow ESA to capitalize on these strengths.

To demonstrate the breadth of support for PRISM, we are in the process of assembling a list of supporters that can be found at the following website:

<http://www.prism-mission.org>

1 Executive summary

PRISM is a large-class mission that will carry out the ultimate survey of the microwave to far-infrared sky in both intensity and polarization, as well as measure its absolute spectrum. PRISM will consist of two instruments: (1) a high angular resolution polarimetric imager with a 3.5 m telescope cooled to around 4 K to reduce thermal noise, particularly in the far-infrared bands; and (2) a low angular resolution spectrometer to compare the sky frequency spectrum to a near perfect reference blackbody. The joint exploitation of the data from these co-observing high-performance instruments will enable PRISM to make breakthrough contributions by answering key questions in many diverse areas of astrophysics and fundamental science. A few highlights of the new science with PRISM include:

(A) The ultimate galaxy cluster survey: The Sunyaev-Zeldovich (SZ) effect is the method of choice for assembling a catalog of clusters at high redshift, of particular interest for cosmology because of the tight correlation between integrated y -distortion and cluster mass. When PRISM flies, all-sky cluster samples (e.g. from eROSITA, *Euclid*) will likely count some 10^5 objects, mostly at $z < 1$. PRISM will find 10 times more clusters extending to deeper redshifts, with many thousands beyond $z = 2$. In fact, PRISM will detect all clusters in the universe of mass larger than $10^{14} M_{\odot}$, and a large fraction of those with mass above $5 \times 10^{13} M_{\odot}$. Owing to its exquisite spectral coverage, angular resolution and sensitivity, PRISM will measure the peculiar velocity of hundreds of thousands of clusters via the kinetic SZ effect, initiating a new research area: the complete mapping of the large-scale velocity field throughout the Hubble volume. In addition, PRISM will also be able to probe the relativistic corrections to the classic SZ spectral distortion spectrum, thus measuring the gas temperature. This cluster sample will allow us to probe dark energy and better understand structure formation at large redshift.

(B) Understanding the Cosmic Infrared Background: Most star formation in the universe took place at high redshift. Hidden from optical observations by shrouds of dust in distant galaxies, it is visible only in the far infrared or in X-rays. Emission from these dusty galaxies constitutes the cosmic infrared background (CIB) which PRISM, owing to its high sensitivity and angular resolution in the far infrared, is uniquely situated to investigate. The survey will sharpen and extend to higher redshifts the determination of the bolometric luminosity function and of clustering properties of star-forming galaxies. Tens of thousands of easily recognizable, bright, strongly lensed galaxies and hundreds of the very rare maximum starburst galaxies, up to $z > 6$, will be detected, providing unique information on the history of star formation, the physics of the interstellar medium in a variety of conditions up to the most extreme, and the growth of large scale structure, including proto-clusters of star-forming galaxies. The survey will also probe the evolution of radio sources at (sub-)mm wavelengths and provide measurements of the spectral energy distribution (SED) of many thousands of radio sources over a poorly explored, but crucial, frequency range.

(C) Detecting inflationary gravity waves: Present precision measurements of cosmic microwave background (CMB) temperature anisotropies lend considerable support to simple models of inflation; however, the most spectacular prediction of inflation—the generation of gravitational waves with wavelengths as large as our present horizon—remains unconfirmed. Several initiatives from the ground and from stratospheric balloons are currently underway to attempt to detect these gravitational waves through the B-mode spectrum of the CMB polarization. However, they suffer from severe handicaps such as limited frequency coverage due to atmospheric opacity, unstable seeing conditions, and far sidelobes from the ground. It is only from space that one may hope to detect the very low- ℓ B-modes due to the re-ionization bump. Because of its broad frequency coverage and extreme stability, PRISM will be able to detect B-modes at 5σ for $r = 5 \times 10^{-4}$, even under pessimistic assumptions concerning the complexity of the astrophysical foreground emissions that must be reliably removed. Moreover, PRISM will be able to separate and filter-out the majority of the lensing signal due to gravitational deflections.

(D) Probe new physics through CMB spectral distortions: The excellent agreement between the microwave sky emission and a perfect blackbody observed by the COBE FIRAS instrument is rightfully highlighted as a crucial confirmation of Big Bang cosmology. However, theory predicts that at higher sensitivity this agreement breaks down. Some of the predicted deviations are nearly sure bets, and others provide powerful probes of possible new physics. The PRISM absolute spectrometer will measure the spectrum more than three orders of magnitude better than FIRAS. y -distortions from the re-ionized gas as well as from hot clusters constitute a certain detection. However, μ -distortions and more general spectral distortions have the potential to uncover decaying dark matter and to probe the primordial power spectrum on very small scales that cannot be measured by other means, being contaminated by the nonlinearity of gravitational clustering

at late times.

(E) Probe Galactic astrophysics: PRISM will have a major impact on Galactic astrophysics by providing a unique set of all-sky maps. It will extend Herschel dust observations to the whole sky and will contribute maps in emission lines that are key to quantify physical processes. The survey will have the sensitivity and angular resolution required to map dust polarization down to sub-arcminute resolution even at the Galactic poles. No project will provide a comparable perspective on interstellar components over such a wide range of scales. The PRISM data will hold unique clues to study the interstellar medium, the Galactic magnetic field and star formation, and will address three fundamental questions of Galactic astrophysics: What are the processes that structure the interstellar medium? What role does the magnetic field play in star formation? What are the processes that rule the composition and evolution of interstellar dust?

These are but a few of the highlights of the rich and diverse physics and astrophysics that PRISM will be able to carry out.

2 Legacy archive

The hundreds of intensity and polarization maps of PRISM will constitute a legacy archive useful for almost all branches of astronomy for decades to come. Combining low resolution spectrometer data and high resolution images from the imager, PRISM will effectively deliver a full spectro-polarimetric survey of the complete sky from $50\,\mu\text{m}$ to $1\,\text{cm}$. The spectral resolution will range from about $0.5\,\text{GHz}$ to $15\,\text{GHz}$ at 1.4° angular resolution, and from $\delta\nu/\nu \approx 0.025$ to 0.25 at the diffraction limit of a $3.5\,\text{m}$ telescope (from $\sim 6''$ to $17''$).

We will build and make public full-sky maps of the absolute temperature of the CMB and of its polarization (at a resolution of about 2 arc-minutes with a sensitivity of order μK or better per resolution element), of the emission of all galactic components in absolute intensity and polarization (including main spectral lines), and several catalogues of various galactic and extragalactic objects, among which a catalogue of about a million of galaxy clusters and large groups up to redshift $z = 3$ or more.

3 Probing the Universe with galaxy clusters

The PRISM mission will exploit the advantages of cluster surveying via the SZ effect in a spectacular way, surpassing in depth any planned cluster survey and, in addition, achieving an objective unattainable in any other way: measurement of the cosmic velocity field throughout the observable universe. In short, we will detect cluster and groups systems throughout the Hubble volume from the moment they first emerge. PRISM will also provide cluster mass determinations out to high redshift through gravitational lensing of the CMB in both temperature [97] and polarization [66], something that is only possible with its high angular resolution and frequency coverage extending into bands unreachable from the ground. The *Planck*, ACT and SPT experiments have transformed SZ from a promising to a useful, mature science; PRISM will turn SZ cluster studies into perhaps our most powerful probe of cosmic large-scale structure and its evolution.

The Cluster Catalog and its Applications: We estimate the content of the PRISM SZ catalog by applying a multi-frequency matched filter [70] to simulations of a typical field at intermediate Galactic latitude. Our detection mass remains below 10^{14} solar masses at *all redshifts* (Fig. 1). Extrapolating from the observed *Planck* counts, this predicts nearly 10^6 clusters with many thousands at $z > 2$. We already know from *Planck* SZ observations [80, 81] that the SZ signal in clusters scales as our adopted relation down to much smaller masses in the local universe, leaving as our main uncertainty poor knowledge of its redshift dependence; this is, of course, one of the very motivations for studying the high redshift cluster population.

Based on this calculation, PRISM will surpass all current and planned cluster surveys, including eROSITA and Euclid – not just in total numbers, but most importantly in numbers of objects at $z > 1.5$. Cluster

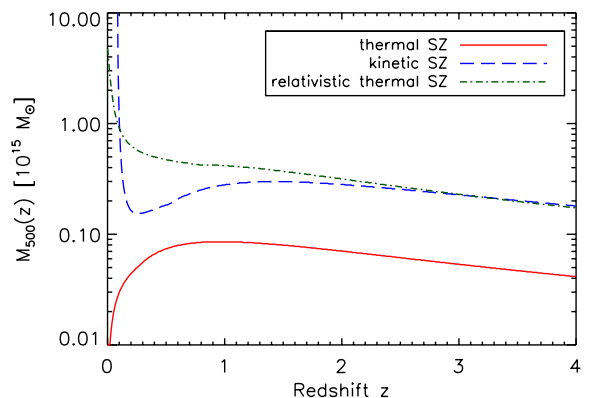


Figure 1: Lower mass limits for detection of the indicated SZ effects at signal-to-noise $S/N > 5$ as a function of redshift.

identification will be vastly more robust for PRISM than Euclid, which will suffer from the much higher contamination rate of optical/NIR cluster searches, especially at redshifts beyond unity. In all cases, only PRISM has the ability to find significant numbers of clusters in the range $2 < z < 3$, the critical epoch that present-day observations identify as the emergence of the characteristic cluster galaxy population on the red sequence. It will also enable us to explore the abundance of the intra-cluster medium (ICM), through the Y – M relation, and its relation to the galaxy population at these high redshifts.

At the time of operation, large imaging (e.g., DES, LSST, HSC) and spectroscopic surveys (e.g., 4MOST, PFS, WEAVE, BigBOSS/MS-DESI, SKA) will have covered the entire extragalactic sky. We will easily be able to obtain redshifts, spectroscopic or photometric, for all objects to $z = 2$, and the two micron cutoff of Euclid’s IR photometric survey (H band) is sufficient to detect the 4000 angstrom break in brighter cluster galaxies at higher redshifts.

The Catalog as a Cosmological Probe: As an example test of the cosmology constraints that can be obtained from the expected cluster catalog, we performed a standard Fisher analysis on the to constrain four parameters, Ω_m , σ_8 and the dark energy equation-of-state parameters w_0 and w_1 , in a standard flat Λ CDM cosmological model. The constraints on the latter dark energy parameters are $w_0 = -1 \pm 0.003$ and $w_1 = 0 \pm 0.1$ after marginalization over the first two parameters. While this remains a simplified evaluation, it nevertheless illustrates the power of the expected cluster catalog as a cosmological probe.

The Cosmic Velocity Field: PRISM will initiate an untapped research area: study of the velocity field through the kinetic SZ effect [103, 91, 8], an independent probe of dark matter and large-scale structure evolution. In Fig. 1 we show mass limit to which we expect to measure a velocity of 300 km s^{-1} to five sigma on individual clusters. This mass limit means that we will obtain velocity measurements for hundreds of thousands of clusters out to the highest redshifts. In addition, by comparing measured velocities to mass concentrations, say from Euclid lensing or galaxy surveys, we can test the theory of gravity on cosmic scales and to high redshift. This science is unattainable by any other means.

Relativistic and non-thermal effects: We will determine the temperature of clusters down to a mass limit just above 10^{14} solar masses through measurement of relativistic corrections to the thermal SZ spectrum [90, 15, 55, 96, 8]. These same characteristics allow us to search for non-thermal signatures in the spectra that could signal the presence of highly energetic particles, perhaps dark matter annihilation products, and even study the temperature structure of the most massive systems.

Diffuse SZ and the cosmic web: The diffuse, unresolved SZ effect probes a different mass and redshift range than observations of individually detected objects. We will study it through the power spectrum and higher order moments of an SZ map of the sky. *Planck* recently extracted the first Compton parameter (y -fluctuations) map [86], but the results are limited by foregrounds and noise. With many more spectral bands and much better sensitivity and resolution, PRISM will significantly improve the results, making possible attempts to directly map the cosmic web, i.e., its filaments, over large scales through its diffuse gas content.

We will explore the gas content of dark matter halos down to very low masses, a research area pioneered by *Planck* by stacking SZ measurements on known objects to detect the signal down below 10^{13} solar masses [80, 81]. The measurement over such a vast range is unique to the SZ effect and a highly valuable constraint on the mysteries of feedback mechanisms at the heart of galaxy formation. PRISM greatly expands this important science area by pushing to the lowest possible masses and by probing gas content as a function of object properties. Coupled with our lensing measurements, we have a new and exceptional tool to study the relation between light and dark matter.

Polarized SZ effect: PRISM will enable enable searches for the polarized SZ effects, giving access to transverse cluster velocities and measurements of the CMB quadrupole at distant locations.

4 Extragalactic sources and the cosmic infrared background

Early evolution of galaxies: Although *Herschel* and *Spitzer* pushed spectacular advance in our understanding of early, dust enshrouded phases of galaxy evolution, our knowledge of star-formation history in the distant universe is still very incomplete. The PRISM mission will make essential progress thanks to its unique properties: coverage of the full sky and unparalleled frequency range. As illustrated by Fig. 2, its unprecedented frequency coverage provides direct measurements of the bolometric luminosities of star-forming galaxies up to high redshifts. At $z \gtrsim 2$, i.e., in the redshift range where both the cosmic star formation and the accretion rate onto supermassive black-holes are maximum, both the IR peak associated with the

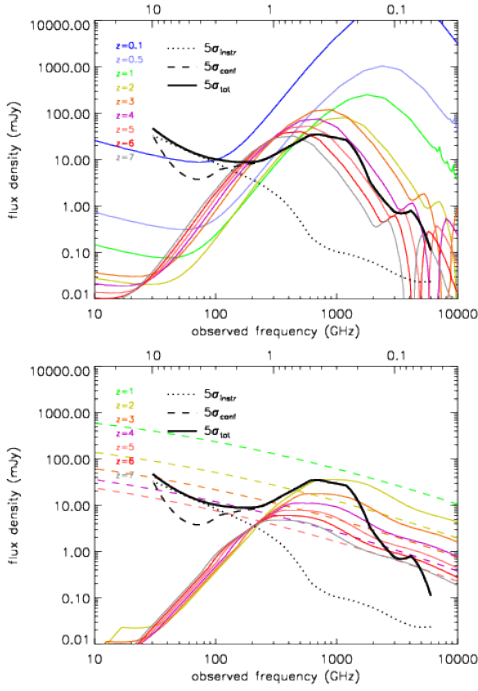


Figure 2: SEDs of dusty galaxies (top panel) and of AGNs (bottom panel) at different redshifts compared with estimated 5σ detection limits (solid black line) taking into account instrumental and confusion noise summed in quadrature. The instrumental noise refers to the full mission. The 5σ detection limits allowing for either component are shown by the dotted and the dashed black lines, showing that PRISM is confusion limited above $\simeq 150$ GHz. We have assumed that component separation techniques, extensively validated both on simulations and on real data, can efficiently remove diffuse emissions such as the CMB (that would otherwise dominate the fluctuation field for $\nu \lesssim 220$ GHz) and Galactic emissions. In the top panel, at $z = 0.1$ and 0.5 we have plotted the Arp 220 SED scaled to an IR (8–1000 μm) luminosity of $10^{12} L_{\odot}$. At $z \geq 1$ we have used the SED of the $z \simeq 2.3$ galaxy SMM J2135-0102 scaled to $L_{\text{IR}} = 10^{13} L_{\odot}$ for $z = 1$ and $z = 2$, and to $L_{\text{IR}} = 3 \cdot 10^{13} L_{\odot}$ [the luminosity of the $z = 6.34$ galaxy detected by *Herschel*/SPIRE, 92] for $z \geq 3$. In the bottom panel, the solid coloured lines represent SEDs of a type-2 QSO (contribution of the host-galaxy subtracted) with $L_{\text{IR}} = 10^{13} L_{\odot}$ at several redshifts ≥ 2 , while the dashed coloured lines show a schematic representation of the SED of the prototype blazar 3C 273 shifted to redshifts from 1 to 5.

dusty torus around AGN ($\lambda_{\text{p,AGN}} \simeq 30 \times (1+z) \mu\text{m}$) and the peak of dust emission in the host galaxy are within the covered range. Moreover, measurements of the complete far-IR to mm-wave SED will allow vastly improve the accuracy of photometric redshift estimates that have a rms error of $\simeq 0.2(1+z)$ with SPIRE alone [65]. This means that the PRISM survey will allow us to characterize to high statistical precision the evolution with redshift of the bolometric luminosity function. At $z \gtrsim 2$ it will be possible to investigate the evolution the relationships between star-formation and nuclear activity: what fraction of the bolometric energy radiated by star-forming galaxies is produced by accretion onto supermassive black holes in active galactic nuclei (AGN)? What are the evolution properties of far-IR selected AGNs? What fraction of them is associated with active star formation? Are the growth of central super-massive black hole formation and the build-up of stellar populations coeval? The substantially higher spatial resolution (thanks to the shorter wavelength channels) and the correspondingly higher positional accuracy compared to *Herschel*/SPIRE will greatly improve identification of reliable counterparts in other wavebands, necessary for a comprehensive understanding of the properties of detected galaxies.

Its all-sky coverage makes PRISM uniquely suited to study rare phenomena. Examples are the ‘maximum starburst’ galaxy at $z = 6.34$, detected by *Herschel*/SPIRE [92], or the most luminous star-forming hyper-luminous IR galaxies, such as the binary one, pinpointing a cluster of star-bursting proto-ellipticals at $z = 2.41$ discovered by Ivison et al. [56]. The $z = 6.34$ galaxy was found looking for ‘ultra-red’ sources with flux densities $S_{250\mu\text{m}} < S_{350\mu\text{m}} < S_{500\mu\text{m}}$. The PRISM survey will allow us to look for even redder sources, potentially at even higher redshifts, and will provide a test of our understanding of the interstellar medium and of star-formation under extreme conditions.

Strongly gravitationally lensed systems have long been very difficult to identify in sufficiently large numbers to be statistically useful. This situation changed drastically with the advent of (sub-)mm surveys. One of the most exciting *Herschel*/SPIRE results was the direct observational confirmation that almost all the galaxies brighter than $\simeq 100$ mJy at $500 \mu\text{m}$ are either strongly lensed or easily identifiable low- z spirals [72]. The surface density of strongly lensed high- z galaxies above this limit is $\simeq 0.3 \text{ deg}^{-2}$, implying that an all-sky survey can detect $\sim 10^4$ such systems. The fact that these sources are very bright makes redshift measurements with CO spectrometers and high resolution imaging with millimeter interferometers relatively easy. This will allow us to get detailed information on obscured star formation in the early Universe and on processes driving it in observing times hundreds of times shorter than would be possible without the help of gravitational amplification and with an effective source-plane resolution several times higher than can otherwise be achieved.

Large numbers of strongly lensed galaxies are also expected from large area optical surveys. It should be noted, however, that sub-mm selection has important distinctive properties. The selected lensed galaxies are very faint in the optical, while most foreground lenses are passive ellipticals, essentially invisible at sub-mm wavelengths so that there is no, or little, contamination between images of the source and of the lens. This makes possible the detection of lensing events with small impact parameters. Also, compared to the optical

selection, (sub-)mm selection allows us to probe earlier phases of galaxy evolution.

Optical spectroscopy of galaxies acting as lenses can be exploited to measure the mass distribution of their dark matter halos as a function of redshift. Note that *Euclid* will directly provide redshifts for the majority of the lenses out to $z \sim 1$ in its area. Taking into account the large number of newly identified strongly lensed galaxies, this will allow a direct test of the evolution of large-scale structure. Large samples of strongly lensed galaxies are also essential for many other astrophysical and cosmological applications [107].

PRISM will study the angular correlation function of detected sources with much better statistics than possible with *Herschel* extragalactic surveys that, altogether, cover little more than 2% of the sky. Also, the accurate photometric redshifts will allow us to follow evolution with cosmic time. Clustering properties measure the mass of dark matter halos associated with galaxies and are a powerful discriminant for galaxy formation and evolution models. Studies of the correlation function or of the power spectrum also establish occupation numbers of star-forming galaxies, and therefore their environments. In particular this study will allow us to detect high- z ‘proto-clusters’ of dusty galaxies. This amounts to investigating an earlier phase of the evolution of the most massive virialized structures we see today than can be done with data in any other waveband.

The PRISM clustering data will extend to much higher redshift than those provided by *Euclid*, whose wide-area survey will accurately map the galaxy distribution up to $z \sim 1$. The PRISM data will provide information at higher z , and primarily over the redshift range $2 < z < 3$, corresponding to the peak in star-formation activity. Moreover, optical and near-IR data severely underestimate the SFR of dust-obscured starbursts and may entirely miss these objects that are the main targets of far-IR/sub-mm surveys such as PRISM. Only the combination of PRISM and *Euclid* data will provide a complete view of the spatial distribution of galaxies and of how star formation is distributed among dark matter halos.

The PRISM sensitivity and spectral coverage will allow substantially improved measurements of the cosmic infrared background (CIB) spectrum with an accurate removal of all contaminating signals. PRISM will also measure, in a uniform way, the CIB power spectrum over an unprecedented range of frequencies and of angular scales (from ~ 10 arcsec to tens of degrees).

Radio sources: PRISM will extend the counts of radio sources, both in total and in polarized intensity, by at least one order of magnitude downwards in flux density compared to *Planck*. Above 217 GHz, the counts will be determined for the first time over a substantial flux density range with good statistics. This will allow the first investigation of evolutionary properties of radio sources at (sub-)mm wavelengths. PRISM will provide measurements of the spectral energy distribution (SED) of many thousands of radio sources and of multifrequency polarization properties for hundreds of them. The vast majority of these sources are expected to be blazars, and the accurate determination of their spectra will allow us to understand how physical processes occurring along relativistic jets shape the SED. For steep-spectrum sources we will obtain the distribution of break frequencies due to electron ageing, allowing an unbiased estimate of the distribution of radio source ages. Moreover, these observations will shed light on the relationship between nuclear radio emission and star formation activity in the host galaxies.

5 Inflation and CMB primordial B-modes

At the heart of modern cosmology is a set of initial conditions generated at very early times by what is known as *cosmic inflation*. During inflation, the Universe undergoes a period of ultra-rapid accelerated expansion, typically driven by a fundamental scalar field ϕ , with a potential energy $V(\phi)$ that dominates over its kinetic energy. Quantum fluctuations of space-time and the scalar field are amplified and stretched to cosmological scales resulting in a quasi-Gaussian stochastic distribution of density perturbations with amplitude A_S , and a scale dependence characterized by the *scalar spectral index*, $n_S \equiv 1 + d \ln A_S^2(k) / d \ln k$. Theory predicts that A_S and n_S will depend on the details of V and hence ϕ ; furthermore, interactions of ϕ with itself and with other fields induce cross-correlations between perturbation modes, leading to non-Gaussianity which can be detected in higher order statistics (bispectrum, trispectrum). Inflation also produces a bath of primordial gravitational waves characterized by an amplitude, A_T and the tensor spectral index $n_T = d \ln A_T^2(k) / d \ln k$. Remarkably, in the simplest models of inflation, the ratio between the tensor and scalar perturbations, r , is a direct probe of V in the early Universe: $r \equiv 16(A_T/A_S)^2 \simeq M_{Pl}^2(V'/V)^2$. Present observations estimate that $V^{1/4} = 3.3 \times 10^{16} r^{1/4}$ GeV, so that measuring r effectively translates into a measurement of the energy scale of inflation. A measurement of r , n_S and n_T can directly probe the physics of the early Universe for which there is a very rich phenomenology. Single field inflation models can relate r directly with the evolution

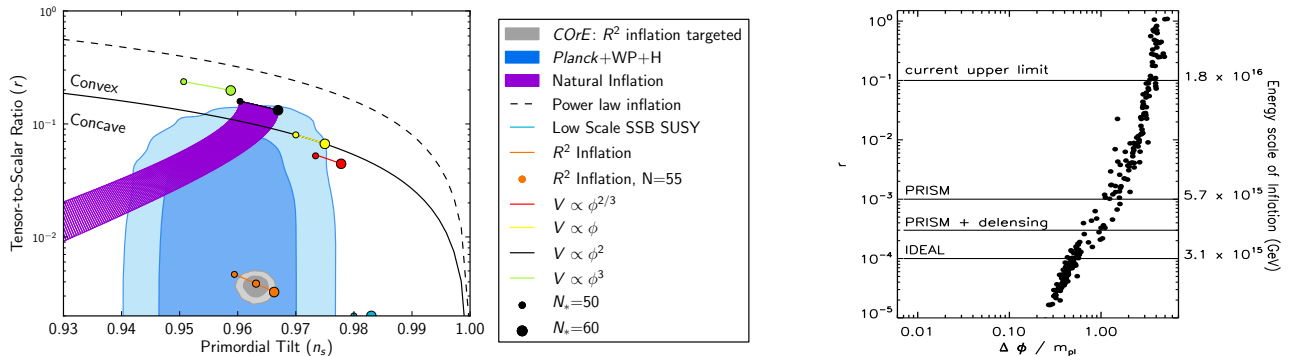


Figure 3: Left: Constraints on inflationary potentials from *Planck* and the predicted constraints from PRISM (not assuming de-lensing) for a fiducial value of $r = 5 \times 10^{-2}$ (adapted from [87]). Right: distribution of inflationary model parameters generated using a model independent approach that Monte-Carlo samples the inflationary flow equations. While these simulations cannot be interpreted in a statistical way (e.g., Kinney [63], Peiris et al. [78], Chongchitnan and Efstathiou [26]), they show that models cluster around attractor regions (adapted from [108]).

of ϕ at early times; indeed, for an inflationary expansion lasting enough to provide the observed level of homogeneity and isotropy, we have that $\Delta\phi/m_{\text{Pl}} \simeq (r/0.01)^{1/2}$. Multiple field inflation models arising in string theory and other proposals for unification at high energies, as well as particle and string production during the inflationary period and can lead to even higher values of r .

Primordial gravitational waves imprint a unique, as yet undetected, signature in the CMB polarization. CMB polarization is a spin-two field on the sky, and is decomposed into the equivalent of a gradient- the E-mode- and a curl- the B-mode. Gravitational wave fluctuations are visible as the B-mode polarization of the CMB and are the only primordial contribution to B relevant at the time of recombination. Hence a detection of B-modes is a direct probe of r and thus, the energy scale of inflation and other primordial energetic processes. Furthermore, in the simple case of slow-roll inflation we have that $r \approx -8n_T$. Additional detailed measurements of the shape of the temperature and polarization spectra will measure higher derivatives of the inflationary potential.

The 2013 release of *Planck* data has significantly improved previous constraints on inflationary models. In particular, and in the context of the simplest Λ CDM scenario, *Planck* results provide $n_s = 0.9624 \pm 0.0075$ and $r < 0.12$. These results are notable for the fact that exact scale invariance (i.e. $n_s = 1$) of primordial perturbations is ruled out at more than 5σ . When specific inflationary models are considered, *Planck* imposes significant constraints on the potential (Fig. 3), as discussed in Ref. [87]. Indeed *Planck* has shown that it is possible to test many inflation models using the CMB temperature data, yet even a forecast *Planck* limit $r < 0.05$ would leave many interesting models unprobed. Given that the stochastic background of gravity waves is the smoking gun of inflation it is therefore crucial to map as accurately as possible the CMB polarization and in particular characterize the B-mode angular power spectrum.

To forecast how well we would be able to measure the power spectrum of the B-modes, it is important to recognize that the foreground signal is likely to dominate the cosmological signal at low ℓ , where the direct information on r comes from. If we propagate the uncertainties connected to foreground contamination into the parameter error forecasts [108, 6, 9], we find that the proposed experimental set-up will enable us to explore most large field (single field) inflation models (i.e. where the field moves for $\geq M_P$) and to rule in or out all large-field models, as illustrated in the right-hand panel of Fig. 3.

As the work by Smith et al. [99] indicates (see Fig. 8), the instrumental sensitivity, angular resolution and, as a result, foreground control and subtraction will enable us to achieve a detailed mapping of the lensing signal, and in particular to implement de-lensing techniques for the measurement of r , improving by a factor of three our constraint on r . This implies that PRISM will detect $r \sim 3 \times 10^{-4}$ at more than 3σ . This performance is very close, within factors $\mathcal{O}(1)$, to what an ideal experiment (i.e. with no noise and no foregrounds) could achieve, allowing PRISM to *directly* probe physics at an energy scale that is a staggering twelve orders of magnitude larger than the center-of-mass energy at the Large Hadron Collider.

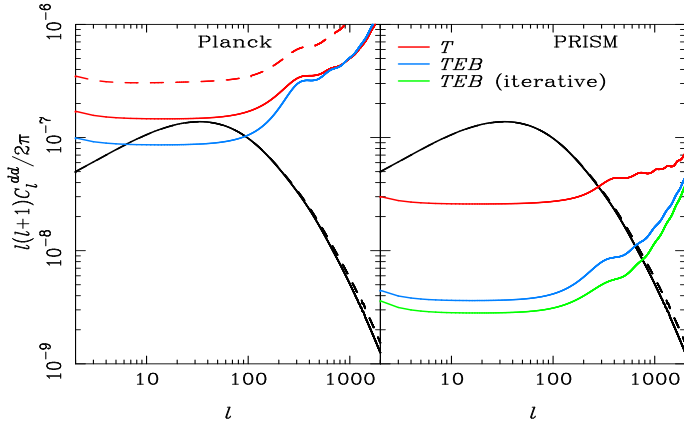


Figure 4: Reconstruction noise on the lensing deflection power spectrum forecast for the full *Planck* mission (four surveys; left) and PRISM (right) using temperature alone (red) and temperature and polarization (blue). For *Planck* we also show the approximate noise level for the temperature analysis of the nominal-mission data (red dashed) [82], and for PRISM, we also show the approximate noise level (green) for an improved iterative version of the reconstruction estimator. The deflection power spectrum is plotted based on the linear matter power spectrum (black solid) and with non-linear corrections (black dashed).

6 CMB at high resolution

The temperature anisotropies of the CMB have proved to be a remarkably clean probe of the high-redshift universe and have allowed the standard cosmological model to be tested to high precision. However, the accuracy of the recent results from *Planck*, based on the temperature anisotropies, are now close to being limited by errors in modelling extragalactic foregrounds. Fortunately, further progress can be made with the polarization anisotropies on small angular scales since the degree of polarization of the anisotropies is relatively larger there (around 4% by $l = 2000$) than the foreground emission. By surveying the full polarized sky in many frequency bands, and with uniform calibration, PRISM will fully exploit the small-scale polarization of the CMB, improving significantly on results currently obtained from the temperature and those conceivably obtainable in the future with ground-based experiments.

Probing the dark universe with CMB lensing: Gravitational lensing of the CMB provides a clean probe of matter clustering integrated to high redshift. Lensing can be reconstructed from the CMB anisotropies via specific non-Gaussian signatures imprinted by the lenses. *Planck* has detected lensing via this route at the 25σ level using the temperature anisotropies, but with low S/N per lensing mode. Polarization-based reconstructions from PRISM will be a major advance over *Planck*, achieving $S/N \gg 1$ over individual multipoles up to $l \approx 600$ over nearly the full sky (see Fig. 4). Significantly, PRISM can extract all of the information in the deflection power spectrum on scales where linear theory is reliable. To illustrate the power of the lensing measurements from PRISM in constraining physics that is inaccessible to the primary anisotropies alone due to degeneracies, we consider the mass of (light) neutrinos. Oscillation data constrain (squared) mass differences, and provide only lower bounds on the total mass summed over eigenstates: 0.06 eV and 0.1 eV for the normal and inverted hierarchy, respectively. These hierarchical limits provide natural targets for absolute mass measurements, but are well below the detection limits of current and future laboratory β -decay experiments. However, masses of these orders can be probed cosmologically via their effect on the clustering of matter. In w CDM models with massive neutrinos, we forecast a 1σ error of 0.04 eV for the summed mass. This constraint can be improved further by combining with near-future BAO measurements, for example by a factor of almost two using BOSS, at which point it becomes possible to distinguish between the normal and inverted hierarchies (in the hierarchical limits) [43].

Lensing constraints from PRISM would be highly complementary to those from upcoming optical cosmic shear surveys, e.g. *Euclid*. The systematic effects are quite different with non-linearities being much less of an issue for CMB lensing and there are no intrinsic-alignment effects. The combination of the two probes of mass is particularly promising, since it allows calibration of multiplicative bias effects such as due to PSF corrections in the optical. Cross-correlating CMB lensing with other probes of large-scale structure, such as galaxies, the Ly α forest or CIB clustering (see Sec. 4), also has exceptional promise, allowing self-calibration of the tracer’s bias relation at the sub-percent level.

Primordial non-Gaussianity: Non-Gaussianity (NG) is now demonstrably a robust quantitative probe of cosmological physics [84]. *Planck* results dramatically improved previous NG analyses, offering the most stringent test to date of inflationary theory (with $f_{\text{NL}}^{\text{loc}} = 2.7 \pm 5.8$) while also detecting for the first time ISW-lensing and diffuse point source bispectra. Already *Planck* offers enticing clues about the nontrivial ‘shape’ of the CMB bispectrum of our universe (see Fig. 5), the origin of which is yet to be explained. PRISM would offer the highest precision reconstructions of the CMB temperature and polarization bispectra and trispectra, which will provide a decisive and unambiguous probe of primordial cosmology back to the *Planck* era. At the same time, PRISM NG data will open new windows for investigating dark energy and gravitational physics,

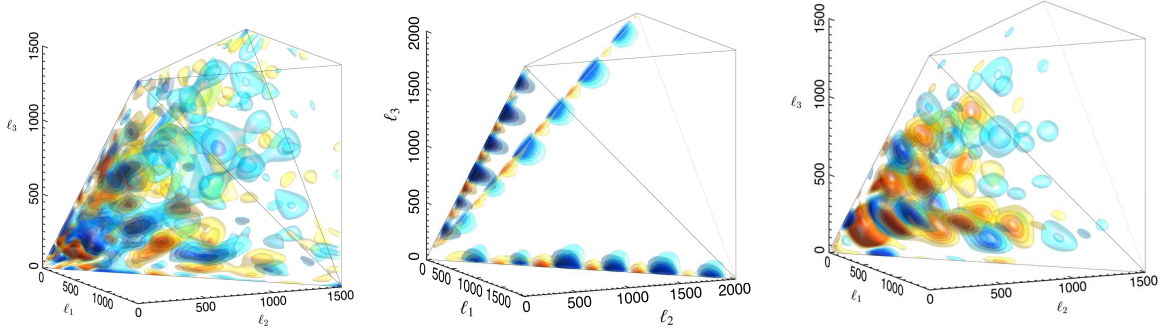


Figure 5: *Planck* CMB temperature bispectrum [84] (left) and primordial (right) and late-time (middle) non-Gaussian shapes [84, 83]. Note the periodic CMB ISW-lensing signal (middle) in the squeezed limit along the edges, which is seen at the 2.5σ level in the *Planck* bispectrum on the left. Scale-invariant signals predicted by many inflationary models are strongly constrained by the *Planck* bispectrum, although ‘oscillatory’ and ‘flattened’ features hint at new physics. An example of an inflationary ‘feature’ model is shown on the right. PRISM will probe these hints with an order of magnitude more resolved triangle configurations.

as well as astrophysical sources, large-scale structure and galactic history.

A unique advantage of the CMB for probing NG is its ability to recognize the distinct patterns that physical mechanisms leave in the *shape* of higher-order correlators (Fig. 5). PRISM will allow a vastly enhanced exploration of physically-predicted NG shapes compared to any other projected probe of NG. For example, the constraint volume in bispectrum space spanned by the local, equilateral and flattened bispectra will reduce by a factor of 75 compared to the current *Planck* volume, and a factor of 30 over that predicted from the full-mission *Planck* data (including polarization). From polarization maps alone (which provide information independent of the temperature maps), we expect a volume reduction factor from the full-mission *Planck* data to PRISM of order 110. Moreover, local-model trispectrum parameters could be measured with a precision $\Delta g_{\text{NL}} = 3 \times 10^4$ and $\Delta \tau_{\text{NL}} = 1 \times 10^2$ [98]. These could investigate consistency conditions between polyspectra, which can be used to test large classes of multi-field inflation models in addition to single-field inflation. There are other alternative inflationary scenarios for which an observable non-Gaussian signal is quite natural, e.g. those with features or periodicity in the inflationary potential (Fig. 5). Each of these models has a distinct fingerprint, many uncorrelated with the standard three primordial shapes and, in all cases, PRISM would significantly improve over present *Planck* constraints, offering genuine discovery potential. Beyond searches for primordial NG, PRISM is guaranteed to make important observations of late-time NG. For example, it will decisively detect and characterize the lensing-ISW correlation, driven by dark energy, achieving a 9σ detection, resulting in a new probe of dark energy physics from the CMB alone.

Parameters from high-resolution polarization spectra: PRISM will measure the CMB angular power spectra with outstanding precision to small angular scales. In particular, in the 105–200 GHz frequency range, the relatively clean EE polarization spectrum is cosmic-variance limited to $l = 2500$ (and the BB spectrum from lensing to $l = 1100$). Such a remarkable measurement of the polarization of the CMB damping tail will be an invaluable source of information on the shape of the primordial power spectrum and the fundamental matter content of the Universe. For example, in Λ CDM models, the spectral index and its running will be measured more precisely than with current *Planck* data by factors of five and three, respectively. The Hubble constant (a point of tension between *Planck* data and direct astrophysical measurements) will be measured a factor of 10 better than currently (and 2.5 times better than expected from the full *Planck* data). Fundamental questions about the matter content include the effective number of relativistic species N_{eff} , for which a non-standard value (which can relieve the *Planck*– H_0 tension) could be due to sterile neutrinos, as advocated in particle physics to explain certain anomalies in the neutrino sector, the helium abundance Y_{P} , which provides a clean test of standard BBN, the neutrino mass, and the dark matter annihilation cross-section.. In one-parameter extensions of Λ CDM, PRISM will measure N_{eff} to 2% precision and Y_{P} to 1%. These values indicate that a 2σ anomaly hinted at by *Planck* could be confirmed decisively with PRISM. Moreover, from its measurement of the B -mode power spectrum, PRISM should extend the range of sensitivity to cosmic strings by an order of magnitude over the recent *Planck* constraints [83, 1].

7 CMB spectral distortions

COBE/FIRAS has shown that the average CMB spectrum is extremely close to a perfect blackbody, with

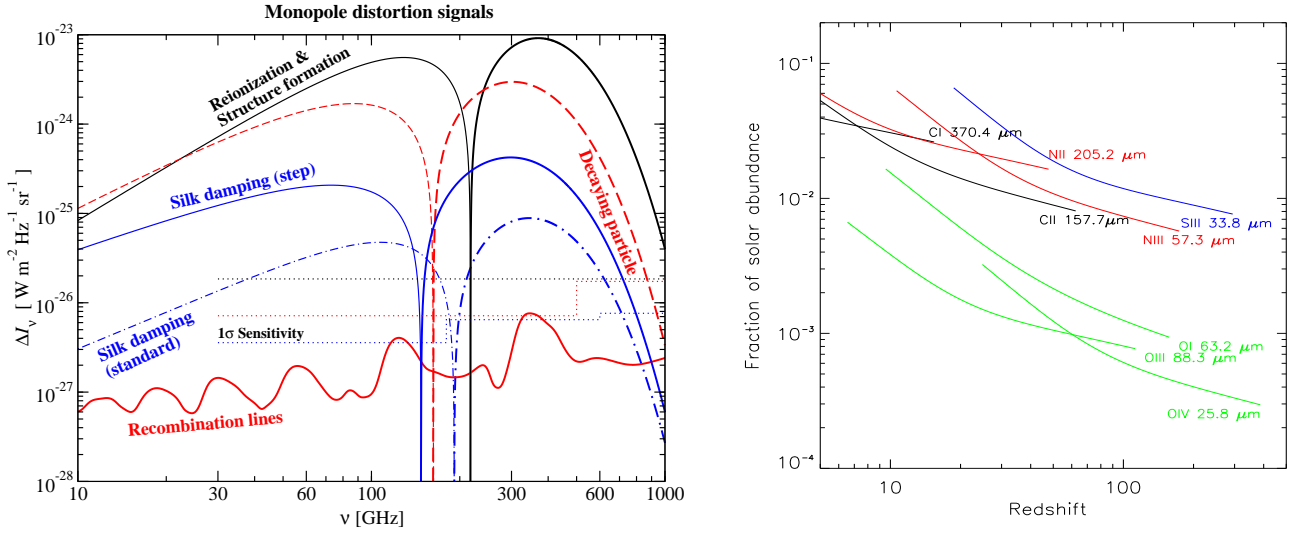


Figure 6: Left: spectral distortions for different scenarios. Thick lines denote positive, and thinner lines negative signal. The 1σ sensitivities of PRISM for different designs are also indicated. Right: projected constraints on different metal ions.

possible departures limited to $\Delta I_\nu/I_\nu \lesssim \text{few} \times 10^{-5}$ [68, 37]. This places very tight constraints on the thermal history of our Universe, ruling out cosmologies with extended periods of significant energy release at redshifts $z \lesssim \text{few} \times 10^6$ [109, 101, 54, 30, 12, 49, 16, 22, 58]. There are, however, a large number of astrophysical and cosmological processes that cause (inevitable) spectral distortions of the CMB at a level that has only come within reach of present-day technology. With PRISM a unexplored window to the early universe will be opened, allowing detailed studies of (see Fig. 6 for illustration):

Reionization and structure formation: Radiation from the first stars and galaxies [53, 2], feedback by supernovae [73] and structure formation shocks [102, 14, 71] heat the IGM at redshifts $z \lesssim 10 - 20$, producing hot electrons that up-scatter CMB photons, giving rise to a Compton y -distortion with average amplitude $\Delta I_\nu/I_\nu \simeq 10^{-7} - 10^{-6}$. This signal will be detected at more than a 100σ with PRISM, providing a sensitive probe of reionization physics and delivering a census of the missing baryons in the local Universe. PRISM furthermore has the potential to separate the spatially varying signature caused by the WHIM and proto-clusters [111]. It also offers a unique opportunity to observe the free-free distortion associated with reionization, providing a complementary way to study the late evolution of inhomogeneities [88].

Decaying and annihilating relics: The CMB spectrum allows placing tight limits on decaying and annihilating particles in the pre-recombination epoch [50, 29, 69, 17, 22]. This is especially interesting for decaying particles with lifetimes $t_X \simeq 10^8 - 10^{10}$ sec, as the exact shape of the distortion encodes when the decay occurred [22, 59, 18, 19]. PRISM therefore provides an unprecedented probe of early-universe particle physics, with many natural particle candidates found in supersymmetric models [36, 35].

Constraining the inflaton: Silk-damping of small-scale perturbations gives rise to CMB distortions [101, 28, 3, 52] which directly depend on the shape and amplitude of the primordial power spectrum at scales $0.6 \text{ kpc} \lesssim \lambda \lesssim 1 \text{ Mpc}$ (or multipoles $10^5 \lesssim \ell \lesssim 10^8$) [24, 62]. This allows constraining the trajectory of the inflaton at stages unexplored by ongoing or planned experiments [23, 89, 60], extending our reach from 7 e-folds of inflation probed with the CMB anisotropies to a total of 17 e-folds. The signal is also sensitive to the difference between adiabatic and isocurvature perturbations [3, 51, 31, 20], as well as primordial non-Gaussianity in the ultra squeezed-limit, leading to a spatially varying spectral signal that correlates with CMB temperature anisotropies at large angular scales [75, 38]. A competing monopole signal, characterized by a *negative* μ - and y -parameter, is introduced by the adiabatic cooling of ordinary matter [16, 22, 61], which PRISM will also be sensitive to.

Metals during the dark ages: Any scattering of CMB photons after recombination blurs CMB anisotropies at small scales, while producing new anisotropies at large scales. Electrons from the reionization epoch are the dominant source of optical depth, causing a signature already detected by WMAP and Planck [7, 85]. The resonant scattering of CMB photon by fine structure lines of metals and heavy ions produced by the first stars adds to this optical depth, making it frequency-dependent [5]. By comparing CMB temperature and polarization anisotropies at different frequencies one can thus determine the abundances of ions such as OI, OIII, NII, NIII, CI, CII at different redshifts [46, 48]. Furthermore, UV radiation emitted by the first

stars can push the OI $63.2\,\mu\text{m}$ and CII $157.7\,\mu\text{m}$ transitions out of equilibrium with the CMB, producing a distortion $\Delta I_\nu/I_\nu \simeq 10^{-8} - 10^{-9}$ due to fine structure emission [41, 47], providing yet another window to reionization within reach of PRISM.

Cosmological recombination radiation: The recombination of H and He introduces distortions [110, 77, 34] at redshifts $z \simeq 10^3 - 10^4$, corresponding to $\simeq 260\,\text{kyr}$ (HI), $\simeq 130\,\text{kyr}$ (He I), and $\simeq 18\,\text{kyr}$ (He II) after the big bang [94, 21, 95]. The signal is small ($\Delta I_\nu/I_\nu \simeq 10^{-9}$) but its unique spectral features promise an independent path to determination of cosmological parameters (like the baryon density and *pre-stellar* helium abundance) and direct measurements of the recombination dynamics, probing the Universe at stages well before the last scattering surface [100]. The effect on the TT power spectrum introduced by resonance scattering of CMB photons by the first lines of the Balmer and Paschen series [93, 45, 48] will also be detectable with PRISM, providing an additional opportunity to directly constrain the recombination history and obtain independent determinations of cosmological parameters (e.g. Ω_b or Ω_m).

Non-Gaussianity: CMB spectral distortions can also open up a new window to primordial NG [76]. We know almost nothing about NG on the small scales that can be probed via these observations. In particular, the cross-correlation between μ -type distortions and CMB anisotropies is naturally sensitive to the very squeezed limit of the primordial bispectrum (probing scales as small as $50 \leq k\,\text{Mpc} \leq 10^4$). Also, the power spectrum of μ -distortions can probe the trispectrum of primordial fluctuations. Such measurements can be particularly constraining for models where the primordial power spectrum grows on small scales (see e.g. [25]), and values $f_{\text{NL}}^{\text{loc}} < 1$ can be achieved. Also, μ -type distortions can shed light into modifications of the initial state of quantum fluctuations. For a large class of inflationary models characterized by a non-Bunch-Davies vacuum (whose bispectrum is enhanced in the squeezed limit with respect to the local form) a high S/N can be achieved [39].

All these examples demonstrate that the CMB spectrum provides a rich and unique source of complementary information about the early Universe, with the certainty for a detection of spectral distortions at a level within reach of PRISM’s capabilities. The CMB spectrum will also establish interesting constraints on the power spectrum of small-scale magnetic fields [57], cosmic strings [74, 104, 105], evaporating primordial black holes [13], decay of vacuum energy density [4, 11, 29], and other new physics [67, 10], to mention a few more exotic examples. Deciphering all these signals will be a big challenge for the future. This area has great potential for new discoveries and for providing new independent constraints on unexplored processes cannot be explored by other means.

8 Structure of the dusty magnetized Galactic ISM

The data analysis is still on-going but it is already clear that Herschel and Planck will have a profound and lasting impact on our understanding of the interstellar medium and star formation. PRISM holds even greater promise for breakthroughs. Dust and synchrotron radiation are the dominant contributions to the sky emission and polarization to be observed by PRISM. Dust emission is an optically thin tracer of the structure of interstellar matter across the ionized, atomic, and molecular states of hydrogen. Synchrotron radiation traces the magnetic field over the whole volume of the Galaxy, while dust polarization traces the magnetic field within the thin star forming disk, where the interstellar matter is concentrated. PRISM will image these two complementary tracers with unprecedented sensitivity and angular resolution. It will also provide all-sky images of spectral lines, which are key diagnostics of interstellar gas physics. No other initiative offers a comparable imaging capability of interstellar components over as wide a range of scales. In the following subsections we detail how PRISM will address three fundamental questions of Galactic astrophysics: (1) What are the processes that structure the interstellar medium? (2) What role does the magnetic field play in star formation? (3) What are the processes that determine the composition and evolution of interstellar dust?

8.1 Structure of interstellar medium

Herschel far infrared observations have provided astronomers new insight into how turbulence stirs up the interstellar gas, giving rise to a filamentary, web-like structure within the diffuse interstellar medium and molecular clouds. PRISM will extend the Herschel dust observations to the whole sky and provide unique data on emission lines key to quantifying physical processes. The spectral range of PRISM includes atomic and molecular lines that serve as diagnostics of the gas density and temperature, its chemical state, and

energy budget. Herschel has observed these lines along discrete lines of sight with very limited imaging. By mapping these lines and dust emission over the whole sky at an angular resolution comparable to that of Herschel, PRISM will probe the connection between the structure of matter and gas cooling across scales.

PRISM sky maps will provide multiple clues to characterize the physical processes that shape interstellar matter. The CII, CI, and OI fine structure lines and rotational lines of CO and H₂O are the main cooling lines of cold neutral medium and molecular clouds and probe local physical conditions and the exchange of energy associated with the formation of molecular gas within the diffuse interstellar medium and of stars within molecular clouds. The NII lines at 122 and 205 μm are spectroscopic tracers of the ionized gas. These lines are essential for distinguishing the contribution of neutral and ionized gas to the CII emission. PRISM will have the sensitivity to image the CII line emission at sub-arcminute resolution even at the Galactic poles. The CII map can be combined with HI and dust observations to study the formation of cold gas from the warm neutral medium through the thermal instability. This analysis will probe the expected link, yet to be confirmed observationally, between the small-scale structure of the cold interstellar medium and gas cooling. The CII line emission is also key to studying the formation of molecular gas by tracing the CO-dark H₂ gas [79]. In star forming molecular clouds, the CO, CI, OI, and H₂O lines are the key tracers of the processes creating the initial conditions of star formation and of the feedback from newly formed stars on their parent clouds.

8.2 Galactic magnetic field and star formation

Star formation results from the action of gravity, counteracted by thermal, magnetic and turbulent pressures [44]. For stars to form, gravity must locally become the dominant force. This happens when the turbulent energy has dissipated and matter has condensed without increasing the magnetic field by a comparable amount. What are the processes that drive and regulate the rate at which matter reaches this stage? This is a long-standing question to which theorists have over the decades offered multiple explanations, focusing on either ambipolar diffusion, turbulence, or magnetic reconnection to decouple matter from the magnetic field and allow the formation of condensations of gas in which stars may form [27].

PRISM observations of the polarization in the far-IR and sub-mm will provide unique clues to understand the role of the magnetic field in star formation. Compared to synchrotron radiation and Faraday rotation, dust polarization images the structure of the magnetic field through an emission process tracing matter. It is best suited to characterize the interplay between turbulence, gravity, and the Galactic magnetic field. The PRISM data will provide unique data to study magneto-hydrodynamical turbulence because it will drastically increase the spectral range of accurately probed magneto-hydrodynamical modes. The data will provide unprecedented statistical information to characterize the energy injection and energy transfer down to dissipation scales.

Polarization data from the PRISM survey will have the sensitivity and angular resolution required to map continuously the Galactic magnetic field over the whole sky down to sub-arcminute resolution even at the Galactic poles. The wide frequency range of the mission will measure polarization for separate emission components with distinct temperatures along the line of sight. PRISM will open an additional perspective on the structure of the magnetic field in molecular clouds, independent of grain alignment, by imaging the polarization of CO emission in multiple rotational lines [40]. No project offers comparable capabilities. Planck has provided the first all-sky maps of dust polarization with 5' resolution but the data is sensitivity limited even at the highest Planck frequency (353 GHz). Ground based telescopes at sub-mm and millimeter wavelengths of bright compact sources at arc-second resolution with e.g. ALMA will be complementary to the whole sky survey of extended emission from the diffuse interstellar medium and molecular clouds that PRISM can uniquely do.

8.3 Nature of interstellar dust

The combination of spectral and spatial information provided by PRISM will provide new tools for studying the interstellar dust, in particular its nature and its evolution. Dust properties (e.g., size, temperature, emissivity) are found to vary from one line of sight to another within the diffuse interstellar medium and molecular clouds. These observations indicate that dust grains evolve in a manner dependent on their environment within the interstellar medium. They can grow through the formation of refractory or ice mantles, or by coagulation into aggregates in dense and quiescent regions. They can also be destroyed by fragmentation and erosion of their mantles under more violent conditions. The composition of interstellar

dust reflects the action of interstellar processes, which contribute to break and re-build grains over timescales much shorter than the timescale of injection by stellar ejecta. While there is broad consensus on this view of interstellar dust, the processes that drive its evolution in space are poorly understood [32]. Understanding interstellar dust evolution is a major challenge in astrophysics underlying key physical and chemical processes in interstellar space. In particular, to fully exploit the PRISM data we will need to characterize where in the interstellar medium grains are aligned with respect to the Galactic magnetic field and with what efficiency.

Large dust grains (size > 10 nm) dominate the dust mass. Within the diffuse interstellar medium, these grains are cold ($\sim 10 - 20$ K) and emit within the PRISM frequency range. Dipole emission from small rapidly spinning dust particles constitutes an additional emission component, known as anomalous microwave emission. Magnetic dipole radiation from thermal fluctuations in magnetic nano-particles may also be a significant emission component over the frequency range relevant to CMB studies [33]. To achieve the PRISM objectives on CMB polarization, it is necessary to characterize the spectral dependence of the polarized signal from each of these dust components with high accuracy across the sky. This is a challenge but also a unique opportunity for dust studies. The spectral energy distribution of dust emission and the polarization signal can be cross-correlated with the spectral diagnostics of the interstellar medium structure to characterize the physical processes that determine the composition and evolution of interstellar dust. The same data analysis will also elucidate the physics of grain alignment.

PRISM will also probe the zodiacal dust emission from within our solar system. The fact that PRISM scans a substantial portion of the sky each day allows for a three-dimensional tomographic mapping of the zodiacal emission. Understanding zodiacal emission is crucial both to understanding our solar system and to carrying out a complete foreground separation.

9 Strawman mission concept

The science program above requires measuring the sky brightness and polarization at high angular resolution and in many frequency bands across a wide spectral range. It also requires measuring the absolute spectrum of the sky background with moderate angular and spectral resolution. As a baseline, we propose to perform the best possible spectro-polarimetric sky survey in the 30-6000 GHz frequency range with two instruments optimized for best joint performance sharing a single platform in orbit around the Sun-Earth L2 Lagrange point: (1) a *polarimetric imager* (PIM) observing with about 30 broad and 300 narrow spectral bands with a diffraction-limited angular resolution and a sensitivity limited by the photon noise of the sky emission itself; and (2) an *absolute spectro-photometer* (ASP) that will measure sky emission spectra with a spectral resolution between 500 MHz and 15 GHz and an angular resolution of about 1.4° . These complementary instruments will map simultaneously the absolute sky intensity and polarization with high sensitivity and with high spectral or spatial resolution. The data from both instruments can be binned (in frequency) and smoothed to obtain matching observations with $\delta\nu/\nu \approx 0.25$ and 1.4° resolution, allowing on-sky inter-calibration on large scales (and hence absolute calibration of the PIM). This will also enable correction of the ASP spectra from foreground contamination using high resolution component maps extracted from PIM data (e.g., large clusters y -distortion in the ASP data and line emission from emitting regions unresolved in the coarse resolution ASP maps).

As the scientific outcome of this mission depends on the complementarity of both instruments and on the control of systematic errors, a careful optimization of the ASP and the PIM (number and bandwidth of spectral bands vs. sensitivity) and of the mission (scanning strategy, joint analysis tools) with comprehensive simulations is an essential future phase of the mission study.

The focal planes of both instruments will be cooled to 0.1K using a cryogenic system adapted from that of *Planck*, with continuous recycling of the gases for an improved mission duration of 4 years (baseline) or longer.

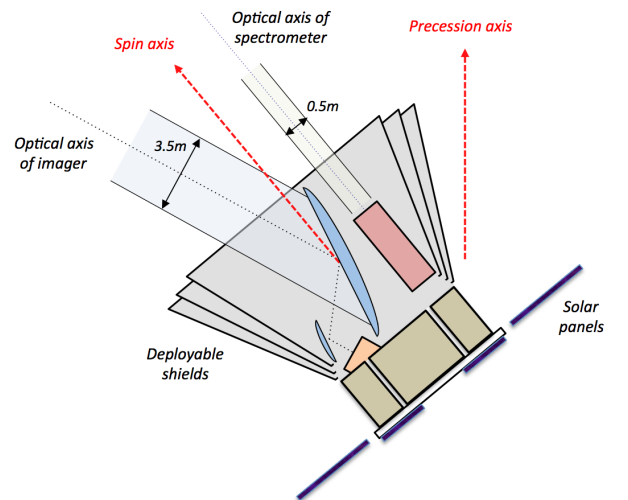


Figure 7: The PRISM spacecraft with its two instruments: PIM, with a 3.5-diameter telescope with a FOV at $\sim 30^\circ$ from the spacecraft spin axis, and ASP, aligned with the spin axis.

9.1 Instruments

ν_0	range	$\Delta\nu/\nu$	n_{det}	θ_{fwhm}	σ_I per det 1 arcmin		$\sigma_{(Q,U)}$ per det 1 arcmin		main molec. & atomic lines
GHz	GHz				μK_{RJ}	μK_{CMB}	μK_{RJ}	μK_{CMB}	
30	26-34	.25	50	17'	61.9	63.4	87.6	89.7	HCN & HCO ⁺ at 89 GHz CO at 110-115 GHz
36	31-41	.25	100	14'	57.8	59.7	81.7	84.5	
43	38-48	.25	100	12'	53.9	56.5	76.2	79.9	
51	45-59	.25	150	10'	50.2	53.7	71.0	75.9	
62	54-70	.25	150	8.2'	46.1	50.8	65.2	71.9	
75	65-85	.25	150	6.8'	42.0	48.5	59.4	68.6	
90	78-100	.25	200	5.7'	38.0	46.7	53.8	66.0	
105	95-120	.25	250	4.8'	34.5	45.6	48.8	64.4	
135	120-150	.25	300	3.8'	28.6	44.9	40.4	63.4	
160	135-175	.25	350	3.2'	24.4	45.5	34.5	64.3	
185	165-210	.25	350	2.8'	20.8	47.1	29.4	66.6	HCN & HCO ⁺ at 177 GHz
200	180-220	.20	350	2.5'	18.9	48.5	26.7	68.6	
220	195-250	.25	350	2.3'	16.5	50.9	23.4	71.9	CO at 220-230 GHz HCN & HCO ⁺ at 266 GHz
265	235-300	.25	350	1.9'	12.2	58.5	17.3	82.8	
300	270-330	.20	350	1.7'	9.6	67.1	13.6	94.9	CO, HCN & HCO ⁺
320	280-360	.25	350	1.6'	8.4	73.2	11.8	103	
395	360-435	.20	350	1.3'	4.9	107	7.0	151	CO, HCN & HCO ⁺ C-I, HCN, HCO ⁺ , H ₂ O, CO
460	405-520	.25	350	1.1'	3.1	156	4.4	221	
555	485-625	.25	300	55"	1.6	297	2.3	420	CO, HCN & HCO ⁺
660	580-750	.25	300	46"	0.85	700	1.2	990	
					nK _{RJ}	kJy/sr	nK _{RJ}	kJy/sr	
800	700-900	.25	200	38"	483	9.5	683	13.4	N-II at 1461 GHz
960	840-1080	.25	200	32"	390	11.0	552	15.6	
1150	1000-1300	.25	200	27"	361	14.6	510	20.7	
1380	1200-1550	.25	200	22"	331	19.4	468	27.4	
1660	1470-1860	.25	200	18"	290	24.5	410	34.7	
1990	1740-2240	.25	200	15"	241	29.3	341	41.5	C-II at 1900 GHz N-II at 2460 GHz
2400	2100-2700	.25	200	13"	188	33.3	266	47.1	
2850	2500-3200	.25	200	11"	146	36.4	206	51.4	O-III at 3393 GHz
3450	3000-3900	.25	200	8.8"	113	41.4	160	58.5	
4100	3600-4600	.25	200	7.4"	98	50.8	139	71.8	O-I at 4765 GHz O-III at 5786 GHz
5000	4350-5550	.25	200	6.1"	91	70.1	129	99.1	
6000	5200-6800	.25	200	5.1"	87	96.7	124	136	

Table 1: The 32 broad-band channels of the polarized imager with a total of 7600 detectors. Sensitivities are averages for sky regions at galactic latitude and ecliptic latitude both higher than 30°. A detector noise level equal to the sky photon noise is assumed. The mission sensitivity per frequency channel is the sensitivity per detector divided by $\sqrt{n_{det}}$.

The polarimetric imager: The optical configuration relies on a dual off-axis mirror telescope with a 3.5 m projected aperture diameter primary and a 0.8 m diameter secondary coupled to a multi spectral band polarimeter. The broad-band PIM comprises 32 main channels of $\delta\nu/\nu \approx .25$ relying on dual-polarized pixel arrays (Table 1. At frequencies below 700 GHz, the emphasis is on the sensitivity and control of systematics for CMB and SZ science.

The whole frequency range will also be covered at higher spectral resolution ($\delta\nu/\nu \approx .025$) to map spectral lines. The ~ 300 frequency channels (not listed in Table 1) will be obtained using antenna coupled bolometers and channelizers to split the spectral band of each broad-band horn into 5-10 narrow frequency bands, with similar numbers of narrow-band and broad-band detectors. The sensitivity to continuum emission per detector is reduced in the narrow-band channels as compared to the broad-band channels, but the sensitivity to spectral lines is better by a factor of about 2-3.

The absolute spectrophotometer: A Martin-Puplett Fourier Transform Spectrometer (FTS) will allow for a large throughput and sensitivity, differential measurements (the sky is compared to an internal blackbody calibrator, as in the COBE-FIRAS), and a variable spectral resolution. Dichroics at the two output ports can optionally split the full 30 - 6000 GHz range into sub-bands with reduced photon noise. The instrument is cooled at 2.7K, so that the bolometric detector sensitivity is limited by photon noise from the sky. Two operating modes are available: high-resolution ($\Delta\nu \sim 0.5$ GHz) and low-resolution ($\Delta\nu \sim 15$ GHz). The sensitivity of the high-resolution mode is 30 times worse than for the low-resolution mode. The instrument beam is aligned with the spin axis of the satellite, so that precession has a negligible effect during the

interferogram scan ($\sim 1\text{s}/10\text{s}$ long in the low-res/high-res mode). The main characteristics for three possible configurations of the instrument are detailed in Table 2.

band (GHz)	resolution (GHz)	$A\Omega$ (cm^2sr)	background (pW)	NEP ν ($\text{W}/\text{m}^2/\text{sr}/\text{Hz}\times\sqrt{\text{s}}$)	global 4-yr mission sensitivity ($\text{W}/\text{m}^2/\text{sr}/\text{Hz}$)
30-6000	15	1	150	1.8×10^{-22}	1.8×10^{-26}
30-500	15	1	97	7.0×10^{-23}	7.2×10^{-27}
500 - 6000	15	1	70	1.7×10^{-22}	1.7×10^{-26}
30-180	15	1	42	3.5×10^{-23}	3.6×10^{-27}
180-600	15	1	57	6.3×10^{-23}	6.5×10^{-27}
600-3000	15	1	20	7.4×10^{-23}	7.6×10^{-27}
3000-6000	15	1	28	1.6×10^{-22}	1.6×10^{-26}

Table 2: Performance of the FTS in three possible configurations for photon-noise limited detectors. With an entrance pupil 50 cm in diameter, the baseline throughput is $\sim 1\text{cm}^2\text{sr}$ and the angular resolution 1.4° . The theoretical monopole sensitivity for each spectral bin is reported in the last column assuming 4 years of observation and 75% useful sky. The actual sensitivity, taking into account efficiency factors can be 2-3 times worse. Line 1 is a configuration with an ultra-wide spectral coverage obtained with one detector in both output ports. In lines 2-3 the detectors at the output ports are sensitive to different bands. In lines 4-7 each output port is split in two sub-bands using dichroics to minimize photon noise in the low-frequency bins.

Using detectors with $A\Omega \sim 1\text{cm}^2\text{sr}$, and angular resolution $\sim 1.4^\circ$, we estimate that the CIB can be measured with $S/N = 10$ in a fraction of a second at 1500 GHz and in ~ 10 seconds at 140 GHz, while a y -distortion $\sim 10^{-8}$ can be measured with $S/N = 10$ at 350 GHz in two hours of integration. Recombination lines could be measured integrating over the whole mission, if the overall stability of the instrument and the quality of the reference blackbody are sufficient.

The main issue for this instrument is the control of systematic effects. The instrument design allows for a number of zero tests and cross-checks on the data. The main problem is to control the blackness of the reference and calibration blackbodies with the required accuracy. Reflectivities lower than $R = -50/-60$ dB have been obtained in the frequency range of interest in the Planck and ARCADE references. We plan to achieve $R < -70$ dB building on these experiences, with a combination of electromagnetic simulations and laboratory emissivity measurements on improved shapes and space-qualified materials.

9.2 Scan strategy

The observing strategy must provide: (1) full sky coverage for both instruments; (2) cross-linked scan paths and observation of all sky pixels in many orientations for all detectors of the PIM; (3) fast scanning for the detectors of the PIM to avoid low-frequency drifts; (4) slow scanning for the ASP field of view (FOV) to allow for few seconds long interferogram scans with negligible depointing; (5) avoidance of direct solar radiation on the payload. All these requirements can be achieved with a spinning spacecraft for which the FOV of the ASP is aligned along the spin axis, while the FOV of the PIM is offset by an angle $\theta_{\text{spin}} \approx 30^\circ$ (Fig. 7). For each rotation of the spacecraft (with a spinning frequency ω_{spin} of a few RPM), the FOV of the PIM detectors scan circles of diameter $\approx 2\theta_{\text{spin}}$ while the FOV of the APS rotates in place and does not move. A slow precession of the spin axis (with a period between a few hours and one day) with a precession angle $\theta_{\text{prec}} \approx 45^\circ$ results in slow scans of the FOV of the ASP on large circles of diameter $2\theta_{\text{prec}}$. Finally, the precession axis is slowly displaced by about 1° per day along the ecliptic plane to maintain the payload away from the solar direction, and slowly moves perpendicularly to the ecliptic plane to map ecliptic poles at regular periods. Deployable screens will isolate the payload from the radiative heat from the Sun, providing a first stage of passive cooling of instruments and telescope to a temperature of ≈ 40 K.

9.3 Experimental challenges

Telescope temperature: There is substantial sensitivity improvement, mainly for frequencies above 200 GHz, if the telescope is actively cooled to 4 K (objective of PRISM) instead of 40 K, which can be achieved by passive cooling. PRISM will benefit from the development activities for the SPICA mission, the telescope of which is based on a 3.5 m diameter primary cooled to 5 K.

Polarization modulation: The baseline, similar to the solution proposed in the previous SAMPAN and EPIC studies, relies on the scanning strategy and the rotation of the entire payload. However alternate

ν_c range	Req. NEP	Req. τ	Focal Plane Technology			
[GHz]	$[10^{-18} \text{ W}/\sqrt{\text{Hz}}]$	[ms]	Detector technology		Optical coupling	
			Baseline	Backup	Baseline	Backup
30-75	3.3 – 5.7	2.96 – 1.18	TES	HEMT	MPA/CSA	HA
90-320	4.6 – 7	1.18 – 0.4	TES	KIDS	HA+POMT	MPA
395-660	0.94 – 3.1	0.4 – 0.13	TES	KIDS	MPA/CSA	LHA
800-6000	0.011 – 0.63	0.13 – 0.01	KIDS	HEB/CEB	MPA/CSA	LHA

Table 3: Required NEP and time constants for various frequency ranges and corresponding baseline and backup focal plane technology. TES: Transition Edge Sensors (Technology Readiness Level 5); HEMT: High Electron Mobility Transistor (TRL 5); KID: Kinetic Inductance Detector (TRL 5); HEB: Hot Electron Bolometer (TRL 4); CEB: Cold Electron Bolometer (TRL 3); HA: Horn Array (TRL 9); LHA: Lithographed Horn Array (TRL 5); MPA: Multichroic Planar Antenna (TRL 4); CSA: Crossed Slot Antenna (TRL 5); POMT: Planar Ortho-Mode Transducer (TRL 5)

strategies such as the use of a half-wave plate in front of the focal plane (the receivers being the major source of instrumental polarization) could be considered during a trade-off analysis.

Detectors: Direct detectors (such as TES bolometers, CEBs or KIDs) are the most sensitive detectors at mm wavelengths. Bolometers have achieved photon noise limited in-flight performance with the Planck [106] and Herschel [42] missions. Large bolometer arrays with thousands of pixels are currently used on large ground-based telescopes. They are currently not proven as a viable technology for 30 to 70 GHz but it is likely that their efficiency will improve in the next few years at low frequencies. For instance studies [64] have shown that 70 GHz CEBs could lead to NEPs of few $10^{-18} \text{ W}\cdot\text{Hz}^{1/2}$. As an alternate solution, the PRISM instruments could take advantage of the breakthroughs recently achieved in cryogenic HEMT technology, with sensitivities predicted to reach 2-3 times the quantum limit up to 150-200 GHz (instead of 4-5 times up to 100 GHz so far). In addition, these devices allowing for cryogenically cooled miniaturized polarimeter designs will simplify their thermo-mechanical design. Hence, while a single detector technology throughout the instruments would be preferable, the option of using a combination of HEMTs and bolometers remains open (Table 3).

Detector time constants: The fast scanning of the PRISM mission requires fast detector time constants, of order 1 ms at 100 GHz, down to $\sim 10 \mu\text{s}$ at 6 THz. These time constants are challenging (especially at high frequencies), but have already been achieved with recent TESs, KIDs or CEBs.

9.4 Ancillary spacecraft

We propose that the mission include a small Ancillary Spacecraft (AS/C) serving the following functions:

Telecommunication: The high resolution mapping of the full sky with the many detectors of PRISM with a lossless compression of 4 gives a total data rate of $\sim 350 \text{ Mbit/s}$ (of which 300 Mbit/s is from the channels above 700 GHz). Further on-board reduction by a factor $\sim 10-20$ can be achieved by averaging the timelines of detectors following each other on the same scan path (after automatic removal of spikes due to cosmic rays) to yield a total data rate $< 40 \text{ Mbit/s}$ (a few times greater than Euclid or Gaia). A phased-array antenna or counter-rotated antenna on the main S/C could be envisioned. Decoupling the communication function from the main spacecraft using an AS/C as an intermediate station for data transmission will allow for a maximally flexible scanning strategy for the best polarization modulation and full sky coverage.

In-flight calibration: The hardest PRISM design problem is ensuring that the performance is limited by detector noise rather than systematic effects and calibration uncertainties. While pre-flight calibration is necessary, an AS/C fitted with calibrated, polarized sources could be used for precise in-flight calibration of the polarization response and polarization angles of the detectors, and for main beams and far sidelobe measurements down to extremely low levels (below -140 dB) at several times during the mission lifetime.

10 Competition and complementarity with other observations

B-mode experiments: Searching for primordial gravitational waves through B-mode polarization is the principal science driver of numerous suborbital experiments (e.g., BICEP, QUIET, SPIDER, ACTPol, SPTPol, QUBIC, EBEX, PolarBear, QUIJOTE) despite considerable limitations due to atmospheric opacity, far-side lobe pickup from the ground, and unstable observing conditions that make controlling systemic errors especially difficult, particularly on the largest angular scales where the B mode signal is largest. Forecasts

of r from ground-based experiments are often impressive but assume very simple foregrounds. For this reason a detection of r from the ground would provide a strong motivation for a confirmation and more precise characterization from space. Moreover, two US space missions concepts, *CMBPol* and *PIXIE*, and one in Japan, *LiteBird*, have been proposed, but none has yet been funded. Among the current space mission concepts, PRISM is the most ambitious and encompasses the broadest science case. *LiteBird* is a highly-targeted, low-cost Japanese B-mode mission concept, in many respects similar to the *BPol* mission proposed to ESA in 2007. With its coarse angular resolution and limited sensitivity, *LiteBird* would be able to detect B-modes assuming that r is not too small and that the foregrounds are not too complicated. *LiteBird*, however, lacks the angular resolution needed to make significant contributions to other key science objectives. The US EPIC-CS mission is the most similar to the present proposal but has considerably less frequency coverage, fewer frequency bands, and no absolute spectral capability. The US mission concept *PIXIE* proposes an improved version of the *FIRAS* spectrometer to measure B-modes and perform absolute spectroscopy simultaneously, but with an effective resolution of only 2.6° .

Cluster observations: When PRISM flies, the *eROSITA* X-ray survey will likely be the only deeper all-sky cluster survey available. 20–30 times more sensitive than ROSAT, *eROSITA*’s principal goal is to explore cosmological models using galaxy clusters. Forecasts predict that *eROSITA* will detect $\sim 10^5$ clusters at more than 100 photon counts, which is sufficient to provide a good detection and in many cases to detect the source as extended in X-rays. The main survey provides a good sample of galaxy clusters typically out to $z = 1$ with some very massive and exceptional clusters at larger distance.

The large majority of these clusters will be re-detected by PRISM and thus provide an invaluable inter-calibration of X-ray and SZE cluster cosmology, provide determinations of cluster temperatures by combining the two detection techniques, and obtain independent cluster distances for many thousands of clusters whose X-ray temperatures and shape parameters can be obtained from the X-ray survey. With $\sim 10^6$ clusters detected with PRISM, one can further exploit the *eROSITA* survey data by stacking in a way similar to the analysis of the X-ray signals from the ROSAT All-Sky Survey for SDSS detected clusters (Rykoff et al. 2008).

Other sub-millimeter/far-infrared initiatives: Existing (APEX, ASTE, IRAM 30m, LMT) and future (CCAT) ground-based single-dish submillimeter observatories are not as sensitive above 300 GHz as PRISM, mainly because of the limitations of observing through the atmosphere. Interferometers (ALMA, CARMA, PdB Interferometer, SMA) are ill-suited to observing large fields. Moreover most interferometers are insensitive to large-scale structure. SKA will span the radio range from 0.07 GHz up to 20 GHz, and will be the perfect complement to PRISM, with more than $10^9 (f_{sky}/0.5)$ HI galaxies in a redshift range $0 < z < 1.5$, and maps of the epoch of reionisation above $z \sim 6$.

PRISM will map the full-sky, large-scale continuum emission at higher sensitivities than ground based single-dish telescopes. Bright compact sources found by PRISM in its all-sky surveys can subsequently be observed in more detail by interferometers. Observations can be combined to produce superior maps of selected sky regions. The Atacama Large Millimeter Array (ALMA), operating in the range 30–1000 GHz, will complement PRISM with follow-up of sources and clusters detected by the PRISM, mapping their structure in total intensity, polarization and spectral line at high angular and spectral resolution.

CCAT will initially have two imaging instruments. At low frequencies, LWCam on CCAT will be able to detect sources below the PRISM confusion limit relatively quickly (see Golwala et al. 2013). However, variations in atmospheric transmissivity and thermal radiation from the atmosphere make it difficult for CCAT to map large scale structures. At high frequencies, SWCam will have difficulty mapping large areas to the confusion limit of PRISM. Based on the specifications from Stacey et al. (2013), CCAT can map an area of 1 square degree at 857 GHz to a sensitivity of 6 mJy (the PRISM confusion limit) within 1 h. To map the entire southern sky to this same depth requires ~ 900 days (24 h) with optimal observing conditions. Such large scale observations will not be feasible with CCAT. PRISM is needed to produce all-sky maps in these frequency bands. PRISM will produce maps at the same resolution as Herschel. However Herschel was only able to map a limited portion of the sky.

Few previous infrared telescopes have performed all-sky surveys in the bands covered by PRISM. Akari was the last telescope to perform such observations, but poor sensitivity and resolution along with limited data access have hindered the usefulness of the telescope’s data. Several other prior telescopes (Spitzer, Herschel) as well as the airborne observatory SOFIA and the future mission SPICA have observed or will observe in the 600–4000 GHz range, but only very limited areas of the sky. Furthermore, except for a few deep fields, they observe objects already identified in other bands. PRISM will be able to perform observations with sensitivities comparable to Herschel or better, but covering the entire sky in many frequency bands.

References

- [1] A. Avgoustidis, E.J. Copeland, A. Moss, et al. *Phys.Rev.Lett.*, 107:121301, 2011.
- [2] R. Barkana and A. Loeb. *Physics Reports*, 349:125–238, July 2001.
- [3] J. D. Barrow and P. Coles. *MNRAS*, 248:52–57, January 1991.
- [4] J. G. Bartlett and J. Silk. *ApJ*, 353:399–405, April 1990.
- [5] K. Basu, C. Hernández-Monteagudo, and R. A. Sunyaev. *A&A*, 416:447–466, March 2004.
- [6] Daniel Baumann et al. *AIP Conf.Proc.*, 1141:10–120, 2009.
- [7] C. L. Bennett, M. Halpern, G. Hinshaw, et al. *ApJS*, 148:1–27, September 2003.
- [8] M. Birkinshaw. *Phys. Rep.*, 310:97–195, March 1999.
- [9] F.R. Bouchet et al. 2011.
- [10] P. Bull and M. Kamionkowski. *ArXiv:1302.1617*, February 2013.
- [11] C. Burigana. In G. L. Chincarini, A. Iovino, T. Maccauro, & D. Maccagni, editor, *Observational Cosmology*, volume 51 of *Astronomical Society of the Pacific Conference Series*, pages 554–+, January 1993.
- [12] C. Burigana, L. Danese, and G. de Zotti. *A&A*, 246:49–58, June 1991.
- [13] B. J. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama. *Phys.Rev.D*, 81(10):104019–+, May 2010.
- [14] R. Cen and J. P. Ostriker. *ApJ*, 514:1–6, March 1999.
- [15] A. Challinor and A. Lasenby. *ApJ*, 499:1–+, May 1998.
- [16] J. Chluba. *Spectral Distortions of the Cosmic Microwave Background*. PhD thesis, LMU München, March 2005.
- [17] J. Chluba. *MNRAS*, 402:1195–1207, February 2010.
- [18] J. Chluba. *ArXiv:1304.6120*, April 2013.
- [19] J. Chluba. *ArXiv e-prints*, April 2013.
- [20] J. Chluba and D. Grin. *ArXiv:1304.4596*, April 2013.
- [21] J. Chluba and R. A. Sunyaev. *A&A*, 458:L29–L32, November 2006.
- [22] J. Chluba and R. A. Sunyaev. *MNRAS*, 419:1294–1314, January 2012.
- [23] J. Chluba, A. L. Erickcek, and I. Ben-Dayan. *ApJ*, 758:76, October 2012.
- [24] J. Chluba, R. Khatri, and R. A. Sunyaev. *MNRAS*, 425:1129–1169, September 2012.
- [25] Jens Chluba, Adrienne L. Erickcek, and Ido Ben-Dayan. *Astrophys.J.*, 758:76, 2012.
- [26] S. Chongchitnan and G. Efstathiou. *Phys.Rev.D*, 72(8):083520, October 2005.
- [27] R. M. Crutcher. *ARA&A*, 50:29–63, September 2012.
- [28] R. A. Daly. *ApJ*, 371:14–28, April 1991.
- [29] L. Danese and C. Burigana. In J. L. Sanz, E. Martinez-Gonzalez, and L. Cayon, editors, *Present and Future of the Cosmic Microwave Background*, volume 429 of *Lecture Notes in Physics*, Berlin Springer Verlag, page 28, 1994.
- [30] L. Danese and G. de Zotti. *Nuovo Cimento Rivista Serie*, 7:277–362, September 1977.
- [31] J. B. Dent, D. A. Easson, and H. Tashiro. *Phys.Rev.D*, 86(2):023514, July 2012.
- [32] B. T. Draine. In T. Henning, E. Grün, and J. Steinacker, editors, *Cosmic Dust - Near and Far*, volume 414 of *Astronomical Society of the Pacific Conference Series*, page 453, December 2009.
- [33] B. T. Draine and B. Hensley. *ApJ*, 765:159, March 2013.
- [34] V. K. Dubrovich. *Soviet Astronomy Letters*, 1:196–+, October 1975.
- [35] J. L. Feng. *ARA&A*, 48:495–545, September 2010.
- [36] J. L. Feng, A. Rajaraman, and F. Takayama. *Phys.Rev.D*, 68(6):063504, September 2003.
- [37] D. J. Fixsen, E. S. Cheng, J. M. Gales, et al. *ApJ*, 473:576–+, December 1996.
- [38] J. Ganc and E. Komatsu. *Phys.Rev.D*, 86(2):023518, July 2012.
- [39] Jonathan Ganc and Eiichiro Komatsu. *Phys.Rev.*, D86:023518, 2012.
- [40] P. Goldreich and N. D. Kylafis. *ApJL*, 243:L75–L78, January 1981.
- [41] Y. Gong, A. Cooray, M. Silva, et al. *ApJ*, 745:49, January 2012.
- [42] et al. Griffin, M. J. *A&A*, 518:L3, July 2010.
- [43] A. C. Hall and A. Challinor. *MNRAS*, 425:1170–1184, September 2012.
- [44] P. Hennebelle and E. Falgarone. *A&A Rev.*, 20:55, November 2012.
- [45] C. Hernández-Monteagudo and R. A. Sunyaev. *MNRAS*, 359:597–606, May 2005.
- [46] C. Hernández-Monteagudo, L. Verde, and R. Jimenez. *ApJ*, 653:1–10, December 2006.
- [47] C. Hernández-Monteagudo, Z. Haiman, R. Jimenez, and L. Verde. *ApJL*, 660:L85–L88, May 2007.
- [48] C. Hernández-Monteagudo, J. A. Rubiño-Martín, and R. A. Sunyaev. *MNRAS*, 380:1656–1668, October 2007.
- [49] W. Hu and J. Silk. *Phys.Rev.D*, 48:485–502, July 1993.
- [50] W. Hu and J. Silk. *Physical Review Letters*, 70:2661–2664, May 1993.
- [51] W. Hu and N. Sugiyama. *ApJ*, 436:456–466, December 1994.
- [52] W. Hu, D. Scott, and J. Silk. *ApJL*, 430:L5–L8, July 1994.
- [53] W. Hu, D. Scott, and J. Silk. *Phys.Rev.D*, 49:648–670, January 1994.
- [54] A. F. Illarionov and R. A. Sunyaev. *Astronomicheskii Zhurnal*, 51:1162–1176, December 1974.
- [55] N. Itoh, Y. Kohyama, and S. Nozawa. *ApJ*, 502:7, July 1998.
- [56] R. J. Ivison, A. M. Swinbank, I. Smail, et al. *ArXiv e-prints*, February 2013.
- [57] K. Jedamzik, V. Katalinić, and A. V. Olinto. *Physical Review Letters*, 85:700–703, July 2000.
- [58] R. Khatri and R. A. Sunyaev. *JCAP*, 6:038, June 2012.
- [59] R. Khatri and R. A. Sunyaev. *JCAP*, 9:016, September 2012.
- [60] R. Khatri and R. A. Sunyaev. *ArXiv:1303.7212*, March 2013.
- [61] R. Khatri, R. A. Sunyaev, and J. Chluba. *A&A*, 540:A124, April 2012.
- [62] R. Khatri, R. A. Sunyaev, and J. Chluba. *A&A*, 543:A136, July 2012.
- [63] W. H. Kinney. *Phys.Rev.D*, 66(8):083508, October 2002.

- [64] L. Kuzmin, G. Yassin, S. Withington, and P. Grimes. In A. Karpov, editor, *Eighteenth International Symposium on Space Terahertz Technology*, page 93, 2007.
- [65] A. Lapi, J. González-Nuevo, L. Fan, et al. *ApJ*, 742: 24, November 2011.
- [66] A. Lewis, J. Weller, and R. Battye. *MNRAS*, 373: 561–570, December 2006.
- [67] K. Lochan, S. Das, and A. Bassi. *Phys.Rev.D*, 86(6): 065016, September 2012.
- [68] J. C. Mather, E. S. Cheng, D. A. Cottingham, et al. *ApJ*, 420:439–444, January 1994.
- [69] P. McDonald, R. J. Scherrer, and T. P. Walker. *Phys.Rev.D*, 63(2):023001–+, January 2001.
- [70] J.-B. Melin, J. G. Bartlett, and J. Delabrouille. *A&A*, 459:341–352, November 2006.
- [71] F. Miniati, D. Ryu, H. Kang, et al. *ApJ*, 542:608–621, October 2000.
- [72] M. Negrello, R. Hopwood, G. De Zotti, et al. *Science*, 330:800–, November 2010.
- [73] S. P. Oh, A. Cooray, and M. Kamionkowski. *MNRAS*, 342:L20–L24, June 2003.
- [74] J. P. Ostriker and C. Thompson. *ApJL*, 323:L97–L101, December 1987.
- [75] E. Pajer and M. Zaldarriaga. *Physical Review Letters*, 109(2):021302, July 2012.
- [76] Enrico Pajer and Matias Zaldarriaga. *Phys.Rev.Lett.*, 109:021302, 2012.
- [77] P. J. E. Peebles. *ApJ*, 153:1–+, July 1968.
- [78] H. V. Peiris, E. Komatsu, L. Verde, et al. *ApJS*, 148: 213–231, September 2003.
- [79] J. L. Pineda, W. D. Langer, T. Velusamy, and P. F. Goldsmith. *ArXiv e-prints*, April 2013.
- [80] Planck Collaboration, N. Aghanim, M. Arnaud, et al. *A&A*, 536:A12, December 2011.
- [81] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv e-prints*, December 2012.
- [82] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv e-prints*, March 2013.
- [83] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv e-prints*, March 2013.
- [84] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv:1303.5084*, March 2013.
- [85] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv:1303.5076*, March 2013.
- [86] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. *ArXiv e-prints*, March 2013.
- [87] Planck Collaboration XXII. *Submitted to A&A*, 2013.
- [88] P. P. Ponente, J. M. Diego, R. K. Sheth, et al. *MNRAS*, 410:2353–2362, February 2011.
- [89] B. A. Powell. *ArXiv:1209.2024*, September 2012.
- [90] Y. Rephaeli. *ApJ*, 445:33–36, May 1995.
- [91] Y. Rephaeli and O. Lahav. *ApJ*, 372:21–24, May 1991.
- [92] D. A. Riechers, C. M. Bradford, D. L. Clements, et al. *Nature*, 496:329–333, April 2013.
- [93] J. A. Rubiño-Martín, C. Hernández-Monteagudo, and R. A. Sunyaev. *A&A*, 438:461–473, August 2005.
- [94] J. A. Rubiño-Martín, J. Chluba, and R. A. Sunyaev. *MNRAS*, 371:1939–1952, (RMCS06), October 2006.
- [95] J. A. Rubiño-Martín, J. Chluba, and R. A. Sunyaev. *A&A*, 485:377–393, July 2008.
- [96] S. Y. Sazonov and R. A. Sunyaev. *ApJ*, 508:1–5, November 1998.
- [97] U. Seljak and M. Zaldarriaga. *ApJ*, 538:57–64, July 2000.
- [98] J. Smidt, A. Amblard, C. T. Byrnes, et al. *Phys.Rev.D*, 81(12):123007, June 2010.
- [99] K.M. Smith, A. Cooray, S. Das, et al. *AIP Conference Proceedings*, 1141:121, 2009.
- [100] R. A. Sunyaev and J. Chluba. *Astronomische Nachrichten*, 330:657–+, 2009.
- [101] R. A. Sunyaev and Y. B. Zeldovich. *Astrophysics and Space Science*, 9:368–382, December 1970.
- [102] R. A. Sunyaev and Y. B. Zeldovich. *A&A*, 20:189–+, August 1972.
- [103] R. A. Sunyaev and Y. B. Zeldovich. *Comments on Astrophysics and Space Physics*, 4:173, November 1972.
- [104] H. Tashiro, E. Sabancilar, and T. Vachaspati. *Phys.Rev.D*, 85(10):103522, May 2012.
- [105] H. Tashiro, E. Sabancilar, and T. Vachaspati. *ArXiv:1212.3283*, December 2012.
- [106] Planck HFI Core Team. *A&A*, 536:A4, December 2011.
- [107] T. Treu. *ARA&A*, 48:87–125, September 2010.
- [108] Licia Verde, Hiranya Peiris, and Raul Jimenez. *JCAP*, 0601:019, 2006.
- [109] Y. B. Zeldovich and R. A. Sunyaev. *Astrophysics and Space Science*, 4:301–316, July 1969.
- [110] Y. B. Zeldovich, V. G. Kurt, and R. A. Syunyaev. *Zhurnal Eksperimental noi i Teoreticheskoi Fiziki*, 55:278–286, July 1968.
- [111] P. Zhang, U.-L. Pen, and H. Trac. *MNRAS*, 355: 451–460, December 2004.



Sub-arcsecond far-infrared space observatory: a science imperative

M. Sauvage (Spokesperson)*

May 24, 2013

*CEA/DSM/Irfu/SAP CE Saclay
91191 Gif-sur-Yvette CEDEX, France,
marc.sauvage@cea.fr,
+33 1 69 08 62 99

A White Paper presented by

R. J. Ivison (UK/STFC-UKATC), F. Helmich (NL/SRON & RUG) & M. Sauvage (FR/CEA)

With contributions from

Afonso, J. (PT/CAAUL)	André, P. (FR/CEA)	Aussel, H. (FR/CEA)
Baes, M. (BE/UGent)	Barlow, M. (UK/UCL)	Baryshev, A. (NL/SRON)
Bendo, G. (UK/JBCA)	Berné, O. (FR/IRAP)	Boulanger, F. (FR/IAS)
Braine, J. (FR/LAB)	Cava, A. (ES/UCM)	Ceccarelli, C. (FR/IPAG)
Contursi, A. (DE/MPE)	de Graauw, T. (ALMA/Santiago)	Durand, G. (FR/CEA)
Galametz, M. (UK/IoA)	Grin, M. (FR/LERMA)	Goicoechea, J. (ES/CAB)
Gomez, H. (UK/Cardiff)	Greaves, J. (UK/SUPA-StAndrews)	Grenier, I. (FR/CEA)
Griffin, M. (UK/Cardiff)	Gruppioni, C. (IT/INAF-OAB)	Guillard, P. (FR/IAS)
Habart, E. (FR/IAS)	Herpin, F. (FR/LAB)	Kerschbaum, F. (AT/UNIVIE)
Leisawitz, D. (US/NASA)	Levrier, F. (FR/ENS)	Lis, D. (US/Caltech)
Madden, S. (FR/CEA)	Maiolino, R. (UK/UCambridge)	Mather, J. (US/NASA)
Minier, V. (FR/CEA)	Molinari, S. (IT/IAPS)	Motte, F. (FR/CEA)
Mullaney, J. (UK/Durham)	Nesvadba, N. (FR/IAS)	Pantin, E. (FR/CEA)
Peretto, N. (UK/Cardiff)	Reveret, V. (FR/CEA)	Rigopoulou, D. (GB/RAL-Oxford)
Rodriguez, L. (FR/CEA)	Savini, G. (UK/UCL)	Schilke, P. (DE/UKoeln)
Serjeant, S. (UK/Open University)	Smail, I. (UK/Durham)	Smith, M. (UK/CU)
Spencer, L. (UK/Cardiff)	Spinoglio, L. (IT/IAPS)	Tremblin, P. (UK/UExeter)
Vaccari, M. (SA/UWC)	van der Tak, F. (NL/SRON-RUG)	van Kampen, E. (ESO/Garching)
Vastel, C. (FR/IRAP)	Viti, S. (UK/UCL)	Walter, F. (DE/MPIA)
Wilson, C. (CA/McMaster)	Wright, E. (US/UCLA)	Wyatt, M. (UK/UCambridge)
Zavagno, A. (FR/LAM)		

And support from

Abergel, A. (FR/IAS)	Audard, M. (CH/UNIGE)	Avison, A. (UK/UManchester)
Benedettini, M. (IT/INAF)	Bernard, J.-P. (FR/IRAP)	Bontemps, S. (FR/LAB)
Cabrit, S. (FR/LERMA)	Caputi, K. (NL/RUG)	Caux, E. (FR/IRAP)
Cavalié, T. (FR/LAB)	Chanial, P. (FR/APC)	Charlot, S. (FR/IAP)
Cimatti, A. (IT/UNIBO-DIFA)	Clements, D. (UK/IC)	Conselice, C. (UK/UNottingham)
Davies, J. (UK/Cardiff)	De Zotti, G. (IT/INAF-SISSA)	Decin, L. (BE/KUL)
Demyk, K. (FR/IRAP)	Doublier-Pritchard, V. (DE/MPE)	Duband, L. (FR/CEA)
Dunlop, J. (UK/ROE)	Elbaz, D. (FR/CEA)	Farrah, D. (US/VirginaTech)
Fich, M. (CA/UWaterloo)	Fischer, J. (US/NRL)	Fletcher, L. (UK/Oxford-AOPP)
Fraser, H. (UK/Open University)	Fuller, G. (UK/JBCA)	Giard, M. (FR/IRAP)
Glenn, J. (US/CASA)	Goldsmith, P. (US/JPL)	Hennemann, M. (DE/MPIA)
Henning, T. (DE/MPIA)	Hill, T. (ALMA/Santiago)	Irwin, P. (UK/UOxford)
Kirk, J. (UK/UCLancas.)	Krause, O. (DE/MPIA)	Laporte, N. (ES/IAS)
Lefloch, B. (FR/IPAG)	Lester, D. (US/UTexas-Dept. Astr.)	Lim, T. (UK/RAL)
Marchetti, L. (UK/Open University)	Marconi, A. (IT/UNIFI)	Meixner, M. (US/STScI)
Murphy, A. (IE/NUI Maynooth)	Naylor, D. (CA/ISIS-ULeth.)	O'Sullivan, C. (IE/NUI Maynooth)
Oliver, S. (UK/Sussex)	Olofsson, G. (SE/UStockholm)	Orton, G. (US/JPL)
Pagani, L. (FR/LERMA)	Pajot, F. (FR/IAS)	Paradis, D. (FR/IRAP)
Perault, M. (FR/LERMA)	Plume, R. (CA/UCalgary)	Poglitsch, A. (DE/MPE)
Pointecouteau, E. (FR/IRAP)	Pozzi, F. (IT/UNIBO)	Rinehart, S. (US/NASA)
Ristorcelli, I. (FR/IRAP)	Ronayette, S. (FR/CEA)	Schaerer, D. (CH/Geneva - FR/CNRS)
Smith, H. (US/CfA)	Starck, J.-L. (FR/CEA)	Tosi, F. (IT/IAPS)
Traficante, A. (UK/UManchester)	Trappe, N. (IE/NUI Maynooth)	Unwin, S. (US/JPL)
Vakili, F. (FR/OCA)	van Langevelde, H. (JIVE/Leiden)	Vandenbussche, B. (BE/KUL)
Ward-Thompson, D. (UK/UCLancas.)	Wilkins, S. (UK/Sussex)	Yorke, H. (US/JPL)

Cover page: The Rosette Nebula, dust seen by *Herschel* in red, H α in blue, and an artist rendering of the propagating UV light. V. Minier(CEA)/Novae Factory - Schneider et al. 2010 - *Herschel*/HOBYS KP.

Executive summary

The far-infrared (FIR) part of the electromagnetic spectrum (25–500 μm) contains around half of the post-big bang energy and nearly all of the photons that come to us from all astrophysical processes. Yet access to this waveband remains challenging for basic reasons: the atmosphere is largely opaque, telescopes must be large and cold, and detector technology has come of age only recently.

Science exploiting the FIR domain is thus relatively young, yet already demonstrates an impressive track record: a succession of facilities – *KAO*, *IRAS*, *COBE*, *ISO*, *Spitzer*, *AKARI*, *Planck* and *Herschel* – allowed us to gaze into the obscured Universe, advancing our understanding of cosmology, star and galaxy formation, and the origin of planetary systems. Despite these developments, FIR observational capabilities remain primitive in comparison with the optical/NIR region. Our most advanced facility, *Herschel*, delivered an angular resolution no better than Galileo’s telescope and was operated against a blinding thermal background.

FIR space facilities have been limited in aperture by the need for cryogenic cooling, which restricted the telescope size to <1 m for *IRAS*, *ISO*, *AKARI* and *Spitzer*. *Herschel* had the largest possible telescope, given technical and budget constraints, but this aperture could only be passively cooled, to ~ 85 K, imposing a fundamental limit to the achievable sensitivity. Despite this, much of its success is attributable to the enhanced angular resolution provided by its 3.5-m aperture.

Indeed, *Herschel* has revolutionised our understanding of the Universe. Observations of the distant Universe have entirely revised the accepted wisdom on high-redshift star formation, from a picture dominated by major galaxy mergers to one where star formation proceeds via a secular mode, albeit inside much larger galaxies than present-day ones. Closer to us, *Herschel* revealed that the initial mass function (IMF, describing the probability of forming a star of a given mass) takes root in the very early stages of structure formation in the interstellar medium (ISM), where turbulence and magnetic fields likely play key roles. For the first time we are also gaining systematic access to a broad range of molecular tracers that reveal the nature of many complex energy exchanges in the ISM, and the richness of the chemistry involved in proto-stellar evolution.

The scientific productivity of *Herschel* has demonstrated the importance of the FIR region to current and future research, and highlights all the more painfully the “FIR gap” – the dramatically poorer sensitivity and an-

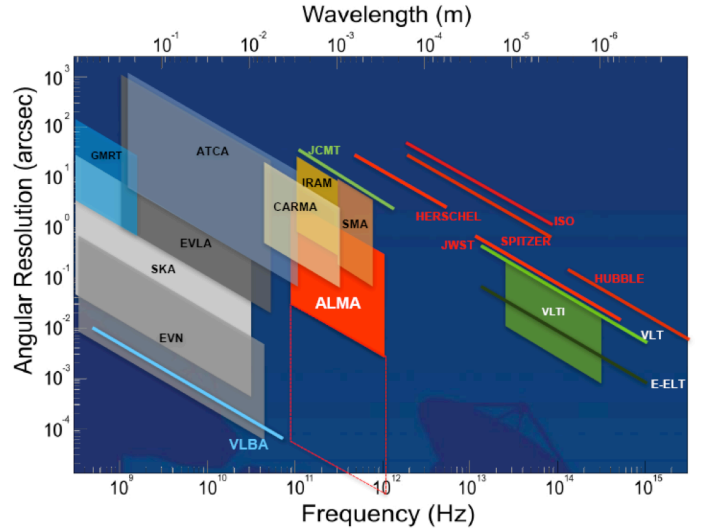


Figure 1: Angular resolutions reached by current or planned facilities, illustrating the growing “FIR gap” (courtesy of Th. de Graaw).

gular resolution of FIR facilities as compared with observatories operating at both longer and shorter wavelengths (e.g. *ALMA*, *JWST*, see Fig. 1). Providing our only access to information critical to our understanding of how galaxies, stars and planets form and evolve, eliminating this FIR gap must be a key objective.

A major leap in FIR capability is promised in the 2020s by the Japanese-led *SPICA* mission, with its cooled 3-m telescope. Zodiacal light and source confusion will then be the limit to sensitivity, and *SAFARI*, operating at 30–200 μm , will provide two orders of magnitude better spectral sensitivity than *Herschel*/PACS, and will map several thousand times faster than *Spitzer* or *Herschel*, enabling new astrophysics.

The FIR sensitivity gap is thus being addressed, but none of the approved facilities will provide any advance in angular resolution, which is an immutable requirement to tackle the fundamental scientific questions highlighted in this document. The scientific imperative for FIR astronomy is clearly to achieve a combination of high *sensitivity* and high *angular resolution*. This is evident from a simple consideration of the physical scales involved in the science areas where *Herschel* has led to significant progress; 100 AU for disks, 0.1 pc for structure in star-forming regions, 1 kpc for large star-forming complexes at cosmological distances. In order to explore these scales, and to fully exploit the synergies with contemporary facilities like *JWST*, *SKA* and *ALMA*, with statistically meaningful samples, one must

reach angular resolutions of 0.1–1" at 100 μm , i.e. 1–2 orders of magnitude better than *SPICA* will deliver.

Furthermore, the FIR domain is one where the very questions raised by ESA's Cosmic Vision can be tackled, namely "*What are the conditions for planet formation and the emergence of life? How did the Universe originate and what is it made of?*" This is first due to the fact that thermal emission from dust associated with star formation peaks in the FIR, but more importantly to the presence of emission signatures that are both unique to this part of the spectral domain, and also unique in the access that they provide to the underlying physics. Via dust features in the continuum emission, we access the dust composition, size distribution and formation scenario, about which we know precious little; through ice features we can study critical evolutionary processes in the dense phases of the ISM; fine-structure lines found in the FIR give access to the thermal balance in the regions observed; for redshifted sources, we access the mid-IR range, where we find unambiguous diagnostics for the presence of active galactic nuclei; finally, the FIR offers a vast array of molecular tracers, among which the simple hydrides that represent a unique diagnostic for the build-up of molecular complexity, and water, whose transitions are key probes of the star-formation process and in assessing habitability for exo-planetary systems.

Together these physical tracers allow the construction of an extremely consistent picture of the objects observed, in a way that is quite unique throughout the electromagnetic spectrum. A FIR facility reaching sub-arcsecond resolution promises formidable breakthroughs in our understanding of such fundamental questions as:

- How do the conditions for planet habitability arise during star, disk and planet formation?
- How do stars of all mass evolve from interstellar clumps to stellar and planetary systems?
- How and why does the FIR/submm spectral energy distribution of galaxies evolve over cosmological timescales?
- What is the interplay between massive black holes in galactic nuclei and star formation in their host galaxies?
- What is the nature of the FIR background and of early, deeply embedded star formation?
- How were the first luminous objects in the Universe ignited? How did the first stars and the first black holes form and evolve?

Indeed, such a facility will make unique and key contributions to our understanding of the Universe. It will

peer through the dust that shrouds stellar nurseries, demystifying the process by which stars and planets are born. It will image proto-stellar and debris disks at the peak of their spectral energy distributions, where the brightness is 1000 \times that at 1 mm, revealing how proto-planetary disks form out of gas, dust and ice. It will break through the confusion limit to determine the properties and structure of distant star-forming galaxies, and to examine the enigmatic symbiosis between host galaxies and their AGN. We may even be able to detect the formation of the earliest stars and galaxies, via redshifted emission from primordial metal-free H_2 , while SKA explores the ensuing re-ionisation of the Universe on a similar timeframe as the L2/L3 missions.

1 FIR science in the L2/L3 era

1.1 Discs - birthplaces of planets

Stars accrete material through discs, created when the massive proto-stellar envelopes collapse. These discs are relatively large (several tens to hundreds AU) and their chemistry is dominated by freeze out and low temperature ion-molecule reactions. Outflows and UV-radiation clear out most of the envelope leaving a small, flaring proto-planetary disc, with an inner gap between disc and star. The mid-planes of these discs are cold and gas is frozen out on dust grains. The outer disc (from a few to tens of AU) surface is irradiated by UV from the young star and shows mostly ions and radicals. Finally, close to the inner rim (less than a few AU) temperatures are high and X-rays heats dust and gas that excites H_2 , fine-structure lines, H_2O , PAHs etc.

Planets form through coagulation of dust and gas accretion in the discs that develop during the collapse and infall of massive proto-stellar envelopes. Although initial accretion rates are high, there is a point when more modest accretion rates allow structures leading to planets to be sustained. The formation of planetary systems is thus an important by-product of the star formation process.

In recent years, remarkable progress has been made on the study of all phases of circumstellar, proto-planetary and debris discs using *Herschel* and *ALMA*. Examples include images of the disc around Fomalhaut by *Herschel* (Acke et al. 2012) and of gas and dust in HD 142527 by *ALMA* (Casassus et al. 2013).

Herschel provided our first samples for which the FIR properties in gas and dust have been probed and *SPICA* will increase our statistics by orders of magni-

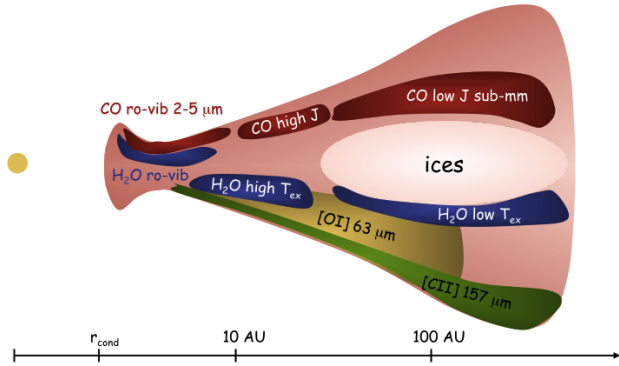


Figure 2: A schematic cut across a proto-planetary disks, highlighting the location of the different tracers of its structure and composition (I. Kamp, priv. comm.).

tude, but neither will resolve the relevant spatial scales in circumstellar discs ($\ll 100$ AU) even at the close proximity of TW Hydrae (at ~ 50 pc with a disc several hundreds of AU in size). This limits our knowledge of the fundamental physical and chemical characteristics of discs, how dusty bodies grow and collide and how the gas is finally cleared out, by winds, outflows and accretion on young planets. Also, we still have a very incomplete understanding of the formation mechanisms responsible for the great diversity of planetary systems detected so far, including our own Solar System.

To progress in this field, direct access to the the basic building blocks of protostars, discs and planets is needed, i.e. the gas, dust and ice. These radiate predominantly in the FIR, where a unique collection of mineral, ice and gas spectral diagnostics that cannot be observed from the ground exist (including water), and that can be complemented by ground-based facilities like *ELT* and *ALMA* (see Fig. 2). Therefore, it is only by observing in the FIR band at high angular resolution that we will be able to unveil fundamental processes that transform the interstellar material into planetary systems, and thus to provide answers to some of the most basic questions about our place in the Universe: Are planets like our own common in the Milky Way and, if so, what implications does this have for the appearance of exo-planets that might sustain life?

The gas mass in planet forming discs, protoplanets in habitable zones. Protoplanetary discs evolve over a timescale of a few million years (e.g. Haisch et al. 2006). This is the critical intermediate stage when plan-

etary formation is believed to take place. Although the dust is relatively easily detected by photometric observations in the FIR range, very little is known about their gas content. It is evident though that too little gas is left at ages > 10 Myr (e.g. Meyer et al. 2008; Dent 2013) to form giant planets. The very fact that these planets are largely gaseous means they must be formed before the gaseous disc dissipated, making the study of the gas in discs essential to understand how and where they formed. We have two possible formation mechanisms: gas giant planets form either via accretion of gas onto rocky/icy cores of a few earth masses (Lissauer 1993; Kornet et al. 2002) or by gravitational instability in the disc that triggers the formation of overdense clumps that afterwards compress to form giant planets (Boss 2003). In the latter scenario, gas giants form quickly and the gas may dissipate early (< 10 Myr), but see also Chiang & Laughlin (2013). The residual gas content in the innermost disc regions at the time rocky protoplanets assemble can also determine their final mass, chemical content and orbit eccentricity, and therefore their possible habitability (Agnor & Ward 2002).

Detailed observations of the radial distribution and amount of gas in a sample of planet forming discs can discriminate between different planet forming theories, or whether these processes all occur within one disc. However accurately constraining the disc gas mass is complex: *ALMA* can observe the optically thin dust emission in the submm and the low-lying lines of carbon monoxide (CO) at high angular resolution. However, those CO lines are optically thick and thus generally trace the disc surface temperature rather than the total mass, whereas the dust-to-total mass conversion factor requires knowledge of the poorly known grain temperature and optical properties (which only be obtained from FIR and MIR observations). The optically thin CO-isotopologues suffer from freeze-out in the mid-plane, although there are indications of competition between freeze-out and vertical mixing. Ground-based interferometers cannot observe either the brightest gas line in protoplanetary discs ([OI] $63 \mu\text{m}$) nor the thermal emission of water vapour (H_2^{16}O) or the emission from the most abundant hydrides in the disc (H_2 , HD, OH...), leaving little room to test differences between accretion and instabilities.

HD (with the lowest energy lines at 112, 56 and $37 \mu\text{m}$ i.e., not observable with *JWST* or *ALMA*) turns out to be the most powerful tracer of the disc gas because it follows the distribution of H_2 and its emission is sensitive to the total mass. Indeed, although HD is

$\sim 10^5$ times less abundant, the FIR lines of HD are many times more emissive than those of H_2 .

The recent first detection of HD ($J=1-0$ at $112\mu\text{m}$) towards the TW Hydrae disc with *Herschel* (total line flux: $6 \cdot 10^{-18} \text{ Wm}^{-2}$; Bergin et al. 2013) implies a disc mass of more than $0.05 M_\odot$, i.e. enough to form a planetary system like the Solar System despite the advanced age of TW Hya ($\sim 10 \text{ Myr}$). There is, however, debate on the CO/HD ratio necessitating more observations including the HD 2-1 line. Unfortunately, even towards the closest protoplanetary disc, the HD line emission is spatially unresolved. With an angular resolution of $\sim 0.5''$ at $56\mu\text{m}$ (HD $J=2-1$), spatially resolved images of the disc mass-distribution can be made, which will bring down errors on the mass determination by orders of magnitude. We will thus replace our current disc-integrated description of the TW Hydrae disc by a spatially resolved one, and we will be able to separate, in the nearest star-forming regions, the outer disc from the inner disc zone, where the majority of planets assemble ($\leq 100 - 200 \text{ AU}$). This will provide the most accurate determination of the gas mass distribution in a sample of planet forming discs around different types of stars and different star forming environments, a critical parameter for any global planet formation theory. Using similar images of [OI] $63, 145\mu\text{m}$, [CII] $158\mu\text{m}$ and of high excitation CO and OH lines, we will separate contributions of jets and outflow from the disc. A high angular resolution observatory will also be unique to study the disc dissipation timescales and resolve the gas pressure radial and vertical disc profiles (constraining the disc thermodynamics), thus possibly providing indirect clues to the potential presence of liquid water.

Resolving the snow line, water transport and chemical composition of primitive planets. The protoplanetary disc is the major reservoir of key species with prebiotic relevance: complex organic molecules and volatiles like ammonia and water. Complex organics are mainly *ALMA* territory, but the light hydrides can only be studied from space in the FIR waveband. Water in particular is the only solvent we know for life, and it is critical that we understand how it transfers from protostellar clouds and protoplanetary discs to more evolved asteroids, comets and planets with oceans like our own.

The handful of detections of cold water vapour towards planet forming discs with *Herschel* (due to the faintness of the spatially unresolved emission) seem to indicate that the primitive Earth was a dry planet and that water was delivered by impacts of icy bodies orig-

inating from a water reservoir in the cold outer disc at $\sim 100 \text{ AU}$ (e.g. Hogerheijde et al. 2011; Podio et al. 2013). On the other hand, warmer water vapour FIR emission, presumably from the inner rocky planet formation zones inside the snow line, has been reported towards a large fraction of the observed discs (e.g. Riviere-Marichalar et al. 2012b), demonstrating that FIR water vapour lines are excellent diagnostics of different disc zones and of their different excitation/abundance regimes. Measurement of the ortho-to-para ratio of water could connect gas and dust temperatures just as $\text{HDO}/\text{H}_2\text{O}$ will. *Herschel* was not sensitive enough and *SPICA* will not resolve emission spatially, thus it will be very difficult to break model degeneracies and fully constrain the water origin and the radial and vertical abundance distribution, which is needed to be able to routinely resolve and map the distribution of water vapour and ice (*ALMA* cannot observe the thermal emission of H_2^{16}O nor the water ice bands), parameters that serve as input for disc dynamical models.

Below $\sim 150 \text{ K}$ water vapour freezes-out onto dust grains and the main form of water in the cold disc mid-plane and at large disc radii will be ice. The physical location at which water freezes out determines the position of the snow line. In the FIR the $\sim 44\mu\text{m}$ and $\sim 62\mu\text{m}$ bands are a powerful tool for the determination of the amorphous/crystalline nature of water ice (e.g. Moore & Hudson 1994). This determination, critical to constrain the formation history of water ice, is best defined in the FIR, and unlike the MIR band of water ice at $\sim 6.1\mu\text{m}$ (accessible to *JWST*), the FIR ice bands are not confused with other features of less abundant ices, and are not influenced by the brightness of the star itself. In addition, in optically thin discs, it is extremely difficult to use MIR absorption to trace water ice and the material is too cold to emit in the NIR/MIR bands. Hence, these strong FIR features are robust probes of the presence/absence of water ice, even in cold or heavily obscured or cold regions without a MIR background, and of the amorphous/crystalline state (Malfait et al. 1999). *SPICA* will be able to observe ice-bands, including those of CO and CO_2 but will not be able to spatially resolve the snow-line except for a handful of systems and will not have sufficient spectral resolution to measure the spectral profile of the gas-lines.

Resolving the snow-line may give vital clues to the understanding of the composition of our Solar System and of exo-systems. Finding the mechanisms for water transport in discs will identify the possibilities for finding life elsewhere in our Galactic neighbourhood.

Structure and architecture of protoplanetary systems. Dust is an important ingredient for planet formation. *JWST* and *ALMA* allow complementary studies, but they have no access to the wavelength region where the dust temperature can be determined accurately; *SPICA* will measure dust at the peak of its spectral energy distribution (SED), but will in most cases not resolve the structures. Only high angular resolution imaging of the dust emission can show gaps, spiral density waves and ring-like structures indicating the presence of planets which “shepherd” the dust and can be compared to the different dust populations probed by *ALMA* and *JWST* images at **similar** spatial scales. The FIR also gives access to features such as the $\sim 69\mu\text{m}$ band of forsterite that will be used to constrain the mineral temperature, crystallinity and rock composition (see e.g. Sturm et al. 2013, for *Herschel* detections).

Discs with ages above ~ 10 Myr are practically devoid of gas and the dust in these older discs is generally not primordial but continuously generated “debris” from planetesimals and rocky body collisions. The smallest dust grains have, at this stage, either been dispersed or have coagulated into larger grains and the disc becomes very optically thin ($\tau_V \ll 1$). Numerous debris disks are known within 10 pc, and volume-limited surveys with *Herschel* are finding that many have *double* dust belts. Such debris discs are thus more massive (and usually younger) analogues of our own asteroid (hot inner disc, $T_{\text{dust}} \approx 200\text{K}$) and Kuiper belts (cool outer disc, $T_{\text{dust}} \approx 50\text{K}$), where this structure is thought to reflect the dynamical influences of a many-planet architecture. Their study at very high angular resolution is vital to place the Solar System in the broader context of other exo-planetary systems. It is here that we can link the *SPICA* measurements of hundreds of Trans Neptunian Objects and Kuiper Belt Objects to the spectral images of exo-zodiacal systems at high spatial resolution: in “Fomalhaut-like” systems (nearby debris discs hosting exoplanets), an angular resolution of $1''$ at $100\mu\text{m}$ will be equivalent to spatial resolutions down to ~ 8 AU. With high sensitivity it will be possible to provide sharp images not only of the $\sim 44\mu\text{m}$ and $\sim 62\mu\text{m}$ water ice bands, crystalline dust, but also of any “secondary” gas content (e.g., traced by [OI] and OH lines in the FIR) that could be produced by the photoevaporation of ice grain mantles, outgassing of comets or even collisional evaporation (see e.g. Riviere-Marichalar et al. 2012a).

All in all, it is clear that only through unprecedented high-angular resolution FIR multi-wavelength observations will we be able to fully understand the complete

picture of the formation and evolution of planetary disc systems by resolving the natural emission of their building blocks (gas, dust and ice). These are the observational capabilities that will ultimately constrain the initial conditions (physical and chemical), the formation timescales and the architecture of habitable planetary systems like our own.

1.2 Star formation

High mass stars control the chemical and dynamical evolution of galaxies, yet how they form remains largely unknown. They are rare and distant, so that *Herschel*’s could not resolve massive dense cores in the Galaxy, and they evolve rapidly, meaning that only with sufficiently large samples will we be able to understand their formation scenario. Contributing to blurring the picture, the initial mass function at the high end is not well constrained, essentially for lack of sufficiently resolved regions on which to study its emergence.

Recently however, progress has been made through both theoretical and observational approaches.

Hennebelle & Chabrier (2008) for instance proposed a theory to derive the Core/Initial Mass Function (CMF, IMF) from the probability distribution function (PDF) of the column density of gas in a gravo-turbulent molecular cloud. Here, the shape of the IMF (Chabrier 2003) can be related to the global properties of the gravo-turbulent cloud: a lognormal distribution (whose width is linked to turbulence) and a high-density power-law tail (from the gravitational instabilities).

Observations also favor a direct relation between the IMF and ISM structuring processes. André et al. (2010) showed that the CMF measured in the *Herschel* column density maps of nearby clouds bears a strong resemblance to the IMF. The clouds are systematically structured in filaments with a characteristic width of 0.1 pc (Arzoumanian et al. 2011, and Fig. 3), independently of their star formation activity, arguing for a universal process at work, and this width is close to the scale of turbulence dissipation. Filaments are also the only star formation site, with a threshold that appears dictated by gravitational instability along the filament (Schneider et al. 2012). Thus observations support a scenario where turbulence and gravitational instabilities play a lead role in the generation of structure and pre-stellar objects. The magnetic field is the missing player in this paradigm, despite the fact that magnetic energy is a significant fraction of the ISM energy budget, but this is because it remains very poorly constrained observations

tionally (e.g. Crutcher 2012). We will show below that we can make significant steps forward here as well.

The *Herschel* studies are however reaching a limit, in that they can only access nearby regions, heavily biased toward the low-mass end of the IMF. *SOFIA* and *SPICA* will not extend our reach, thus exploration of the high-mass end of the IMF, and of the possible impact high-mass star formation can have on the IMF, i.e. the feedback, requires a leap in resolution in the FIR, combined with access to a significant survey area.

Stellar and cloud feedback processes In high-mass star-forming regions, expanding ionized gas from the HII regions into the molecular cloud compresses the material and leads to the formation of dense layers, and clumps. *Herschel* imaging clearly revealed this impact: the PDFs of all clouds surveyed in Tremblin (2012) can be fitted with two log-normal distributions, where the lower column density distribution describes the unperturbed molecular gas while the second peak corresponds to a compression zone induced by the expansion of the ionized gas into the molecular cloud.

Thus there is clear evidence that feedback is at work in high-mass star forming regions (see for instance Fig. 3. Furthermore, considering the relative timescales between the molecular clouds lifetime (~ 10 Myr) and the HII region development time (~ 1 Myr), clouds hosting massive star formation should spend almost all their lifetime in a state where the ionized gas compresses the cold gas. Therefore any scenario that derives the initial mass function (IMF) from the cloud structure must include feedback as a structuring process as well.

Herschel observations of high-mass star-forming regions reveal additional strikingly different structural properties from low-mass counterparts. They show the dominance of very dense filaments, called “ridges” (10 times the critical density required for pre-stellar core formation) over a network of smaller filaments connected to the ridge (Hill et al. 2011; Hennemann et al. 2012; Minier et al. 2013). These sub-filaments are aligned with the ambient magnetic field according to available optical/NIR polarization vectors. This is suggestive of mass accretion along field lines into the main filaments, supporting the view that magnetic fields may at least partly control the growth of interstellar structures (see magnetic field project below). How these ridges manage to reach these densities while resisting fragmentation is not yet understood, partly because our current resolving power makes it impossible to evidence substructure in the ridges.

These observations point to a very complex inter-

play for structure formation at the high-mass end that demands deeper study. Unfortunately *Herschel* could not resolve the high-density regions in high-mass star forming clouds, due to their distance. *SPICA* will not improve on *Herschel*’s resolution, while *ALMA*, working in the Rayleigh-Jeans part of the dust emission, cannot allow the correct focus on the warmest phases of the clouds where feedback processes leave their mark.

On the contrary, with a $1''$ resolution at $100\ \mu\text{m}$, and considering the wavelength coverage needed to disentangle temperature from density in emission maps, we can resolve the pre-stellar core stage out to 5 kpc, encompassing all of the Galactic high-mass star forming regions studied by *Herschel*. This will shed light on the PDF at the high-mass end, constraining the IMF. It will allow exploring the process of mass accretion by ridges (feeding through the network of filaments, or merging of filaments, or both). We will search for revealing substructure in the most massive ridges in complete and unbiased surveys, in regions too bright to be observed with *SPICA*. Low-velocity shocks, induced by the merging of filaments, produce H_2O and SiO emission, providing yet another way to study their formation. Such studies would be complex with ground-based telescopes due to the difficulty of mapping large-scale structures through a variable atmosphere (e.g. *CCAT*). While *ALMA* offers a superior spatial resolution, it will have difficulty restoring all the physical scales, yet those hold crucial clues for the emergence of structure in dense clouds.

Understanding high-mass star formation. Massive dense cores observed today are our best candidates for the high-mass equivalent of proto-stellar cores. However, while current or planned surveys (e.g. with *SPICA*) will generate lists of such massive cores in the Galaxy, ambiguities will remain regarding their nature, i.e. proto-cluster, or massive proto-star. This limits our capacity to build a scenario for massive star formation and it can only be solved with higher spatial resolution. *ALMA* can provide case-by-case answers but statistically significant and unbiased samples will be hard to obtain (and indeed interferometric observation of these cores typically show substructure, which indicates that a large number of candidates have to be observed to isolate true massive proto-stars, see Bontemps et al. 2010). On the opposite, a facility with the capacity to resolve these cores over a significant fraction of the Galactic plane promises to advance significantly our understanding of the high-mass star formation scenario by giving us access to all the stages of the process.



Figure 3: A *Herschel* PACS & SPIRE false color map of Vela C (blue for the short wavelengths, red for the long wavelengths), highlighting regions of intense feedback (e.g. the butterfly-shaped RCW 36 in blue), the filamentary network, with the massive vela ridge in the RCW 36 region, and cores forming stars (from Hill et al. 2011).

Resolving protostellar structures is also important to measure the star formation rate (SFR). A key parameter in extragalactic studies, the SFR surface density is observed to correlate with the molecular gas surface density in external galaxies (the Kennicutt-Schmidt, KS, law), but little physical support exist to explain this relation. In the Galaxy, the SFR can be estimated using direct *Herschel* counts, rather than relying on indirect tracers. These studies reveal SFR surface densities of the order of $10 - 100 \text{ M}_{\odot} \cdot \text{yr}^{-1} \cdot \text{kpc}^{-2}$ on $1 - 10 \text{ pc}^2$ areas, i.e. worthy of starburst galaxies (see e.g. Nguyen Luong et al. 2011, where the uncertainties are driven by the unknown core to massive star efficiency), and show that these regions already fall on the KS relation.

Evidences suggest that enhanced star formation follows the formation of the ridge through colliding flows (Nguyen Luong et al. 2011). Such a link between the formation of structure and star formation forms a natural context within which to explain the KS relation, yet these SFR studies, combining source counts and structure characterization are limited to the closest clouds due to the resolving power of our instruments. They also rely on crude assumptions about the mass transfer from the core to the star. While the latter will see much progress with targeted *ALMA* studies, unbiased surveys can only be achieved with a significant gain in angular resolution in the FIR. Indeed, with $1''$ at $100 \mu\text{m}$ in the MIR/FIR regime, we can at the same time survey the

complete structure of star-forming clouds, and identify the complete population of massive protostellar objects from the closest Galactic arm ($1 - 3 \text{ kpc}$), to the tip of the Galactic bar (6 kpc), and the central molecular zone. This diversity of targets samples the cloud characteristics as well as the cloud evolutionary stage, and will provide important clues regarding 1) the universality of the IMF and 2) the cloud size or mass scale beyond which the notion of IMF applies.

Highlighting activity with water. The water molecule is a key constituent of interstellar matter with a threefold influence on the formation of stars and planets. Gaseous water acts as a coolant of collapsing interstellar clouds, solid water enhances the coagulation of circumstellar dust grains into planetesimals, and liquid water on planetary surfaces brings organic molecules together which helps to start the emergence of life. The first role is especially important for high-mass star formation, which depends on the balance between the collapse of a massive gas cloud and its fragmentation; this balance depends strongly on the temperature. The great sensitivity of the water abundance to the gas temperature, which is much larger than it is for CO, makes water a useful probe of the high-mass star formation process, which is currently only partially explored due to observational difficulties.

Thermal water lines give information on the physical conditions and chemical evolution of star-forming regions, but require space-based platforms for observation. Following pioneering work with *ISO* at low spatial and spectral resolution, and with *SWAS* and *Odin* of a single line at high spectral but low spatial resolution, *Herschel* has revolutionized our view on interstellar water, by making a survey of water in ~ 100 protostars spanning a large range in luminosity (from ~ 1 to $10^5 L_{\odot}$) and evolutionary stage (from pre-stellar cores to proto-stellar objects and proto-planetary discs), see van Dishoeck et al. (2011).

This survey showed that bright water emission is limited to the embedded protostellar phase, where it contributes up to 50% of the total far-infrared cooling power, and is dominated by dense warm shocked material. It also indicated that most interstellar water is formed on dust grains. But the biggest surprise came from the low water abundances measured in hot cores, where icy grain mantles evaporate into the gas phase. The grain mantles are mostly water ice, but the gas phase abundance is $\sim 100\times$ lower than the solid state abundance. One possibility is incomplete evaporation,

because the grains are shielded from radiative heating, for example in a disc configuration.

Analysis was greatly limited by the lack of spatial resolution. High angular resolution observations would reveal the places where water really radiates and could complete our understanding of activity in the star-forming process. The spectrally resolved water lines around $300\mu\text{m}$ have been proven to be the most useful, and the $987\text{ GHz}/304\mu\text{m}$ line appears to be a good tracer of the mean weighted dust temperature of the source, which may explain why it is readily seen in distant galaxies.

The *Herschel* results demonstrate the use of water as a tracer of energy input into interstellar gas, that can be greatly enhanced at higher angular resolution. High-J CO lines give similar information, but are outside the reach of *ALMA* and require very broad-band spectra. The $203\text{ GHz}/1.48\text{ mm}$ line of H_2^{18}O with *ALMA* will probe warm water, but with large uncertainties as only one transition is observed. Operating in space around $300\mu\text{m}$ one could image the ground state lines of ortho and para H_2O , and the 304 and $273\mu\text{m}$ lines of excited para and ortho H_2O . This will give immediate access to areas with enhanced activity, highlighted in water emission. The H_2O^+ lines near $269\mu\text{m}$ trace the cosmic-ray ionization rate and the ortho-para ratio of water, crucial to understand its formation process.

Finally access to the water ice at 44 and $62\mu\text{m}$ will complete our picture on the interplay between physical conditions, activity and the chemistry of water for regions which are too bright to be observed with *SPICA* but which are crucial for understanding the details of high-mass star-formation.

Magnetic fields in warm regions In the diffuse interstellar medium, outside star-forming molecular clouds, the kinetic energy from interstellar turbulence and the magnetic energy are comparable. Both are much larger than the cloud gravitational binding energy and the gas internal energy. For stars to form, gravity must become locally the dominant force. This occurs where the turbulent energy has dissipated and matter has condensed without increasing the magnetic field flux in comparable proportions. When and how frequently this occurs is a key question for our understanding of what regulates the efficiency of star formation. This is a long standing conundrum that theorists have addressed in many ways over decades, focusing on either ambipolar diffusion, turbulence or magnetic reconnection to allow the formation of “supercritical” dense cores in which stars may form. The actual role the magnetic field plays in

star formation remains quantitatively debated due to the paucity of high angular resolution data.

The magnetic flux problem also lies at the heart of the formation of stars and of proto-planetary disks. The magnetic flux of the pre-stellar core must be decoupled from the matter that enters the star. When and how the decoupling occurs is a fundamental problem of star formation that has yet to be resolved. Magnetic braking is required to draw out the angular momentum of the collapsing core. However, it may also prevent formation of centrifugally supported circumstellar disks. Again, this is a topic highly debated among theorists but there are no data on the field structure on relevant scales to test proposed solutions.

Data are slowly becoming available from Submm observations of dust polarisation at a few arcseconds resolution with interferometers (e.g. Hull et al. 2013, and references therein). This will become a key science topic for *ALMA* but there are fundamental aspects for which FIR polarimetry will be essential. First, the FIR is necessary to isolate the warmer environments next to the proto-stars from the colder dust further away. It is at FIR wavelengths that we can probe the field structure close to the star where the key action for the formation of stars and disks occurs. For heated dust, a FIR space observatory is vastly more sensitive than *ALMA* in the Submm. The gain in sensitivity is one order of magnitude for 18 K dust observed at $100\mu\text{m}$. Second, it is only from space we will get the sensitivity to map filament’s polarization over substantial areas, to evidence the role magnetic fields play in the emergence of structures. Even for 10 K dust, space observations at $300\mu\text{m}$ still gain a factor of 5 in the column densities that can be mapped in polarization w.r.t. *ALMA*, and we will overcome its limited wide-field imaging capabilities, so that we will be able to make the connection with the B-fields in the parent, more diffuse molecular clouds.

1.3 The nearby Universe

The Galaxy will remain the target of choice when access to the smallest physical scales is required, however it also creates a number of fundamental limits to our studies: distance ambiguities can sometimes not be resolved, line-of-sight pile up confuses our interpretation, and the galaxy itself does not offer a wide range of physical conditions. High angular resolution offers the unique prospect of allowing to expand studies currently performed only in the Galaxy to objects in significantly different evolutionary or energetic stages. In the FIR,

the pressing issues that can be tackled using galaxies in our direct environment deal with the dust enrichment in galaxies, the impact of interstellar medium physics on the emerging spectral energy distribution, the star formation process, and the physics in the direct environment of active galactic nuclei.

The dust budget in galaxies. While dust masses are routinely estimated in galaxies using e.g. *Herschel*, how the Universe became so dusty is still very uncertain. The major sources of dust in galaxies had been assumed to be low- and intermediate-mass stars (LIMS), during their AGB mass-loss phase. However, Submm detections of very dusty high redshift galaxies, some observed less than 10^9 yr after the Big Bang (e.g. Bertoldi et al. 2003), have focussed attention on the potential contribution of massive stars to the dust budgets of galaxies (e.g. Dwek et al. 2007), specifically dust formation in the ejecta from core-collapse supernova (CCSN). In addition, evidence is mounting that LIMS may not be able to fully account either for dust in nearby galaxies (e.g. Matsuura et al. 2009).

According to models for supernova dust formation and for dust evolution in galaxies (e.g. Bianchi & Schneider 2007; Dwek et al. 2007), CCSNe should create between 0.1 and $1.0 M_{\odot}$ of dust per event in order to account for observed dust masses in galaxies. However, mid-infrared studies of CCSNe typically found that less than $0.001 M_{\odot}$ of warm dust (200-450K) formed during the first 3 to 4 years after outburst (e.g. Meikle et al. 2011), i.e. $100\text{--}1000\times$ lower than required.

Herschel has allowed us to probe for much cooler dust (< 100 K) in young supernova remnants (SNRs), in particular in SNRs young enough (< 1500 yrs) that the amounts of swept-up interstellar dust are negligible. *Herschel* observations of the 330-yr old Cas A CCSNR showed it to contain $0.075 M_{\odot}$ of cool (35 K) dust (Barlow et al. 2010) which together with the $0.025 M_{\odot}$ of warmer dust found with Spitzer (Rho et al. 2008) implied a total of $0.10 M_{\odot}$ of dust in this remnant. *Herschel* observations of the 950-yr old Crab Nebula CC-SNR (Gomez et al. 2012), and of the much younger 25-yr old CCSNR SN 1987 A (Matsuura et al. 2011) turned out respectively $0.1\text{--}0.2 M_{\odot}$ of cool dust (< 34 K) and $0.4\text{--}0.7 M_{\odot}$ of cold dust (~ 20 K) detected. Whether a significant amount of colder dust exists is currently impossible to test given the confusion with foreground dust emission.

The question of whether CCSN are significant contributors to the dust found in galaxies is thus re-opened.

However, only the above three young CCSNRs could be studied by *Herschel*, and it is not clear either whether any of the dust will survive their passage through the shock wave into the general ISM. Further progress depends on observing a significantly larger sample of young CCSNRs at far-infrared wavelengths in different galaxies and at different ages. Significantly better sensitivities and angular resolutions than provided by *Herschel* will be needed to separate out the different components that emit in the FIR beside dust formed in the SNR. *JWST*/MIRI will represent a substantial leap in capacities, but it will only be sensitive to hot to warm dust, which constitute a very small fraction of the dust present in CCSNRs.

SPICA, with the confusion limits, will not detect the far-IR emission from young CC-SNRs in nearby galaxies against their strong ISM background. *ALMA* will not suffer from this confusion, however it will detect mostly emission dominated by the cold interstellar environment and the non-thermal emission from the remnant, and thus will provide little constraints on the dust mass formed in the SNR (see Barlow et al. 2010, on Cas A). A FIR instrument in space with an angular resolution $0.1\text{--}1''$ at $100 \mu\text{m}$ would transform this outlook: placed at a distance of 10 Mpc, the ejecta of Cas A correspond to an angular size of $\geq 0.1''$, which offers the prospect of separating CCSNR from their complex environment in a substantial number of galaxies. This would allow the investigation of dust formation and chemistry in SNRs at different stages of their evolution, by resolving the nearest very young (< 30 yr) remnants, as well as completing the 30-300 year gap in SNR age in the current “sample”. Considering that the spiral galaxy NGC 6946 (6 Mpc) has hosted nine CCSNe in the past 100 years, a total sample of 50-60 young CCSNRs is in reach and would enable the rate at which dust is formed by massive star SNe to be meaningfully quantified for the first time, and could help pin down the timescales in which dust is formed and destroyed in the SNRs.

Thermal balance in galaxies. Radiative transfer within galaxies, where ultraviolet and optical light from stars and other sources is absorbed by dust and reradiated as infrared light, is the fundamental process through which the SED of galaxies is built. Properly understanding this process is key given the vast number of interpretations that are derived solely from galaxy SEDs.

Herschel revealed that dust in nearby galaxies is found in different thermal phases (see for example Bendo et al. 2010, 2012; Boquien et al. 2011; Groves et al.

2012; Smith et al. 2012). A warmer phase observed at $\lambda \leq 160 \mu\text{m}$ is heated by star forming regions, while a colder phase observed at $\lambda > 250 \mu\text{m}$ microns is heated by the diffuse interstellar radiation field from the total stellar population. Infrared emitting regions can thus be represented as shells around star forming regions. In the inner parts, the dust is heated by the young stars while in the outer parts it is shielded from these stars, and is heated by older stars in the vicinity of the cloud. The balance between these two situations depends upon the properties of the galaxies: in M81, bulge stars heat the dust observed down to $70 \mu\text{m}$, while in the late-type M83, most of the dust observed at $\lambda < 250 \mu\text{m}$ is heated by star forming regions (Bendo et al. 2012). Thus only a proper coverage in the FIR range can allow a quantitative connection between the dust luminosity of galaxies and their young star content.

This work is however limited by the angular resolution of *Herschel*: molecular clouds ($\sim 100 \text{ pc}$) are only marginally resolved in the closest spiral galaxies. Yet we need to resolve star forming regions to examine how exactly the dust is heated. Unfortunately, the situation will not change with *SPICA*, as it has the same angular resolution as *Herschel*. On the opposite, a resolution of $\sim 1''$ in the FIR ($\sim 25 \text{ pc}$ at 5 Mpc), will allow us to look at the detailed structures of dust at the molecular cloud scale, resolving the processes of radiative transfer in nearby objects such as strong starbursts (M82, M83, NGC 253), AGNs (Cen A, Circinus), or galaxies with large bulges (M81), environments covering the diversity found in the more distant Universe.

With a spectral resolution of $R \geq 1000$, these analyses can be greatly enhanced by observations in the FIR fine structure lines ([OI] $63 \text{ \& } 145 \mu\text{m}$, [OIII] $88 \text{ \& } 52 \mu\text{m}$, [CII] $158 \mu\text{m}$, [NII] $122 \text{ \& } 205 \mu\text{m}$, [NIII] $57 \mu\text{m}$). *Herschel* revealed both the exquisite level of insight that can be gained on the ISM and how simple our preconceptions can be before we actually model the emitting regions. For instance, Cormier et al. (2012) and Lebouteiller et al. (2012) show that the association of the [CII] line with photo-dissociation regions, and hence star formation, is much more complex than thought, meaning that significant “calibration” work is required before it can be used as a star formation tracer at high redshift. *SOFIA* and *SPICA* will bring this nascent field to a much more established base, but the modeling of these lines is very dependent on the actual geometry (filling factors), and involves a multitude of phases for even a single line (see e.g. Pineda et al. 2013), thus large ambiguities will persist as long as the characteris-

tic physical scales involved are not resolved.

This is precisely the promise of the architectures proposed here, coupling ideally these spectral studies with dust continuum studies to fully constrain the interplay between radiation sources and the ISM in galaxies. Furthermore, reaching spatial resolutions comparable to *JWST* and *ALMA* will amplify the impact these facilities will have on our understanding of the ISM. With the *JWST*, we will enrich the spectral studies with MIR fine structure lines, related to the more active part of the star forming regions, e.g. [NeII], [NeIII], as well as with the PAH bands that trace directly young massive stars. Combining with *ALMA* surveys in CO and CI will allow to address fundamental issues in ISM physics. Of particular interest is the question of CO-dark molecular gas, i.e. H_2 gas that is not traced by CO. Recent FIR studies in the Milky Way have revealed the importance of this phase (30% on average, Pineda et al. 2013), while joint *Fermi-Planck* studies show this CO-dark gas to be widespread (Grenier, in prep.). With a high-angular resolution FIR facility, combined with *EVLA* and *SKA* to constrain the neutral and ionized phases better, the same studies can be deployed in nearby galaxies, and we can understand how the abundance of CO-dark H_2 is influenced by physical conditions, thus providing more control on the CO/ H_2 ratio that will still be heavily in use for high-redshift studies.

Massive star formation beyond the Milky Way.

It is well known that the star-formation rate per comoving volume decreased by an order of magnitude between $z = 1$ and now (e.g. Madau et al. 1996; Heavens et al. 2004). There was certainly not $10 - 30\times$ more molecular gas present at that time, so this suggests that star-formation efficiency (SFE) was significantly higher then. The star-formation rate today is dominated by large spiral galaxies, but this was much less true in the past when spirals were smaller (at least in their stellar part), bluer, more gas-rich, and chemically poorer.

This shows that star formation must be sampled in the widest range of environments and Local Group galaxies provide this test-bed. The Magellanic Clouds, with metallicities $\sim 1/3$ and $\sim 1/5$ Solar, and masses $\sim 1/10$ and $\sim 1/50$ of the Milky Way, are only 55 and 70 kpc away, and have a higher gas-to-stellar mass ratio than the Milky Way. A high angular resolution FIR instrument could enable us to study proto-stellar condensations in these galaxies in a manner analogous to Galactic work. In the Galaxy however, we are restricted to sampling star formation in a solar metallicity envi-

ronment, in the gravitational potential of a large spiral with differential rotation, shear, etc. Furthermore, in the Galaxy, the stellar surface density dominates over the gas out to quite large radii, which could have an effect on the structure of molecular clouds, the efficiency of star formation, and thus potentially the IMF. The conditions in the Magellanic clouds are different both dynamically, chemically, and in terms of the gas/star mass ratio, and their HII regions are much more active sites of star formation than anywhere in the Milky Way.

At 55 kpc, 0.1'' probes sizes down to 5000 AU, e.g. near solar-mass condensations, while 1'' resolves the 0.1 pc scale of Galactic star forming filaments. The FIR is the crucial wavelength regime here because most of the energy of proto-stellar condensations is emitted at wavelengths shorter than can be feasibly detected with *ALMA*. The FIR thus offers the only method of measuring temperatures and luminosities, and thus mass, so that, by mapping star-forming regions in a range of galaxies, the clump mass function can be constructed and compared to its Galactic counterpart, to investigate possible departures from universality.

With moderate resolving power we can map the feedback of cosmic rays in starburst regions. There is observational (e.g. Schilke et al. 1993) and theoretical evidence (e.g. Papadopoulos 2010) that massive star formation implies regions of extremely high cosmic ray energy density with higher heating rates and higher ionization fractions. As starbursts are episodic and localized, cosmic ray ionization rates are probably not constant in time nor space. Yet, it is of paramount importance to determine the cosmic ray ionization rate variations as they regulate the ionization fraction of the molecular gas and hence star formation (Papadopoulos 2010), and the best tracers of cosmic ray ionization rate are H_2O , OH, C, C^+ , CO, accessible preferentially in the FIR.

The AGN/host relationship. Active Galactic Nuclei (AGN) are of fundamental importance in astrophysics. By converting gravitational energy into radiation, they are the most powerful objects in the Universe. More importantly, they play a fundamental role in galaxy evolution, as they can impact the global properties of their host. It is thus crucial to study in the local Universe not only the accretion physics, but the AGN feeding and feedback on the galaxy as well, in forms of winds, jets and outflows. To do so we need extinction-free spectroscopic tracers, moderate spectroscopic resolving power and high spatial resolution. Disentangling the complex interactions occurring near AGNs can

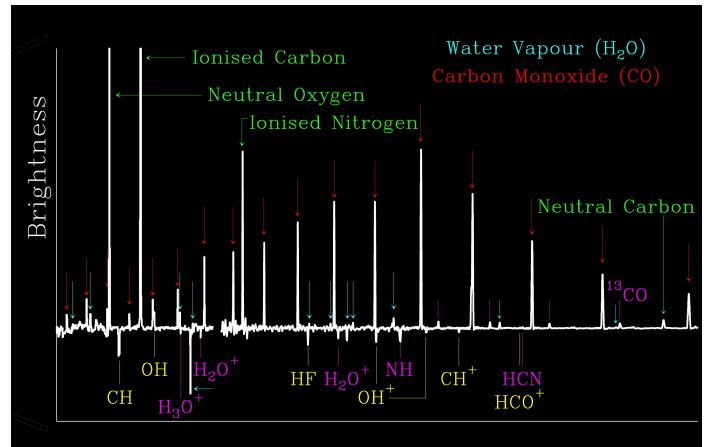


Figure 4: The rich *Herschel* PACS and SPIRE spectra from the central region (15'' Goicoechea et al. 2013).

only be performed at MIR/FIR wavelengths because 1) the regions are often heavily obscured at shorter wavelengths and 2) this spectral region contains a large amount of atomic and molecular lines covering a wide physical parameter space. The MIR/FIR ionic fine structure lines produced in the AGN emission regions can characterise not only the gas conditions but also the primary spectrum from the ionising source, disentangling AGN from stellar ionisation (e.g. Spinoglio et al. 2005); the FIR high-J CO lines can also trace the nature of the source that is pumping energy into the BH environment, and thus separate the impact of the AGN, through X-rays, shocks and winds, from that of the central starburst regions, contrary to the situation with submm low-J CO lines, that trace low excitation material. Other molecular tracers such as H_2O , OH, HCN, OH^+ , H_2O^+ accessible only in the FIR, can be used to derive fundamental properties of the BH environment, as *Herschel* observations of a few nearby AGN (González-Alfonso et al. 2013) and of the GC (Goicoechea et al. 2013, and Fig. 4) have demonstrated. These molecular tracers revealed very massive outflows (Sturm et al. 2011), however elucidating their driving mechanism requires probing their structure at sub-arcsecond spatial resolution, beyond what *Herschel* provided (or what *SPICA* or *SOFIA* will deliver). This is crucial to reveal how BH-powered outflows can entrain most of the molecular gas in the host, as required by the quenching paradigm for the evolution of galaxies.

With 0.1-1'' angular resolution almost all the components of nearby AGNs (Cen A, NGC 1068, Circinus) can be resolved, from the molecular torus (10 pc) out to the outflow and jet structure (1 kpc). FIR studies of

these objects will provide the interpretation guidelines needed for application to higher redshift objects, where the AGN/host interaction is a key component in galaxy evolution. While some of these interactions will be accessible to *ALMA* for the redshifted objects, it is only with the FIR domain in the local Universe that we can put the physics of the interaction on a firm basis.

Turning closer to us to the Galactic Center (GC), the potential of high angular and spectral resolution is enormous. The GC hosts a supermassive black hole ($4 \times 10^6 M_\odot$) and represents a unique environment to study a broad variety of phenomena (from black hole accretion, star formation and IMF studies in extreme physical conditions, to the impact of shocks, turbulence, magnetic fields, high energy radiation and cosmic-rays in the environment). The distribution of gas and dust in the GC shows the presence of a central cavity of ~ 1.5 pc size and a circumnuclear disc (CND) between ~ 1.5 pc and ~ 5 pc. The CND is the reservoir for material accreting into the central parsec of the Galaxy and it is very likely a fundamental feature of most galaxies. However, despite the proximity of the GC compared to extragalactic nuclei, the angular resolution of *Herschel* does not allow resolving the fundamental structures of the innermost regions of the galaxy, and sub-arcsecond resolution in the FIR will be necessary.

1.4 The evolving Universe

The FIR waveband is key to our understanding of galaxy evolution. Such a claim rests on the foundation that the most dramatic phase of evolution for AGN and their host galaxies occurred between $z \sim 3$ and the present day, a period of 11.5 Gyr or 84% of the age of the Universe. For much of this interval, the thermal continuum peak that constrains T_{dust} and hence M_{dust} , L_{IR} and ultimately the obscured star-formation rate, as well as fine-structure lines of ionised atoms such as [O III] $\lambda 88 \mu\text{m}$ and [C II] $\lambda 158 \mu\text{m}$, fall in a wavelength domain, the FIR, that only a space mission can capture. It is only at very high redshift, $z > 3$ that these enter into the atmospheric windows accessible to *ALMA*.

Furthermore, reaching, as we propose, arcsecond to sub-arcsecond resolution in this key domain will represent a significant turning point. First, it essentially breaks free of the confusion limit that marred the *Herschel* surveys on nearly all the FIR domain. Second, thanks to the turnover in angular diameter distance in concordance cosmology, a spatial resolution of $0.1''$ samples sub-kpc structure at any redshift ($\sim 0.8''$ at

$z = 1-3$; rather better at $z < 1$). Finally, when working beyond the local Universe we can rely on gravitational amplification to achieve even higher spatial resolution, with surface brightness conserved. Thanks to large-area surveys with *Herschel* and the South Pole Telescope, we have hundreds of strongly lensed galaxies, selected at $250 \mu\text{m}$ and beyond (e.g. Negrello et al. 2010). but post *Euclid* and *eROSITA*, we will have discovered and created high-precision mass models for many hundreds of galaxy clusters. Thus we will no longer need to rely on extreme FIR-submm brightness to select strongly lensed galaxies, meaning that we can construct samples representative of unlensed populations. The vast majority of these galaxies will be amplified by the smooth, cluster-wide gravitational potentials, rather than the cuspy potentials of individual galaxies which can give rise to differential amplification. Such clusters offer magnifications ranging from a few to $\sim 30\times$ (e.g. Swinbank et al. 2010), yielding milli-arcsecond resolution. Thus while we cannot rule out a future requirement for $< 0.1''$ spatial resolution for studies of the distant Universe, it is not yet obvious that it will be needed.

Disentangling star formation and the obscured growth of black holes. Galaxy evolution is characterised by the interplay of three main phenomena across cosmic time: accretion onto black holes in AGN, star formation, often occurring in energetic bursts, and the feedback related to both. These processes jointly determine the energy budget of a galaxy throughout its evolution and an evolutionary sequence from starburst-dominated through active nuclei has been suggested (e.g. Sanders et al. 1988; Storchi-Bergmann et al. 2001). The growth of bulges through star formation may be directly linked to the growth of black holes through accretion, resulting in the tight local correlation between the mass of the stellar spheroid and the central black hole (e.g. Merritt & Ferrarese 2001).

Much of this evolution is hidden by dust, causing up to hundreds of magnitudes of optical extinction. Uniquely, rest-frame MIR-FIR spectroscopy is able to trace these physical processes. The heavily obscured ISM is energised by the host's star formation and its growing AGN, and IR spectroscopy provides the diagnostics to distinguish between and quantify the two, measuring the separate luminosity functions of accretion and star formation as a function of cosmic time. The FIR waveband offers a unique tool to study the effect of both radiative and dynamical feedback: high-resolution imaging spectroscopy ($< 1''$, $R \sim 1000$) at rest-frame $20-60 \mu\text{m}$ (ob-

served FIR) allows us to distinguish between the AGN-heated and starburst-heated components and constrain the possible joint evolutionary scenarios for galaxies and QSOs (e.g. Farrah et al. 2007). Imaging spectroscopy in the FIR waveband is thus the most important observational tool required to measure star formation as a function of redshift, whilst disentangling the effects of black-hole accretion, thereby elucidating what we know of galaxy/QSO evolution during the period when galaxies and AGN undergo their most significant evolution.

The relevant methods are well developed from *ISO*, *Spitzer* and *Herschel* results (e.g. Genzel et al. 1998; Dale et al. 2006; Spoon et al. 2007). AGN are identified in the rest-frame MIR through high-excitation NLR lines and by hot dust re-radiating absorbed primary AGN emission. This is a key advantage since both methods can detect Compton-thick sources ($N_{\text{H}} > 1.5 \times 10^{24} \text{ cm}^{-2}$) that largely escape X-ray surveys, even at hard X-ray energies. However, diagnostics inferred from integrated light provide information only on the dominant source of energy in each galaxy. Low- and intermediate-luminosity (Seyfert-like) AGNs are completely missed in starbursting galaxies, because of the overwhelming emission and dilution from the latter; moderate star formation is difficult to detect in powerful quasar host galaxies. Furthermore, low angular resolution also means sources rapidly fall below the confusion limit. These limitations hamper our ability to understand galaxy–BH co-evolutionary processes.

Only with high spatial resolution can sensitive FIR spectroscopy ($R \sim 100\text{--}5000$) fulfill its potential, measuring both star-forming and AGN components in distant galaxies, irrespective of dust obscuration, as a function of cosmic time, over a representative range of luminosities, metallicities and stellar masses. We can also trace binary BHs, on separation scales from $\sim 1\text{--}100$ kpc, providing a robust estimate of the birth rate of BH mergers for detection via gravitational waves. We can thereby constrain BH–galaxy co-evolutionary models, which provide detailed but very different predictions on the relative roles of star formation and BH accretion as a function of stellar mass and cosmic epoch. Some models suggest that joint BH–spheroid growth is triggered by major mergers, which drives gas towards the central BH, and by the eventual coalescence of multiple BHs from the merging galaxies; other models show that minor mergers and galaxy bar instabilities can fuel efficient accretion onto the BH. Given the copious extinction, these hypotheses can only be tested with FIR observations of distant AGN and star-forming galaxies

at high spatial resolution.

Probing galaxy evolution via turbulent molecular hydrogen. H_2 is the most abundant molecule in the Universe, representing a significant fraction of the total baryonic mass. It makes up the bulk of the mass of the dense gas in galaxies and is found in all regions where the shielding of UV photons, responsible for its photo-dissociation, is sufficiently large (i.e. where $A_V > 0.01\text{--}0.1$ mag). Star-forming galaxies are known to have large quantities of highly turbulent molecular gas (e.g. Swinbank et al. 2011). Dissipation of kinetic energy through shocks must play a major role in regulating star formation and galaxy growth: energy released by gas accretion, mergers and AGN – and associated feedback – must be dissipated in order for stars to form.

Pure-rotational H_2 line emission is an ideal tracer of this energy dissipation and dominates the cooling of shock-heated molecular gas with densities, $\sim 200\text{--}1000 \text{ cm}^{-3}$, at temperatures below 10^4 K, where the gas is too cold to emit in the UV. Line luminosities, ratios and kinematics allow us to clearly identify regions of gas that are dominated by shock heating, as *Spitzer* has shown with detections of H_2 (at the few $10^{-18} \text{ W m}^{-2}$ level – e.g. Fiolet et al. 2010) for a diverse set of nearby extragalactic environments, including interacting galaxies (Guillard et al. 2009), mergers (Herrera et al. 2012), powerful AGN (Nesvadba et al. 2010), cool-core and forming galaxy clusters (Egami et al. 2006; Ogle et al. 2012). Since H_2 contributes up to 90% of the gas cooling budget in the relevant temperature/density range (Lesaffre et al. 2013), it is the only tracer which allows us to quantify the total amount of turbulent energy dissipated. And as we go to higher redshifts, more of the H_2 lines shift into the FIR domain.

Tracing turbulence dissipation with H_2 offers a method to study the dynamical interactions at work during the formation of galaxies. To identify the driver(s) of the turbulence – cold accretion, mergers/interactions, feedback – and their roles in triggering and regulating galaxy growth, and to disentangle H_2 regions dominated by shocks from those dominated by UV heating, *sub-arcsec spatial resolution will be critical* (see for instance the example of the ‘Antennae’, Herrera et al. 2012). High-redshift, gravitationally lensed galaxies will allow us to probe H_2 and numerous other IR lines in individual star-forming regions, at spatial resolutions akin to those probed in nearby galaxies, exploring the kinematic link between the ISM on kpc scales and individual star-forming regions. Being contemporary with SKA and

E-ELT, we will also probe the gas cycle of galaxies, from the diffuse, atomic phase (with SKA), through the molecular phase in young star clusters (with E-ELT and ALMA), to the metal enrichment of the intergalactic and intracluster medium (through absorption-line work with E-ELT), for high-redshift environments as fundamental and diverse as massive, forming galaxy clusters, starburst galaxies and AGN hosts.

Molecular hydrogen emission from the first epoch of star formation. H_2 is key to our understanding the earliest evolution of primordial gas: its formation on grains initiates the chemistry of the gas, where it then plays a major cooling role. Locally, as we have seen, H_2 provides key insights into the physical processes that lead to its excitation (PDRs, XDRs, etc.).

Its role is even more significant in the early Universe, however: its importance increases with look-back time. The first stars form by gravitational collapse of primordial gas clouds, after the dark ages, several hundred million years after the Big Bang ($z \sim 20$), induced by H_2 line cooling (e.g. Saslaw & Zipoy 1967; Kamaya & Silk 2002; Mizusawa et al. 2004); in this metal-free gas, H_2 then serves as the only effective coolant. The formation of H_2 thus plays a key role in the early Universe and these first-generation stars should be bright in mid-IR H_2 line emission. Understanding exactly how this happened is one of the most important and exciting goals of modern astrophysics.

Theoretical work predicts that the ground-state transition lines, 0-0 S(1), S(2) and S(3), should have a $> 10^{35} \text{ erg s}^{-1}$ luminosity, within reach of the mission concepts proposed here. Adopting a different approach to feasibility, based on H_2 detections of nearby disk galaxies (Roussel et al. 2007, e.g.), the sum of 0-0 S(0) and 0-0 S(1) corresponds to 20–25% of the luminosity emitted in cooling lines such as [Si II] or [S III], which corresponds to $3 - 3.5 \cdot 10^{-4}$ of the total 8–1000- μm luminosity within the same area; again, this is within reach.

2 Mission requirements and straw-man concepts

While all the science cases presented above are built upon the necessity to reach the 0.1-1'' angular resolution at 100 μm , the requirements in terms of spectral resolution vary. We recap them in Table 1. The mission concepts attached to this paper enable a significant fraction of the science presented here, but not all. Progress

Table 1: Spectral resolution type required by each science case (see corresponding sections). Heterodyne stands for R 10^6 – 10^7 , medium for 1000-5000, and SED for broadband.

Science case	Resolution
Gas mass in disks	heterodyne
Water transport	heterodyne
Sotoplanetary systems	SED, medium
Feedback	SED
Massive star formation	SED
Highlighting activity with water	heterodyne
Magnetic field	SED
Dust budget	SED, medium
Thermal balance	medium
Massive star formation	SED, medium
AGN/host relationship	medium
Galaxy and AGN co-evolution	medium
H_2 for galaxy evolution	medium
First stars	medium

toward an actual mission implementation will require (1) selection of this science domain for an L-class mission and a timeline for implementation, (2) involvement of the whole community, and (3) feasibility studies for each of the concepts that only ESA can perform.

2.1 TALC - A deployable large aperture telescope

Fairings put a strict limit to the size of a single dish aperture that can be launched in space. TALC (Thinned Aperture Light Collector) is a 20 m diameter deployable concept that explores some unconventional optical solutions (between the single dish and the interferometer) to achieve a very large aperture. Its collecting area is 20 times larger than *Herschel*'s, giving access to very faint and/or distant sources. With an unconventional optical design comes the necessity to combine data acquisition with unconventional data processing techniques, which are being developed today, based on the notion of sparsity in astronomical signals (e.g. Starck et al. 2010).

The deployable mirror structure exploits the concept of *tensegrity*, i.e. when structural rigidity is achieved through compression. The TALC mirror (see Fig. 5) is a segmented ring of 20 m diameter and 3 m width. For launch the identical mirror segments are stored on top of each other and a deployable mast pulls a series of cables that deploy the stack of mirror into the required

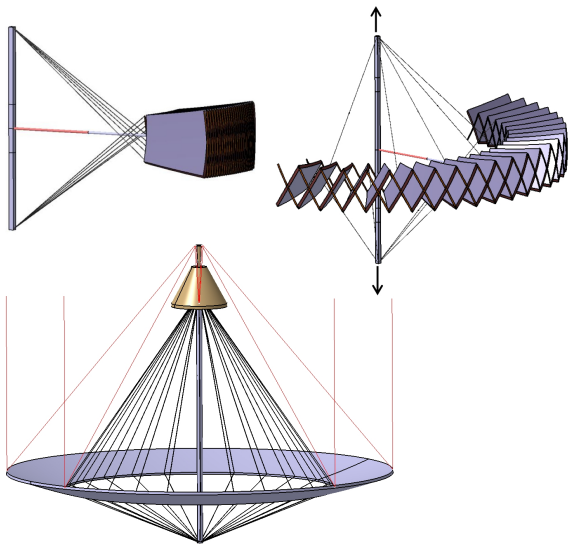


Figure 5: TALC deployment (clockwise). At top left the mirror segments are still stowed, but have been pushed away from the central mast, which extends, deploying the whole structure. Red lines indicate the optical path to the instrument platform (figure from A. Bonnet).

shape. Tension on the cables applied by the central mast provides stiffness to the inner diameter of the deployed ring. On the outer diameter, a degree of freedom persists that allows optimization of the mirror geometric shape by adjusting each of the segments with respect to its neighbors. We foresee an active system using reference stars in the NIR to optimize the segments' position for FIR operations.

Because the aperture is not filled, TALC exhibits a main beam size that is narrower than that of a 20 m single dish, and reaches $0.9''$ at $100\,\mu\text{m}$. While this main beam contains 30% of the total energy, simulations of typical observations campaigns demonstrate that we have the numerical tools at hand to restore a clean map at nominal resolution.

The mirror surface is passively cooled, with a concept that borrows from the *JWST* sunshield. Sensitivity estimations have been performed with a mirror temperature of 80 K. Simulations with an 80 K mirror show that a sensitivity of $0.1\,\text{mJy } 5\sigma \text{ } 1\,\text{hr}$ is reached.

The available field of view is $\sim 2'$ and the instrument bay can find ample room just below the secondary, allowing for a suite of instruments to be implemented on the telescope. TALC is currently foreseen to embark imaging instruments (to exploit the field of view and the sky accessibility), with polarimetric and medium spectroscopic in-pixel capacities (such as those studied in

the context of the FOCUS collaboration¹). Preliminary investigations show that implementation of very-high resolution heterodyne spectroscopy can be envisioned as well.

2.2 FIRIT - Direct imaging interferometer

A conservative design with sub-arcsecond spatial resolution, the direct-detection interferometer option is based on a pre-existing ESA CDF study (Lyngvi 2006, to which we refer for many of the technology details) and an extensive NASA study (Leisawitz et al. 2007). This straw-man concept (see Fig. 6) would see the combination of input beams from two telescopes, moving along an unfolding or telescopic boom, in a third hub unit with the detecting part of the payload. A promising candidate for the payload identified in both Lyngvi (2006) and Leisawitz et al. (2007) is the double-Fourier modulation (DFM) technique (Mariotti & Ridgway 1988) which allows spectral and spatial interferometry to be performed simultaneously in a single instrument, providing integral field spectroscopy. This technique was recently demonstrated at sub-mm frequencies (Grainger et al. 2012).

A four-spectral band configuration covering the range $25\text{--}400\,\mu\text{m}$ would use Nyquist sampling to its advantage by employing a single scanning mechanism delay line to produce multiples of the optical delays for each band (Leisawitz et al. 2007; Wilson et al. 2007) to achieve a $\lambda/\Delta\lambda > 3000$ spectral resolution in each spatial resolution element.

With two 1 m dishes and a boom extending to 36 m (Leisawitz et al. 2007) or potentially more depending on the mission timeframe, an angular resolution of $2.8''$ at $400\,\mu\text{m}$ and $0.18''$ at $25\,\mu\text{m}$ would be achieved for a 1×1 field of view. A trade-off study would then establish the possibility of increasing the sensitivity (with 2 m dishes) or angular resolution (increased boom length) in order to fit in the Ariane 5 fairing which Leisawitz et al. (2007) has shown to be limiting in volume rather than mass. With cooled mirrors (technology currently planned for *SPICA*) a point-source sensitivity of $\sim 10\,\mu\text{Jy}$ for the continuum and $10^{-19}\,\text{W.m}^{-2}$ per spectral resolution element could be achieved at $100\,\mu\text{m}$.

Current detectors meet the NEP and array size requirements for FIRIT, but further advances are needed to decrease the detector response time while holding the NEP to a few $10^{-19}\,\text{W}/\sqrt{\text{Hz}}$. Focal plane cooling to $\sim 30\,\text{mK}$ may solve this problem. In addition,

¹ipag.osug.fr/Focus-Labex/

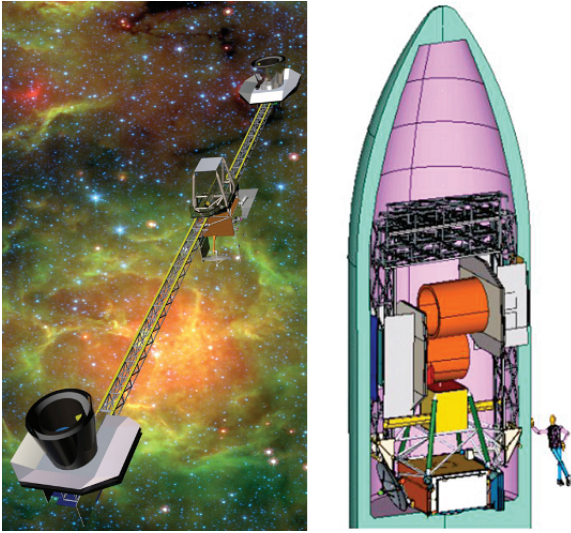


Figure 6: The FIRIT concept. Left: spacecraft concept design; right: the stowed configuration in both showing how the 1 m-class primary mirror design would fit into the medium fairing Ariane 5.

two recent REA-sponsored activities have begun (FP7-SPACEKIDS² and FP7-FISICA³) to respectively investigate the further development of Kinetic Inductance Detectors for space applications and to advance beam-combination, cryogenic delay lines, position metrology and to create a representative instrument simulator for a direct detection interferometer such as FIRIT.

2.3 ESPRIT - A heterodyne interferometer for the FIR

The concept here is based on HIFI/ALMA experience and the ESPRIT study (Wild et al. 2008, and see Fig. 7), adapted to 1'' angular resolution. To achieve this 4 dishes of 3.5m diameter will be employed with baselines upto 50m in a free-flying configuration. The uv -plane will be filled by letting the dishes drift with respect to each other. The use of phase-closure and self-calibration will guarantee optimal image quality, while the heterodyne principle creates a long coherence length and allows for relaxed positioning requirements

The ESPRIT concept consists of the following parts: Dishes for collecting the FIR waves; high sensitivity, large bandwidth, sensitive heterodyne mixers for down-converting the THz signals; Local Oscillator (LO) system for controlling the phases within the system and

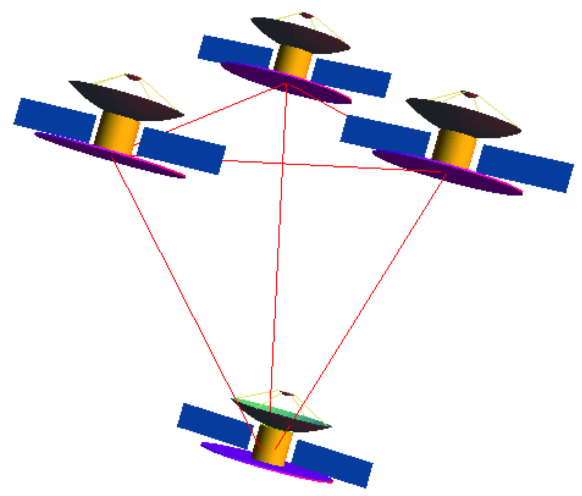


Figure 7: A schematic view of the free-flying ESPRIT interferometer (figure from A. Baryshev).

providing the ultrastable signal needed for heterodyne operation; the correlator system including optical data links and the cooling system.

The dishes should have a small wavefront error in order to optimize the shape of the main beam. The system is preferably off-axis to avoid standing waves. Lightweighthing of the dishes is a necessity, since 4 are needed to be brought in space.

Progress in SIS (now available upto 2THz) and HEB technology has brought the noise temperature of these mixers down to a few times the quantum limit (Kloosterman et al. 2013). Stability of HEB devices used to be problematic. However, new ways of controlling the output power of THz LO systems have been found, improving the stability. These methods allow accurate measurements of continuum, besides the targeted spectral lines, to be made (Hayton et al. 2012). Employing small arrays would increase the field-of-view.

New concepts for Local Oscillators have now been proven to work. These include Quantum Cascade Lasers (QCL) for the highest frequencies, besides the HIFI-technology which is now reaching 2.7 THz.

Because the data rates in interferometry are large, correlation has to be done in space. Wild et al. (2008) show that a distributed correlator system is the optimal system for resources and redundancy. Only the visibilities (amplitude and phase) will need to be downloaded to Earth.

Cooling of the dishes is not essential but can be done passively. Only the detecting elements need to be cooled down to around 4K, which can be achieved by small cryocoolers (e.g. Planck heritage).

²www.astro.cardiff.ac.uk/newsandevents/?page=news_detail&news=0136

³www.fp7-fisica.eu/

References

- Acke, B., et al. 2012, *A&A*, 540, A125
- Agnor, C. B. & Ward, W. R. 2002, *ApJ*, 567, 579
- André, P., et al. 2010, *A&A*, 518, L102
- Arzoumanian, D., et al. 2011, *A&A*, 529, L6
- Barlow, M. J., et al. 2010, *A&A*, 518, L138
- Bendo, G. J., et al. 2012, *MNRAS*, 419, 1833
- Bendo, G. J., et al. 2010, *A&A*, 518, L65
- Bergin, E. A., et al. 2013, *Nature*, 493, 644
- Bertoldi, F., et al. 2003, *A&A*, 406, L55
- Bianchi, S. & Schneider, R. 2007, *MNRAS*, 378, 973
- Bontemps, S., et al. 2010, *A&A*, 524, A18
- Boquien, M., et al. 2011, *AJ*, 142, 111
- Boss, A. P. 2003, *ApJ*, 599, 577
- Casassus, S., et al. 2013, *A&A*, 553, A64
- Chabrier, G. 2003, *PASP*, 115, 763
- Chiang, E. & Laughlin, G. 2013, *MNRAS*
- Cormier, D., et al. 2012, *A&A*, 548, A20
- Dale, D. A., et al. 2006, *ApJ*, 646, 161
- Dent, W. R. F. 2013, *PASP*, in press
- Dwek, E., et al. 2007, *ApJ*, 662, 927
- Egami, E., et al. 2006, *ApJLett*, 652, L21
- Farrah, D., et al. 2007, *ApJ*, 667, 149
- Fiolet, N., et al. 2010, *A&A*, 524, A33
- Genzel, R., et al. 1998, *ApJ*, 498, 579
- Goicoechea, J. R., et al. 2013, *ApJLett*, 769
- Gomez, H. L., et al. 2012, *ApJ*, 760, 96
- González-Alfonso, E., et al. 2013, *A&A*, 550, A25
- Grainger, W. F., et al. 2012, *App. Opt.*, 51, 2202
- Groves, B., et al. 2012, *MNRAS*, 426, 892
- Guillard, P., et al. 2009, *A&A*, 502, 515
- Haisch, Jr., K. E., et al. 2006, *AJ*, 132, 2675
- Hayton, D. J., et al. 2012, in *SPIE Conference Series*, Vol. 8452
- Heavens, A., et al. 2004, *Nature*, 428, 625
- Hennebelle, P. & Chabrier, G. 2008, *ApJ*, 684, 395
- Hennemann, M., et al. 2012, *A&A*, 543, L3
- Herrera, C. N., et al. 2012, *A&A*, 538, L9
- Hill, T., et al. 2011, *A&A*, 533, A94
- Hogerheijde, M. R., et al. 2011, *Science*, 334, 338
- Hull, C. L. H., et al. 2013, *ApJ*, 768, 159
- Kamaya, H. & Silk, J. 2002, *MNRAS*, 332, 251
- Kloosterman, J. L., et al. 2013, *Applied Physics Letters*, 102, 011123
- Kornet, K., et al. 2002, *A&A*, 396, 977
- Lebouteiller, V., et al. 2012, *A&A*, 548, A91
- Leisawitz, D., et al. 2007, *Adv. Sp. Res.*, 40, 689
- Lesaffre, P., et al. 2013, *A&A*, 550, A106
- Lissauer, J. J. 1993, *ARA&A*, 31, 129
- Lyngvi, A. 2006, *FIRI ESA CDF Study Report*
- Madau, P., et al. 1996, *MNRAS*, 283, 1388
- Malfait, K., et al. 1999, *A&A*, 345, 181
- Mariotti, J.-M. & Ridgway, S. T. 1988, *A&A*, 195, 350
- Matsuura, M., et al. 2009, *MNRAS*, 396, 918
- Matsuura, M., et al. 2011, *Science*, 333, 1258
- Meikle, W. P. S., et al. 2011, *ApJ*, 732, 109
- Merritt, D. & Ferrarese, L. 2001, *ApJ*, 547, 140
- Meyer, M. R., et al. 2008, *ApJLett*, 673, L181
- Minier, V., et al. 2013, *A&A*, 550, A50
- Mizusawa, H., et al. 2004, *PASJ*, 56, 487
- Moore, M. H. & Hudson, R. L. 1994, *A&AS*, 103, 45
- Negrello, M., et al. 2010, *Science*, 330, 800
- Nesvadba, N. P. H., et al. 2010, *A&A*, 521, A65
- Nguyen Luong, Q., et al. 2011, *A&A*, 535, A76
- Ogle, P., et al. 2012, *ApJ*, 751, 13
- Papadopoulos, P. P. 2010, *ApJ*, 720, 226
- Pineda, J. L., et al. 2013, *A&A*, in press
- Podio, L., et al. 2013, *ApJLett*, 766, L5
- Rho, J., et al. 2008, *ApJ*, 673, 271
- Riviere-Marichalar, P., et al. 2012a, *A&A*, 546, L8
- Riviere-Marichalar, P., et al. 2012b, *A&A*, 538, L3
- Roussel, H., et al. 2007, *ApJ*, 669, 959
- Sanders, D. B., et al. 1988, *ApJ*, 325, 74
- Saslaw, W. C. & Zipoy, D. 1967, *Nature*, 216, 976
- Schilke, P., et al. 1993, *ApJLett*, 417, L67
- Schneider, N., et al. 2012, *A&A*, 540, L11
- Smith, M. W. L., et al. 2012, *ApJ*, 756, 40
- Spinoglio, L., et al. 2005, *ApJ*, 623, 123
- Spoon, H. W. W., et al. 2007, *ApJLett*, 654, L49
- Starck, J. L., et al. 2010, *Sparse image and signal processing* (Cambridge University Press)
- Storchi-Bergmann, T., et al. 2001, *ApJ*, 559, 147
- Sturm, B., et al. 2013, *A&A*, 553, A5
- Sturm, E., et al. 2011, *ApJLett*, 733, L16
- Swinbank, A. M., et al. 2011, *ApJ*, 742, 11
- Swinbank, A. M., et al. 2010, *Nature*, 464, 733
- Tremblin, P. 2012, *PhD thesis*, AIM Paris-Saclay
- van Dishoeck, E. F., et al. 2011, *PASP*, 123, 138
- Wild, W., et al. 2008, in *SPIE Conference Series*, Vol. 7013
- Wilson, M. E., et al. 2007, in *SPIE Conference Series*, Vol. 6687



White Paper

The Case for an ESA L-Class Mission to Volatile-Rich Asteroids

Spokesperson:

Geraint H. Jones
Mullard Space Science Laboratory, University College London,
Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK

Email: g.h.jones@ucl.ac.uk
Tel: +44 1483 204100



Mission to a Volatile-Rich Asteroid

Contributors and Supporters

Primary contributors marked in red

United Kingdom

Geraint H. Jones, Christopher S. Arridge, Andrew J. Coates, Anne Wellbrock
Mullard Space Science Laboratory, University College London

Neil Bowles, Ian Thomas
Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford

Mark Burchell, Samuel Duddy, Mark Price, **Stephen C. Lowry**
Centre for Astrophysics and Planetary Science, University of Kent, Canterbury

Michele Dougherty
The Blackett Laboratory, Imperial College London

Alan Fitzsimmons,
Queen's University Belfast

Germany

Jessica Agarwal, Hermann Boehnhardt, Martin Hilchenbach, Holger Sierks, **Colin Snodgrass**, Cecilia Tubiana
Max Planck Institut fuer Sonnensystemforschung, Katlenburg-Lindau

R. Srama
University of Stuttgart

M. Paetzold
Universitaet zu Koeln

Mario Tieloff
University of Heidelberg

Olivier Hainaut
European Southern Observatory, Garching

Italy

Monica Lazzarin, Ivano Bertini
Universita' di Padova

Maria Teresa Capria
INAF, IASF Sezione di Roma

Francesco Marzary
Istituto Nazionale di Fisica Nucleare Sezione di Padova

Spain

J. Licandro

IAC

Fernando Moreno Danvilla, Luisa Lara
IAA

Switzerland

Yann Alibert, Kathrin Altwegg, Bernhard Schlaeppli, Andre Bieler
University of Bern

France

Wlodek Kofman, **Alain Herique**
University of Grenoble

Finland

Antti Penttilä, Evgenij Zubko, Karri Muinonen
University of Helsinki

Sweden

Bjorn Davidsson
Uppsala Astronomical Observatory

Netherlands

Aurelie Guilbert-Lepoutre, Sebastien Besse,
ESTEC – Research Fellowship Programme

Belgium

Emmanuël Jehin, Damien Hutsemékers, Jean Manfroid, Cyrielle Opitom, Alice Decock
Institut d'Astrophysique de l'Université de Liège

Israel

Dina Prialnik
Tel Aviv University

USA

Henry Hsieh
University of Hawaii

Carey Lisse
Johns Hopkins University Applied Physics Laboratory

George Hospodarsky
University of Iowa

Amara Graps, M. Davis, T. Greathouse, K. Retherford
Southwest Research Institute

Japan

Adam Masters
ISAS, JAXA

A full list of supporters is available at
<http://bit.ly/volatileasteroids>

1. Executive Summary

Where does Earth's water come from? There are few questions about our planet's history that are more fundamental. Here, we present the case for a future ESA L-class mission that aims to address this question, and key uncertainties about the solar system's history in general, by visiting for the first time at least one volatile-rich body in the asteroid belt. Ideally, this body would be a member of a newly-discovered family of objects in our solar system: Main Belt Comets (MBCs).

Asteroids and comets in general are of interest because they provide a way to probe the protoplanetary disk from which our solar system formed. By determining the chemical and physical properties of various populations of small bodies in our solar system, we can gain insight into the chemical and thermal conditions in different areas of the disk and also investigate the chemical, thermal, collisional, and dynamical processes that have shaped those populations since their formation.

MBCs are perplexing objects that have stable orbits within the asteroid belt, but during certain seasons behave like comets, possessing a dust coma and tail. This strongly suggests that volatiles at their surfaces are sublimating, driving off the dust; this volatile material is likely to be water ice. Dynamical models suggest that bodies from this region could have brought water to Earth, and hence the remaining MBCs hold a frozen record of the source of Earth's water.

MBCs are of particular interest to the planetary science community as they represent the “missing link” between rocky asteroids and icy comets. Although the first MBC was discovered in 1995, the population as a whole was only recognized eight years ago and has since been the subject of high activity in the astronomical community. MBCs' dynamical stability strongly implies that they formed where we observe them today. Their activity also suggests that volatile water ice has been preserved in the subsurface layers since its formation.

Recently, this research has taken on added significance with the revelation that these minor bodies may have been responsible for the delivery to Earth of key volatile materials essential for the development of life, particularly water. Most importantly, MBCs a likely source of water and therefore, ultimately, of life on Earth. This presents a unique opportunity to probe pristine material from a known location in the protoplanetary disk. The proposed mission scenarios would provide powerful constraints to protoplanetary disk models, and answer key questions about how the Earth became the water-rich planet it is today and ultimately suitable for life. The mission would also contribute to a greatly improved understanding of the formation and development of habitable planetary systems around our Sun and elsewhere.

Motivated by a number of recent observational and theoretical developments, we propose an ESA mission to investigate one of the likely sources of exogenous terrestrial water, and in doing so, greatly advance the understanding of our origins.

A mission with this science theme would address several of ESA's Cosmic Vision Themes, in particular “*What are the conditions for planet formation and the emergence of life?*”, and “*How does the solar system work?*”.

2. Science Objectives

Below, we outline the scientific rationale for the proposed mission. We explain what is known about the origin of Earth's water and more generally of volatiles within the inner solar system. We present why a mission to volatile-rich asteroids would address numerous, fundamental questions regarding the early

development of the solar system, as well as helping determine the process through which water and other volatiles were delivered to Earth and to the other terrestrial planets.

2.1 The Origins of Earth's water

2.1.1 Background

One of the major areas of interest in planetary science and astrobiology concerns the origin of water and other volatile material on the Earth (e.g. Izidoro et al. 2013). The composition of planets and small bodies is generally interpreted with respect to the heliocentric distance at which they formed, and in particular by considering whether they formed inside or outside the snow line. The snow line corresponds to the inner boundary of the water ice condensing region: beyond this limit, the solar nebula was cool enough so that volatiles would condense in icy grains, which were then accreted into planetesimals. This process occurs at a temperature in the range from 145 to 170 K depending on the partial pressure of water vapour. In the optically thin solar nebula, the snow line is estimated to be located at 2.7 AU. Before planet formation however, protoplanetary disks are likely to be optically thick. Thus, the snow line could have been further from the Sun due to stronger viscous heating. Furthermore, the snow line can migrate, as the disk evolves. For instance, Davis et al. (2005) showed that the snow line can reach ~ 0.6 AU, which is the minimum radius considered in their calculation.

Given our current understanding of the solar system's formation, the Earth likely formed well inside the snow line, and as a result, should not have accreted a significant amount of water ice and should be largely dry (e.g., Boss 1998). Drake (2005) suggests that "wet accretion" could also be possible, but the more generally-held view of "dry accretion" is corroborated by the distribution of asteroids in the Main Belt, where evidence of hydration is only seen on asteroids outside of 2.5 AU from the Sun (Gradie & Tedesco 1982), well beyond Earth's orbit. This suggests that the Earth accreted from mostly anhydrous planetesimals, and should therefore be anhydrous itself. Since the present-day Earth is not dry, we are left to determine the source of our current water.

The prevailing hypothesis for the origin of Earth's water involves the post-accretion bombardment of the young Earth by volatile-rich planetesimals from beyond the snow line, i.e., either comets from the outer solar system or objects from the main asteroid belt (e.g., Anders & Owen 1977; Owen & Bar-Nun 1995), from hereon referred to as the Main Belt. This is referred to as the "late veneer" model. The large ice content of comets from the outer solar system makes them natural candidates for delivering water to the Earth. However, as we now know, asteroids from the Main Belt also contain ice and therefore could have also delivered water to the Earth.

Planet migration may have been the cause for increased impact rates on Earth of debris bodies from the planetary disk and it has been shown that material from the asteroid belt may have had a significant if not dominant share (Gomes et al. 2005). Dynamically, it is easier to deliver material from the asteroid belt to Earth than from the outer solar system.

Although most models indicate that the Earth formed dry, our planet is clearly not. For decades, the source of volatiles on Earth has been a matter of great debate. This topic is not only important in order to understand the origin of life on our planet, but also because it can hold crucial clues on the early evolution of the solar system.

Volatiles could have been accreted during the Earth's formation. Based on dynamical models, Morbidelli et al. (2000) indeed proposed that the Earth accreted "wet": only a few impacts of C-type asteroids would have been sufficient to import the oceanic mass on Earth. Alternatively, volatiles could have been accreted at a latter stage of Earth's evolution. Indeed, Javoy et al. (2010) showed that enstatite chondrites are the only chondrite group isotopically identical to Earth's bulk composition. This implies that the proto-Earth was dry, and that volatiles were delivered later in the evolution of Earth, through the accretion of volatile-rich asteroids and comets (Owen & Bar-Nun 1995). Regardless of the

scenario we consider for delivering volatiles on Earth, it is clear that the composition of our planet is governed by the volatile inventory of planetesimals which were accreted by Earth. The oceans hold about 97% of the surface water on Earth. About the same amount of water may be stored in the surface crust. Hence, the total volume of water on Earth may be between 1.55 and $3.1 \cdot 10^9 \text{ km}^3$. Although small in relation to the total volume of the planet ($1.1 \cdot 10^{12} \text{ km}^3$) it fills a sphere of about 700 to 900 km radius which is in the range of the larger representative of minor bodies in the planetary system (radius of 1 Ceres is about 470km, that of Pluto is about 1200km). Liquid water may have existed on Earth just 200 Million years after the planet's formation (Mojzsis et al., 2001).

2.1.3 Water Delivery Scenarios

There are three primary scenarios for the origin of water on Earth:

A) Endogenic: In this case, the water was not brought to Earth; it was already here, as part of the nebular gas, either (a) adsorbed inside the dusty building blocks that agglomerated to form our planet, an idea known as “wet accretion,” (Muralidharan et al. 2008; Stimpfl et al. 2006) or else, b) acquired after the Moon-forming event, an idea known as the “nebular origin” model (Genda & Ikoma 2008; Ikoma & Genda 2006). In the first case, 1–3 Earth oceans of water could be adsorbed by associative adsorption onto perfect forsterite surfaces, while later studies of imperfect forsterite suggest more Earth oceans of water were adsorbed prior to planetary accretion onto the building blocks of terrestrial planets (Muralidharan et al. 2008). Two remaining questions are whether this model can account for Earth's iron core and whether the chronology of the accretion and nebula gas dispersal can explain noble gas fractionation values.

In the nebular origin model, Earth attracted the surrounding nebula gas to have a massive ($>10^{21} \text{ kg}$) hydrogen-rich atmosphere. At the time, our planet was covered in a magma ocean containing oxides such as FeO that reacted with the hydrogen to produce water. The surrounding nebular gas then began to disappear, the dense atmosphere and proto-Earth cooled, and 1-100 ocean masses of liquid water formed through the condensation of steam in the atmosphere (Genda 2009). When the nebular gas dissipated enough to permit the extreme-ultraviolet and far-UV radiation from the Sun to penetrate the Earth's atmosphere, the lightest gases, hydrogen and helium, escaped into space.

B) Exogenic (a “stochastic, big splashes model”) (O'Brien et al. 2006; Raymond et al. 2004; Morbidelli et al. 2000): The water came from agglomerated bodies, ranging in size from planetesimals (size $\sim 1 \text{ km}$) to embryos (size $\sim 1000 \text{ km}$, lunar-sized), on solar orbits inside Jupiter's orbit. One, or, more likely, a few, large embryos arrived from the Main Belt (parent bodies). This chance occurrence, after Earth had formed but before the Moon-forming impact, within the first few tens of millions of years (My) of solar system evolution, brought the water and volatiles to Earth in a few large splashes. This terrestrial planet formation scenario works best after 10 My, by which point the nebular gas has dissipated. The amount of water accreted by their modelled terrestrial planets, in the 1.0-1.4 AU region, ranges from 1-10 Earth oceans (O'Brien et al. 2006).

C) Endogenic + Exogenic (a “dual-origin model”) (Dauphas 2003): The proto-Earth picked up water from nebular gas or planetesimals, but then the lighter gases hydrodynamically. After the nebular gas dissipated, but before Earth's last major impact (the moon-forming event), a few impacts of (cold) comets travelling from the *outer* solar system supplemented Earth's noble gas inventory.

Estimates for the contribution to Earth's oceans from Oort cloud comets range from less than 10% (Robert et al. 2000) to 50% (King et al. 2010). Theoretical arguments rule them out as the only source of Earth's water because many more comets would exist in the outer parts of the solar system than is currently thought possible.

If not Oort Cloud comets, what brought the water to the Earth? Jupiter family comets are a possibility and indeed, the D/H ratio of 103P/Hartley 2, a Jupiter family comet, was found to match that of

terrestrial ocean water (Hartogh et al. 2011). However, as discussed above, certain Main Belt asteroids could also contain substantial amounts of ice and therefore should also be considered.

Thermal vacuum chamber experiments by the KOSI consortium suggest that isotopic enrichment may occur in the upper layers of an active water sublimating surface (Roessler et al. 1992). This makes the interpretation of observational results at active comets rather difficult, since it may mean that the primordial D/H ratio would be smaller than measured while the comets are active.

2.2 Volatiles in the Inner Solar System.

As well as their importance to determining the origin of Earth's water, the fraction and composition of ice in solar system bodies are of interest due to the temperature constraints that they provide. The exact location of the snow line in our protoplanetary disk was dependent on various poorly constrained conditions including opacity, mass density, and accretion rate in the disk. As the disk evolved, the snow line is thought to have shifted as planetesimal accretion progressed. Bodies formed in the outer solar system between Jupiter and Neptune (the original accretion zone of current Oort Cloud objects) and beyond (the Kuiper Belt), are well beyond the snow line and thus are certainly icy. Closer to the Sun, the situation is less certain. Observations of asteroids suggest that the snow line probably existed around 2.5 AU from the Sun (Gradie & Tedesco 1982; Jones et al. 1990), but theoretical studies (e.g., Sassolov & Lecar 2000) have placed it possibly closer than the orbit of Mars. If true, objects throughout the Main Belt could have incorporated water ice at the time of their formation.

Since their formation, asteroids in the Main Belt have undergone substantial heating from solar radiation, the radioactive decay of ^{26}Al , and perhaps electromagnetic induction from the solar wind that long ago caused that primordial water ice to liquefy and drive hydration reactions within the parent bodies (Grimm & McSween 1989; Cohen & Coker 2000; Mousis & Alibert 2005). Spectroscopic evidence of these reactions exists in the form of absorption features in the infrared portion of the spectrum at $3\mu\text{m}$ (e.g., Lebofsky 1980; Lebofsky et al. 1981) and in the visible portion of the spectrum at 0.7 μm form of absorption features Barucci et al. 1998). Hydrated minerals have also been found in CI and CM carbonaceous chondrite meteorites that have been determined to have originated from C-type asteroids in the Main Belt (e.g., Hiroi et al. 1996).

Given the ubiquity of primordial water implied by spectroscopic and meteoritic evidence, Jones et al. (1990) argued that a decline in hydrated silicate detections on asteroids with increasing semimajor axis could indicate not a lack of ice, but insufficient heating to melt that ice and to drive hydration reactions. That ice could still exist today. By combining HST observations of the shape of asteroid 1 Ceres, the largest object in the asteroid belt, with other physical properties such as its rotation rate and bulk density, Thomas et al. (2005) demonstrated that Ceres is most likely a differentiated body consisting of a dense core of rocky material surrounded by a water-ice mantle comprising 16-26% of its total mass.

More recently, water ice has been directly detected on the surface of an asteroid (Rivkin et al. 2010; Campins et al. 2010). Centred at $3.1\mu\text{m}$, an absorption band unlike those caused by hydrated minerals in carbonaceous chondrites or C-type asteroids was observed in the near-IR spectrum of asteroid 24 Themis. Modelling of the absorption band indicated that it was likely due to fine-grained water ice and organic material, although this has been disputed (Beck et al. 2011). The absence of hydrated minerals on the surface of 24 Themis suggests that there is a large continued presence of ice within the asteroid that could survive longer than the age of the solar system (Campins et al. 2010). Exposure of this ice could be explained by impact events overturning the surface. A similar process occurs on the Moon at a rate of 1m/Gyr (Heiken et al. 1991). Alternatively, the slow sublimation of water ice could cause escaping water to recondense on the asteroid surface as a frost.

Hydrated minerals such as phyllosilicates have been identified both at the surface of asteroids and in meteorites studied in the laboratory. These hydrated minerals can contain structurally bound OH or H_2O (some carbonaceous chondrites can contain up to 12wt% in water), which show a characteristic

spectroscopic band near $3 \mu\text{m}$. The fraction of asteroids showing the feature varies with heliocentric distance (see Takir & Emery 2012 for a review). The presence of phyllosilicates at the surface of Main Belt asteroids can be understood with respect to the early radiogenic heating of these objects (Grimm & McSween 1993), used to explain the compositional zoning of objects in the Main Belt.

Inside 2.4 AU, asteroids might have formed very quickly, and the internal heating could have been so intense that the resulting material would be completely devolatilized. No phyllosilicates can be observed on those asteroids because they would be destroyed.

Between 2.4 and 3.4 AU, asteroids might have formed rapidly enough so that the internal temperature would rise above the water melting point, but slow enough so that the material would not be devolatilized. The chemical reactions between liquid water and rock produced phyllosilicates, which are indeed observed both on asteroids and in meteorites. In addition, fluid inclusions have been reported in the Monahans and Zag chondrites (Zolensky et al. 1999, Rubin et al. 2002). The 4.5 Gyr old brine contained in these inclusions is a sample of the fluid presumably responsible for the aqueous alteration of early solar system material.

Beyond 3.4 AU, asteroids would have formed too slowly. The heating would not have been sufficient to produce any liquid water, much less hydrated minerals.

Some water ice spectral features have been reported for several asteroids. Themis and Cybele for example, exhibit a spectral feature around $3 \mu\text{m}$, which has been attributed to the presence of water ice at their surface (Rivkin & Emery, 2010, Campins et al. 2010, Licandro et al. 2011a). Jewitt & Guilbert-Lepoutre (2012) suggested a scenario in which repeated impacts could steadily bring buried ice at their surface (provided water ice could have survived up to now). The spectrum of Ceres also displays this spectral feature, which has tentatively been attributed to a very thin water ice frost (Lebofsky et al. 1981), NH_4 -bearing phyllosilicates (King et al. 1992), a mixture of organics and crystalline water ice (Vernazza et al. 2005), or iron-rich clays (Rivkin et al. 2006). Ceres has however a very low density ($2.077 \pm 0.036 \text{ g/cc}$), which is compatible with a large fraction of internal water ice. Furthermore, hydroxyl emissions, produced by photodissociation of H_2O (A'Hearn & Feldman, 1992) have been observed near the polar region of Ceres, and were interpreted as an evidence of the presence of internal water ice close to the surface. Thermal models of Ceres are also compatible with mantles containing ~15 to 25 wt% of water ice (Fanale & Savail 1989, McCord & Sotin, 2005).

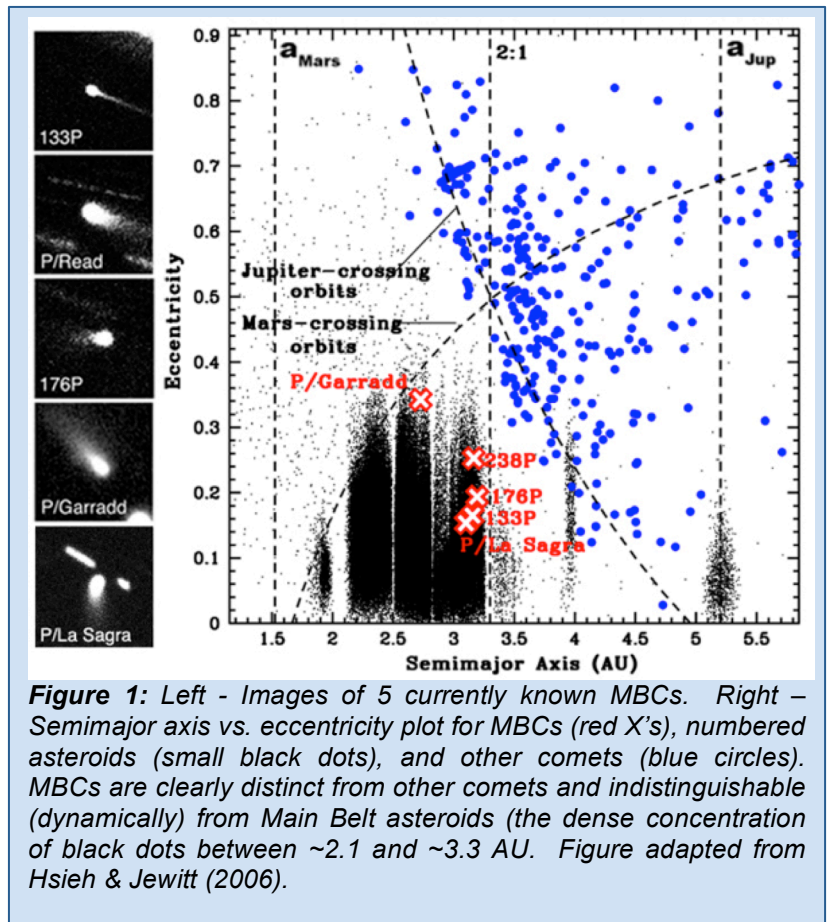


Figure 1: Left - Images of 5 currently known MBCs. Right - Semimajor axis vs. eccentricity plot for MBCs (red X's), numbered asteroids (small black dots), and other comets (blue circles). MBCs are clearly distinct from other comets and indistinguishable (dynamically) from Main Belt asteroids (the dense concentration of black dots between ~2.1 and ~3.3 AU). Figure adapted from Hsieh & Jewitt (2006).

2.3 Main Belt Comets: A new class of Solar System object

2.3.1 Discovery

A strong indication of the existence of present-day ice in Main Belt asteroids came with the recognition of the new cometary class of MBCs (Hsieh & Jewitt 2006). These objects occupy stable orbits in the main asteroid belt indistinguishable from those of other Main Belt asteroids, yet exhibit cometary mass loss in the form of dust comae and tails, indicative of the sublimation of volatile ices (Jewitt 2012). MBCs differ from other comets in that their stable orbits indicate they are most likely native to the inner solar system, whereas other comets originate in the colder and icier outer solar system.

Motivated by their study of comet 133P/Elst-Pizarro, the first known MBC (Hsieh et al. 2004), Hsieh and Jewitt set out to try to explain its strange comet-like physical behaviour but asteroidal dynamical properties by putting the two lead hypotheses for 133P's origin to the test. If 133P was a "lost comet", i.e., an ordinary comet from the outer solar system that had somehow dynamically evolved onto a Main Belt orbit, it could be unique in the asteroid belt, given how difficult such a dynamical transition is considered to be. If, on the other hand, 133P was an icy asteroid that was native to the Main Belt, more 133P-like cometary asteroids would be expected to exist, since if 133P was an otherwise ordinary asteroid that happened to exhibit cometary activity, other otherwise ordinary asteroids could do so as well.

In 2005, whilst conducting a targeted survey of the outer Main Belt, Hsieh and Jewitt, discovered another active comet, 176P/LINEAR. Together with the serendipitous discovery of another comet in the asteroid belt, 238P/Read, 133P and 176P constituted a new class of cometary objects, dubbed MBCs (**Figure 1**; Hsieh & Jewitt 2006, Hsieh 2009). The small scale of their survey (just 600 objects observed, out of more than 400,000 Main Belt objects currently known) further suggested that the total size of the MBC population could be much larger, possibly in the hundreds (Hsieh 2009). While fulfilling a key prediction of the "icy asteroid" hypothesis (that 133P would not be unique), the discovery of multiple objects in the asteroid belt exhibiting cometary behaviour rendered the "lost comet" hypothesis extremely unlikely, given the vanishingly low likelihood of *multiple* outer-solar system comets transitioning onto Main Belt orbits. Two more MBCs have been found since 2006, validating the prediction that the first three MBCs represent a small sample of a larger population.

Also the only known reflectance spectra of MBCs, that of 133P, 174P and (300163) 2006 VW139, are different from the spectra of comets (Licandro et al. 2011a, Licandro et al. 2013). Spectra of MBCs are typical of asteroids with similar semi-major axis and corresponds with primitive B- and C-type asteroids. The link between MBCs and primitive asteroids needs to be further explored.

In addition, some asteroids – MBCs - are ejecting dust, producing comet-like activity features. These have all the dynamical and spectroscopical characteristics of asteroids but presents the coma and tail that are characteristic of comets. They appear to be a third distinct reservoir of water ice in the solar system, in addition to the Kuiper Belt and the Oort Cloud. Although different processes could account for mass loss among asteroids, including impact ejection, rotational instability, electrostatic forces, thermal fracture, thermal dehydration, shock dehydration, radiation pressure sweeping, some objects have shown a repetitive activity close to perihelion, which strongly points to ice sublimation (Jewitt 2012). Thermal evolution models have shown that water ice could survive in shallow subsurface layers, either very close to the surface (Schorghofer 2008), or ~50 to 150m deep (Priainik & Rosenberg 2009) in these active asteroids.

2.3.2 The Thermal Evolution of MBCs

Dust ejection due to sublimation implies the existence of both ice and dust particles. This leads to imagine that MBCs are comet-like bodies, and so the thermal modeling codes used to simulate and study the activity of comets, KBOs and icy bodies in general can be also applied to this class of objects. Schorghofer (2008) assumes that the ice, in order to survive for such a long time in spherical bodies at

heliocentric distances of 2-3.2 AU, must have been buried under an insulating mantle, and gives estimations of the thickness of this dust layer. Studies of the survivability of ice in the Main Belt suggest that water ice can survive in subsurface layers. The models indicate that ice can persist in the top few metres of a dusty asteroid surface, if the surface temperature remains below 145K. Although this favours ice retention on slowly rotating asteroids ice can still survive at high latitudes on fast rotating asteroids. The same author argues against the possibility that the ice be mixed with rocky material rather than dusty material. Rocky surfaces are seldom able to retain ice, due to the larger thermal conductivity and the larger molecular free path of rocks with respect to dust grains.

Prialnik & Rosenberg (2009) applied their comet nuclei thermal evolution model to 133P/Elst-Pizarro. They demonstrate that deep-buried ice could have survived since the Main Belt formation time, and that this ice is most probably composed, at least nowadays, of crystalline water ice. It would have been almost impossible for other ices, sublimating at temperatures well under 130 K to survive even under an insulating mantle. Modelling of the survival of subsurface ice within the snow-line indicates that only water ice will survive, and more volatile ices will be lost (Prialnik & Rosenberg 2009; Capria et al. 2012). The parent ice species for CN in normal comets is not certain, but a strong candidate is HCN, which is considerably more volatile than water (with a sublimation point of 95K), and is not expected to survive in MBCs.

Capria et al. (2011), making the assumption that the MBCs were icy bodies on which a stable cometary activity has been induced by a recent impact that excavated a crater in a dust mantle, studied the possible duration of this activity. The authors demonstrated that, given the existence of buried ice, the activity observed in MBCs can be explained as cometary activity, that means dust emission sustained by ice sublimation. They also estimated the time for having a collision with a projectile ≥ 8 m in size on a body with the size of Elst-Pizarro (~ 5 km), and found that impacts occur every ~ 10 yr.

Dust at the surface of asteroids in the Main Belt can act as an insulating layer protecting ice possibly present in their interiors. Two major mechanisms can produce such a dust mantle. Irradiation by cosmic rays, hard UVs and other energetic particles lead to the formation of a refractory layer up to a few meters thick (Strazzulla et al. 1991, Baratta et al. 1994). Progressive devolatilization induced by cometary activity can also produce a very porous dusty crust, formed by particles too heavy relative to their cross-section to be entrained by gas outflow. Although the properties of such a rubble mantle at the surface of a small body are not very well constrained, it is assumed that it should be very porous, with a very low thermal inertia. With this idea in mind, Schorghofer (2008) introduced the concept of a buried snow line: the rubble mantle could be such a poor heat conductor that the loss of ice is sufficiently slow over the age of the solar system. Ice could then survive within the top few meters of the surface, under the top dusty surface layer, over billions of years. By accounting for gas flow through the porous nucleus and sublimation in the deep porous interior, Prialnik & Rosenberg (2009) suggested that this depth should rather be between 50 and 150 m.

The repeated activity observed for some MBCs strongly points toward a comet-like activity produced by the sublimation of water ice. Given that ice could have steadily survived buried in the interiors of these objects, some orbital or physical change must trigger the outgassing. Both Prialnik & Rosenberg (2009) and Capria et al. (2012) suggested that water ice could be exposed at the surface by a recent impact. Impact rates in the Main Belt are indeed compatible with this scenario. It is worth noting that Jewitt & Guilbert-Lepoutre (2012) also invoked a collision scenario to explain the presence of water ice at the surface of asteroids (24) Themis and (65) Cybele. In order to test this hypothesis, in situ measurements of the thickness of this inferred rubble mantle, or more generally of the internal structure of MBCs, should bring valuable constraints. Marzari et al. (2011) suggest that small asteroids in the Main Belt are accelerated beyond the breakup limit. The consequent rotational fission may explain the formation of a dust tail like that of some MBCs. This scenario is an alternative, and possibly a more frequent one, to the collisional model.

The interior structure of MBCs could hold important evidence for the nature of the bodies' surroundings at the time of formation. The distribution of subsurface ice; whether there exist lenses of ice in a rocky

body, deep icy core mask by a thick crust, and the global ice / rock ratio would be key issues to address. The deep structure of the body is a consequence of the accretion mechanisms and is a parameter that allows insights into the mechanical, thermodynamical and compositional condition in the protoplanetary disk. The data from an object likely formed in the MBC accretion zone would complement the interior structure observations made by Rosetta at Comet 67P.

2.3.3 Observational evidence of volatiles at MBCs

The question of whether or not MBCs contain ice has been convincingly settled for 133P, as it has returned to activity after each perihelion passage since its discovery, meaning that sublimation of ice is the only reasonable explanation (Hsieh et al 2010). Other recent discoveries confuse the picture: While 238P/Read also shows repeated activity and is likely a bona-fide comet (Hsieh et al 2011), other objects with comet-like appearance have been shown to be due to collisions or rotational break up (e.g. Snodgrass et al 2010, Jewitt et al 2010, Moreno et al. 2011, Moreno et al. 2012, Stevenson et al 2012). Modelling of the dust morphology can be used to differentiate between tails from comet-like activity, which has lasted for many months, and trails of debris from single events. Repeated activity remains the best evidence for the comet-like nature of some of these objects. Ideally we would like direct confirmation that sublimating ice drives their activity – this requires detection of a gas coma.

For normal comets spectroscopy reveals the presence of the gas and allows it to be identified. Water is found to be the main constituent of comets, which is photodissociated into OH and H; a strong signature of emission by OH at 308 nm is seen in comet optical spectra (water itself is very difficult to detect from the ground due to Earth's atmosphere). The next strongest feature in the optical is the group of CN band at 389 nm, which is far easier to observe, as the OH band is strongly affected by terrestrial atmospheric absorption, so it is difficult to detect from the ground. For this reason, ground based spectroscopy of MBCs to date has concentrated on the CN band, but has proven unsuccessful (e.g. Jewitt et al 2009, Licandro et al 2011b, Hsieh et al 2011). The upper limits on water production resulting from these works (calculated using an assumed 'typical' CN:H₂O ratio based on traditional comets) show that MBCs are very weakly active comets, in agreement with the low dust production rates inferred from their faint tails. There is a potential problem with assessing water production via CN emission band, as the underlying assumption, that MBCs have the same proportions of volatile ices to other comets, is likely to be incorrect (see next section on survival of ice in MBCs).

Two attempts were made to directly detect water around MBC candidates using the ESA Herschel space telescope. This telescope operates at thermal infrared wavelengths, and is sensitive to water emission at 557 GHz. Unfortunately, in the first case 176P/LINEAR did not return to activity when expected (the observations were scheduled to coincide with the same near-perihelion point in the orbit that activity was seen at in 2005), and no water was detected (de Val-Borro et al 2012). The upper limit on water production rate from these observations, $Q(\text{H}_2\text{O}) < 4 \times 10^{25}$ molecules s⁻¹, is more sensitive than most of the limits from CN line observations, and would have been sensitive enough to detect water if MBC production rates follow the same empirical relationship to total brightness found for other comets (Jorda et al 2008). A second Herschel observation of an MBC, P/2012 T1 (PANSTARRS) was made in January 2013. The comet was clearly active in optical images obtained at the same time, but again there was no detection, implying a lower than expected water production (O'Rourke et al., in prep.). An attempt to observe the water around this comet via the OH emission line at 308nm from the ground (using the VLT and the X-SHOOTER spectrograph) also produced only an upper limit (Snodgrass et al., in prep.). The very low water production rate that must be responsible for MBC activity supports the idea that they show activity from only a small patch of their surface, which has recently been excavated to reveal previously buried ice.

3 Measurements to be made by the mission

3.1 Measurement of the D/H ratio

3.1.1 Rationale

Measurement of the deuterium-to-hydrogen (D/H) ratio has the potential for constraining the origin of volatiles on Earth. Deuterium was synthesized during the Big Bang (Wagoner et al 1967), and there is no known mechanism that produces significant amounts of D in galaxies or stars (Epstein et al 1976). The D/H ratio in water is very sensitive to the conditions in which it was formed, in particular the kinetic temperature of the medium. In the solar nebula, isotopic exchange reactions between H_2 and HDO molecules would have led to a gradual reduction of the D/H ratio in water. The efficiency of these reactions depends on the turbulent mixing in the solar nebula, which is correlated with gas density and temperature. The D/H ratio in water is thus predicted to increase with heliocentric distance (see Robert 2006 for a review). Ices accreted in planetesimals would have then preserved the D/H ratios from this early epoch, and current small body populations are expected to exhibit different D/H ratios depending on their formation heliocentric distance.

The Earth's oceans provide an interesting and paradoxical case of enhancement compared to the solar nebula value. The estimated D/H ratio value in Standard Mean Ocean Water (SMOW) (Balsiger et al. 1995) is 1.6×10^{-4} , about 6-7 times the value from the primitive Sun, deduced from the solar wind implanted onto lunar soils or from the atmosphere of Jupiter (Robert et al. 2000, 2001). In-situ measurements of the D/H ratio were obtained for Comet 1P/Halley by the Giotto space probe (Balsiger et al. 1995; Eberhardt et al. 1995) and more recently from material returned to Earth by the Stardust mission to Comet 81P/Wild 2 (McKeegan et al. 2006). The D/H ratio of Comet Halley was found to

be significantly higher than that of the terrestrial value measured from sea water. In contrast the D/H ratio for Comet Wild 2 was determined to cover a wide range, from terrestrial to Halley-like values. Radio-spectroscopy was used to measure D/H in Comets C/1995 O1 (Hale-Bopp) and C/1996 B2 (Hyakutake) during their near-Earth passes (Meier et al. 1998; Bockelee-Morvan et al. 1998). In both cases the D/H ratio was found to be approximately twice that of the terrestrial value.

More recently, D/H was measured in comet C/2002 T7 (LINEAR) from high resolution spectra of OH and OD in the near UV using UVES at the ESO VLT (Hutsemékers et al. 2008). This ratio was also determined in C/2001 Q4 (NEAT) using the Space Telescope Imaging Spectrograph on HST (Weaver et al. 2008). This was achieved by measuring the flux of emission lines produced by D and H. Finally, the Cryogenic Infrared Echelle Spectrograph (CRIRES) at the VLT was used to observe the spectrum of Halley-family comet 8P/Tuttle (Villanueva et al. 2009). The D/H ratio was measured by modelling the emission spectrum of HDO and extracting the flux at the expected wavelengths. In each case, the D/H ratio was comparable to previously measured values for comets.

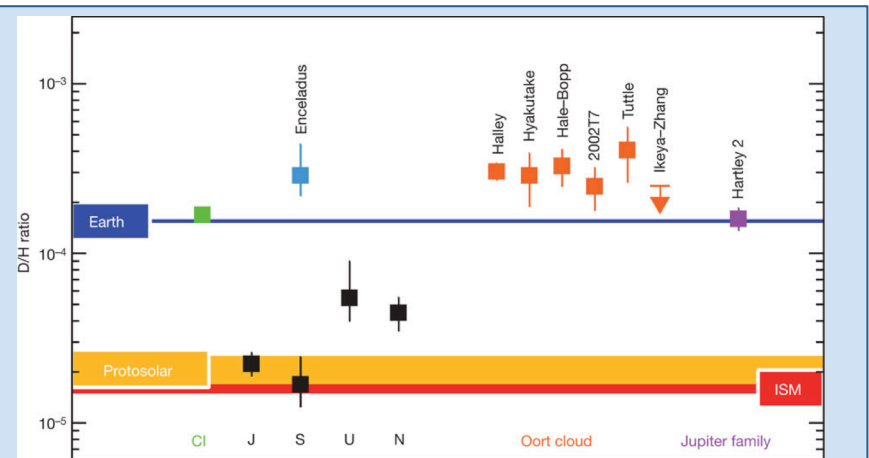


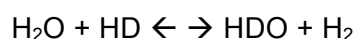
Figure 2: Summary of relevant D/H measurements, from Hartogh et al. (2011). The D/H determinations in comets originating from the Oort cloud are twice the value for the Earth's ocean (blue line) and ~10 times larger than the protosolar value in H_2 (broad yellow line). The D/H ratio in the Jupiter-family comet 103P/Hartley 2 matches Earth's ocean value and the chondritic CI value.

The SMOW value is very similar to the D/H ratio measured in carbonaceous chondrites¹ originating from beyond 2.5AU ($D/H = 1.4 \pm 0.1 \times 10^{-4}$). The measurements in Oort Cloud comets result in $D/H = 2.96 \pm 0.25 \times 10^{-4}$ on average. This has led to models in which asteroids must have been the major suppliers of volatiles on Earth, with less than 10% of water being delivered by comets. However, the recent measurement in Jupiter family comet (JFC) 103P/Hartley 2, with $D/H = 1.61 \pm 0.24 \times 10^{-4}$ (Hartogh et al. 2011; **Figure 2**), has raised many new questions and the need of re-addressing these models. Furthermore, this result is incompatible with the scenario of small body formation via the Nice model (Tsiganis et al. 2005, Gomes et al. 2005, Morbidelli et al. 2005), which suggest that Oort Cloud comets might have been formed closer to the Sun than JFCs. Resolving this issue include two different studies:

- Models of the D/H gradient in the solar system, which are mostly predictions given the scarcity of accurate isotopic measurements, may need to be revisited. For instance, the Grand Tack model (Walsh et al. 2011) suggests a large-scale mixing of material between the inner and outer solar system. Therefore, a significant fraction of Main Belt asteroids might have originated in the primordial Kuiper Belt, with different D/H ratios. The measurement of the D/H ratio on MBCs, which are a key population to disentangle the different models of the solar system dynamical evolution, would provide critical data in this regard.

- Processes yet to be determined may also be occurring during the history of small bodies. Indeed, by recalculating the D/H ratio in comet 1P/Halley at different positions in the coma, Brown et al. (2012) revealed a gradient with the distance from the nucleus. Depending on environmental conditions, the coma can be enriched, depleted, or equivalent in D/H ratio relative to the bulk composition of the comet. It is therefore highly unlikely that D/H ratios observed in comae are representative of the bulk D/H ratio in their nuclei. Consequently, in situ measurements on the bulk of small bodies could again provide crucial data.

The deuterium to hydrogen (D/H) isotopic ratio provides a chemical signature of the source region of water. Hydrogen and deuterium (hydrogen which consists of a proton and neutron) were created during the Big Bang. The second largest reservoir for deuterium in the solar system is water. Substitution of one hydrogen atom in water results in deuterated water, HDO and proceeds by the reversible reaction



In the early solar nebula, this reaction occurred while water was still in the vapour phase, leading to a reduction in HDO. After condensation and formation of ice grains, the reaction halts. Therefore, the reaction slows and halts more rapidly at larger heliocentric distances. This leads to an enrichment of the D/H ratio at larger heliocentric distances (Mousis et al. 2000; Horner et al. 2007). The ice grains continue to grow and are incorporated into the planetesimals of the early solar system. The D/H ratio of the region in which the planetesimals formed is essentially frozen into the body at the time of water ice formation. Current models indicate that the D/H ratio shows very little variation below 5AU, before rapidly increasing with heliocentric distance to 45 AU; beyond this distance the variation becomes very small.

Coupled with the inferred and observed existence of ice in the Main Belt, the MBCs present themselves as excellent candidates to probe the nature of water ice in the Main Belt and thus determine the exogenic origin of Earth's water. By measuring the D/H ratio we will be able to validate or dismiss this hypothesis. Although telescope observations are being used increasingly to measure the D/H ratio, it is only possible when the targets are exceptionally bright. Measurement of this ratio for MBCs via telescopic observations is impossible due to their faintness.

¹ The D/H ratio of hydrated minerals can be used to determine the D/H ratio of water ice accreted in asteroids, because the isotopic fractionation between hydrated minerals and liquid water is negligibly small at the scale of the isotopic variations in the solar system (Robert 2006).

3.1.2 Interpretation of Possible Findings

Measuring D/H at an MBC will allow us to constrain its formation to the Main Belt. Although this is highly likely given known MBCs' current physical and dynamical properties, recent dynamical models by Levison et al. (2009) have suggested that material from the Kuiper Belt could have been captured in stable orbits within the asteroid belt during the planetary migration phase of the early solar system. If D/H in MBCs were found to be comparable to that measured for comets, it would corroborate that these dynamical models are accurate and cast doubt on the MBCs as a dominant source of Earth's water.

We may find that the D/H ratio is comparable to or greater than that of SMOW and smaller than that of the comets. In the case that it is comparable, we will have demonstrated that current models of D/H variation with heliocentric distance are correct and provided constraints on the chemistry of the proto-planetary disk. On the other hand, if the D/H ratio is greater than SMOW but less than comets, it would suggest that MBCs did not make a significant contribution to Earth's water, leaving only MBCs as the last potential exogenic sources.

Since the impact of comets would have raised the D/H ratio on Earth, it could be expected that D/H measured in the MBCs would be lower than SMOW. Coupled with noble gas measurements at the target, we would be able to determine the likelihood that the MBCs are the sources of Earth's water. We would also be able to constrain the cometary contribution further.

In each case, the data returned by the mission will allow us to constrain where in the solar system the MBCs formed and the subsequent dynamical evolution of the solar system, the chemical and thermal properties of the protoplanetary disk, and most importantly, we will determine the likely delivery method of water to Earth.

Independent of the terrestrial water issue, the measurement of the D/H ratio of an MBC as a member of the asteroid belt would impose important constraints on the primordial snow line and the enhancement relation of D/H versus distance to the Sun. This will allow the formation scenario to be tightly constrained.

3.2 Other measurements

Below is listed an example strawman payload for a mission to characterize an MBC. This assumes that the spacecraft would rendezvous with the target object. Requirements would differ in the case of a flyby mission.

Instrument	Scientific observations	Comments
Imager	Body morphology, shape, density, craters, active regions, reflectance properties, possible coma structures.	Strong heritage
Mass spectrometer	Instrument capable of determining D/H ratio as well as those of other key volatile species.	Strong heritage
Mid-IR spectroscopy	Compositional maps to complement those in visible range. Physics nature and structure of near-surface layers.	Existing heritage, and high TRL for new variations on instrumentation.
Dust instrument	Mass range and velocity of dust grains, plus compositional information.	Strong heritage, but should be noted that unless a flyby, spacecraft will essentially be at rest with respect to MBC

Mission to a Volatile-Rich Asteroid

		and its dust coma and tail. Instrument cannot therefore rely on ram speed of particles to determine particle parameters.
Radar subsurface sonder	Internal structure of MBC	Heritage from Rosetta, Curiosity.
Plasma instrument	Energies and directions of positive and negative ions, and electrons.	Strong heritage.
Magnetometer	Magnetic field strength and direction. Measures nature of interaction with solar wind as well as searching for an internal magnetic field.	Strong heritage.

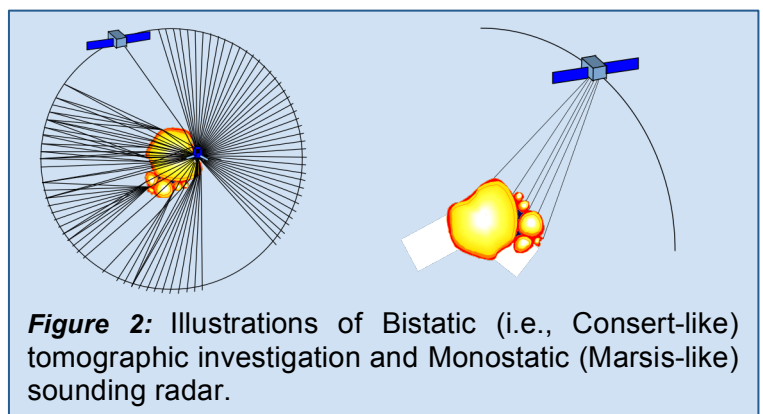
The key measurement of the D/H ratio would be primarily achieved by a mass spectrometer, for which the instrument heritage is strong, such as the ROSINA instrument aboard Rosetta. Estimates of the maximum gas production rates of active MBCs are in the range of 10^{22} - 10^{24} molecules per second. This is considerably lower than comets that have been encountered to date, however, estimates indicate that the flux of emitted neutrals is clearly sufficient to allow the determination of D/H purely from sampling the gas coma from a spacecraft near the MBC. Maximised exposure to the coma during the spacecraft rendezvous period would of course increase the signal to noise ratio of the result, but it is clear that this key measurement is within the range of existing technologies. Complementary measurements could be made using an UV spectrometer, however, given the large heliocentric distance, such a measurement may be challenging. The most precise isotope ratio measurements would be achieved if a landing on the target object is possible. However, this would most likely require a targeted landing at an active region of potentially very limited area, thus raising the complexity of the mission.

Plasma instrumentation could potentially provide an independent measurement of the activity level of the MBC. In situ observations of ions near Saturn's moons have detected activity at levels lower than that anticipated for MBCs (Teolis et al. 2010).

Defining the internal structure from submetric scale to global scale:

Characterizing interior of an MBC, its bulk ice to dust ration, its bulk porosity and its internal structures (cometesimals, layering), is one way to understand how MBCs are the missing links between comets and asteroids. It could address questions such as: Are they icy cores insulated by a thick crust or rocky bodies with a few ice lenses? How their internal structures differ from 67P/Churyumov-Gerasimenko as it will be fathomed by Consert/Rosetta? (Kofman, 2007)

The deep interior structure of the nucleus reveals the accretion mechanism: the internal variation of the porosity and of the composition, the presence and the size of cometesimal, the ice distribution are a unique way to access to the mechanical, thermodynamical and compositional condition in the region of the protoplanetary disk within which MBCs formed.



It is also a key point to model thermal activity and activation mechanisms of such bodies: Because of its expected high porosity, the dust is thermally insulating and so thermal modelling requires estimation of porosity and ice distribution. The determination of such structure is especially important in the crust (near subsurface - first tens meters) in order to identify lateral variation of porosity / composition, layering, presence of ice lens, voids or channels. So, high resolution characterization of the crust allows the reconnection of the surface which will be well characterized by other remote sensing instruments to the deep interior.

Radar is the only technique capable of achieving this science objective of characterizing the internal structure and heterogeneity of MBCs:

- The deep interior structure tomography requires low-frequency radar to penetrate throughout the complete body. The radar wave propagation delay and the received power are related to the complex dielectric permittivity (i.e the composition and microporosity) and the small scale heterogeneities (scattering losses) while the spatial variation of the signal and the multiple paths provide information on the presence of heterogeneities (variations in composition or porosity), layers, ice lens. A partial coverage will provide "cuts" of the body when a dense coverage will allow a complete tomography. Two instruments concepts can be envisaged: A monostatic radar like Marsis/Mars Express (ESA) (Picardi, 2005) operating at ~10MHz with ~2 MHz Bandwidth that will analyze radar waves transmitted by the orbiter and received after reflection by the MBC, its surface and its internal structures. A bistatic radar like Consert/Rosetta (ESA) (Kofman, 2007) operating at ~90MHz that will analyze radar waves transmitted by a lander, propagated through the body and received by the orbiter with a time resolution of up to 50 ns (7 m). The instrument concept selection will depend in the mission constraints (presence of a lander, antenna accommodation) and target characteristics (size).

- Imaging the first ~50m of the subsurface with a decimetric resolution to identify layering and to reconnect surface measurement to internal structure requires a higher frequency radar (Orbiter only) like Wisdom developed for ExoMars Rover (ESA) (Ciarletti, 2010) with a frequency ranging from 300 MHz up to 2.7 GHz.

3.3 A Summary of Potential Targets.

Object	Activity type	Suitable Target?
133P/Elst-Pizarro	Repeating	Compatible
238P/Read	Repeating	Compatible
176P/LINEAR	Single active period	Unlikely
P/2008 R1 (Garradd)	One active period, awaiting return	Unknown
P/2010 R2 (La Sagra)	Single active period	Incompatible
300163	Single active period	Incompatible
P/2010 A2 (LINEAR)	Single active period - impact?	Incompatible
596 Scheila	Single active period – impact	Incompatible
P/2012 F5 (Gibbs)	Single active period – impact?	Incompatible
P/2012 T1 (Pan-STARRS)	One active period so far	Unknown

In the table above are listed the primary MBC targets for a mission to a volatile-rich asteroid. All objects have displayed the presence of a dust coma and/or tail, however, the activation mechanism is key to determining whether they are “true” MBCs or not. Elst-Pizarro and Read have displayed the same activity during multiple perihelion passes, indicating that the period during which they eject dust are seasonal in nature. These objects therefore are by far the most likely to display cometary activity when visited by a spacecraft. Other targets have only shown activity during one orbit, suggesting that their active periods may have been “one-offs”. Visits to these targets would be much more risky in terms of science output from the mission. Finally, we list for completeness asteroids that have displayed

cometary activity at least once, but are strongly suspected, or known, to have been caused by the impact of a much smaller body.

Currently, our most likely target is the first-known MBC, 133P/Elst-Pizarro, primarily because it is the best-characterised member of the population to date (Hsieh et al. 2004; Hsieh et al. 2010), having been observed for over 17 years through three active episodes.

Its active seasons are well constrained, and have been used to model the activity as due to a single active patch at mid latitudes on the surface. Its approximate shape is known from light-curves, and its albedo has been estimated (Hsieh et al. 2004; Hsieh et al. 2009; Bagnulo et al. 2010). The properties of its surface material have also been probed using polarimetry (Bagnulo et al. 2010). Additionally, there are ongoing observation campaigns to further improve our knowledge of the properties of MBCs, including surveys to find more of these objects and follow up studies on all known ones. In the case of 133P, further ground based observations will allow the determination of a more complex shape model

and pole solution using the light-curve inversion method (Kaasalainen & Torppa 2001), refinement of the rotation period, and more precise measurements of the exact dates of active periods. Together these will allow more detailed modelling of the size and location of the active area(s) ahead of in situ observation by the spacecraft. These studies will also give similar details on other MBCs, including the 5 known today and others that will be discovered by surveys in the next years, which will potentially produce other suitable alternative targets for the mission.

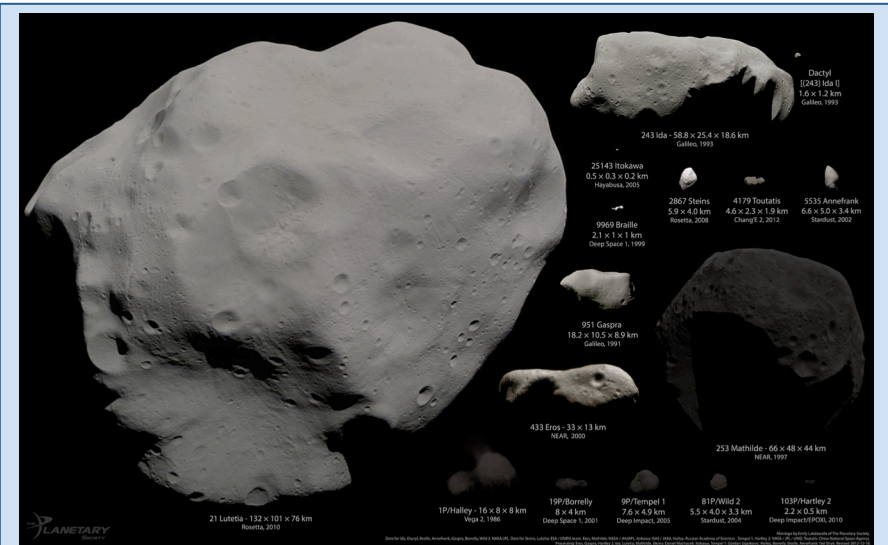


Figure 4: All asteroids and comets imaged at close range, shown to scale, with the exception of Vesta. MBC Elst-Pizarro has a size comparable to the nucleus of 103P/Hartley 2 (lower right). Montage by Emily Lakdawalla. Data from NASA/JPL/JHUAPL/UMD/JAXA/ESA/OSIRIS team/Russian Academy of Sciences/China National Space Agency. Processed by E. Lakdawalla, D. Machacek, T. Stryk, G. Ugarkovic.

4 Mission scenarios

Several mission scenarios, of varying complexity, and hence cost, can be envisaged to

address the science goals described above. We outline the primary scenarios below, concluding with a discussion of more ambitious missions which are probably beyond the scope of the L2/L3 missions with existing technology.

4.1 Detailed characterisation of a single target

For a lander spacecraft a very sensitive mass spectrometer could be landed at the active region to measure D/H and the isotopes of the noble gases. This may work even without a drill, by just sampling the sublimating gas.

4.2 A multiple target mission

As there are several potential targets, it is conceivable that a well-instrumented spacecraft could target several of them. Such a mission in the Main Belt is currently underway: Dawn is a NASA ion-propulsion spacecraft that, after one year spent orbiting Vesta, is now on its way to Ceres. It represents a kind of very flexible and efficient mission. In the case of an MBC mission, the benefits of a multi-targets mission are obvious. Besides visiting more than one MBCs, the option “one MBC + 24 Themis” would maximize

the scientific results. Another possibility would be visit an MBC and one of the bodies considered to be dead comets. The comparison of the two bodies would be extremely useful and interesting.

4.3 A sample return of hydrated mineral grains.

As well as *in-situ* science, there is the option of a sample return mission. The great value of such an approach is that samples can be analysed on Earth using multiple instruments without regard for payload constraints. Instruments used in the subsequent analysis can include synchrotrons, or even instruments not yet available as the samples will be available for curation and long term analysis. The successes of the Stardust (Brownlee et al., 2006) and Hayabusa (Nakamura et al., 2011) missions (to comet 81P/Wild 2 and asteroid 25143 Itokawa respectively show the scientific value of sample return missions.

The simplest technique for collecting samples is a passive capture system deployed during a fly-by of an MBC whilst it is actively emitting dust. This is what was done on the Stardust mission at a comet, where two collection media were used, aerogel and aluminium foils (see Tsou et al., 2003). Aerogel is a very low density, highly porous media (see Burchell et al., 2006 for a discussion of its use in space to capture cosmic dust). Small dust particles which impact it at speed tunnel into it and can be found semi-intact at the ends of tracks in the aerogel, or as fragments lining the walls of the tracks. Stardust captured dust at a fly-by speed of 6 km s^{-1} and has yielded thousands of grains for analysis. SiO_2 based aerogels are transparent, and locating and extracting the captured particles is relatively straightforward. The use of aluminium foils in the capture cell (10% of the total area on Stardust) is to permit impacts on an effectively semi-infinite dense target. Classical hypervelocity impact craters result from these impacts and are lined with residue. Although highly processed during capture at 6.1 km s^{-1} on the Stardust mission (Burchell and Kearsley, 2009, estimated peak shock pressures of 60 – 90 GPa) it has been shown that some residue can retain its original crystallinity, elemental analysis is possible on the residues and even dust particle shape can be inferred (Burchell et al., 2008, Kearsley et al., 2008, Wozniakiewicz et al., 2012).

The key parameters for dust collection are the (i) fly-by speed (which should not exceed $8 - 10 \text{ km s}^{-1}$ before capture effects become too severe, (ii) the emission rate of the dust and (iii) the distance of closest approach (which influences the dust flux that is intercepted), and (iv) capture media collection area. As well as collection during the fly-by of the target MBC, more capture media can be exposed during the mission cruise phase as was done on the Stardust mission. This permits selected targeting of different dust sources in the solar system. For example, if the capture media were exposed whilst in the asteroid belt it would deliver both samples of asteroid dust in general, plus a measure of the time integrated flux experienced in that region. Use of microphones on the capture system or elsewhere on the spacecraft can provide real-time data on impact rates both during the encounter and cruise phases (e.g. see Tuzzolino et al., 2003; 2004. After collection, the spacecraft needs to fly past the Earth, delivering its sample holder for re-entry. The samples can then be examined and curated for distribution to teams around the world. A disadvantage of this sampling technique is that volatiles themselves may not be captured in the sampling medium, i.e. it is hydrated grains that would be collected.

4.4 Other potential mission scenarios

Many more ambitious mission scenarios can also be explored, but are likely to be beyond the cost envelope of the L2/L3 missions. These include a Deep-Impact style mission, where an impactor spacecraft would target a body of interest for a high velocity impact, while another spacecraft would remotely observe the results of the collision. Such a mission would provide invaluable insights into the near-surface volatile content of the target body. However, targeting a small body at Main Belt distances would be incredibly challenging. In addition, it would be prudent to carry out a reconnaissance of the target body prior to impact to select a suitable impact area, and then to target that location with the impactor, further raising the technological challenges. A pre-impact reconnaissance would require that the surveying spacecraft would arrive at the target well before the impactor, almost certainly requiring two launches, or at least separate departure of the two mission elements from Earth orbit. More realistic means of attempting to initiate fresh activity at an MBC could involve the detonation of explosives at the

surface, similar to the explosive-accelerated kinetic impactor that will be included in the JAXA Hayabusa II mission.

The ultimate in sample return would be the delivery to Earth of a gathering of volatile-rich material from the target body obtained from the surface or sub-surface of the target body. Ideally, this material would be kept at cryogenic temperatures throughout its journey to Earth. However, achieving this would be immensely challenging. A compromise sample return mission could require a hermetically sealed container without such severe temperature controls imposed on the storage of its contents. Although volatiles contained therein would evaporate, a well-sealed container would allow the full analysis of the volatiles with well-equipped Earth-based laboratory instrumentation.

5 Summary and Conclusion

MBCs are intriguing since they are dynamically indistinguishable from Main Belt asteroids yet periodically exhibit cometary activity. They are also particularly interesting for astrobiology because recent research suggests that the outer asteroid belt, where most MBCs are found, may have played an important role in the primordial delivery of water to Earth.

By sampling and analysing volatile material at an asteroid, the mission would:

- Gather data critical for evaluating hypotheses concerning the primordial delivery of water and other volatile materials to Earth, shedding light on the question of how life itself came to be possible;
- Explore a member of a newly discovered class of objects that challenges classical definitions of both asteroids and comets;
- Study of some of the most pristine material remaining in the modern solar system; and
- Greatly increase our understanding of the compositional and temperature structure of the protoplanetary disk, in turn providing insights into conditions that may give rise to habitable extrasolar planets.

Although we have stressed the importance of the determination of D/H to ascertain the origin of water at Earth, we also emphasize that measurements by an in situ mass spectrometer would allow the determination of other key isotope ratios of direct relevance to the origin of Earth and its atmosphere, such as those of nitrogen. Previous missions to the Main Belt have visited inert asteroids but not volatile-rich bodies, while cometary missions have visited Jupiter family and Halley-type comets, all relatively close to the Sun when the comets have been highly active. Rosetta will rendezvous with a comet far from Sun and follow it as its activity evolves, but it is already clear from ground-based observations that the activity of Jupiter family comets and MBCs must be very different.

It is natural to ask whether, in a period of major enhancements in the capabilities of ground- and space-based observatories, whether such facilities could address the primary science goals of the proposed mission. As yet, in the absence of the positive detection of an MBC gas coma, the remote measurement of D/H and other isotope values at MBCs is unattainable. Although it is conceivable that facilities such as ALMA, ELT, and JWST could provide valuable data on these objects within the timeframe of the L2/L3 missions, in situ measurements would still be strongly preferable. As well as the increased precision of in situ observations, a spacecraft could place isotope measurements in context, allowing the characterization of this new class of solar system bodies, and in turn, key data on the conditions under which they were formed.

The mission scenarios presented above clearly meet the requirements of ESA's Cosmic Vision programme. They would primarily address the Cosmic Vision thematic questions of "What are the conditions for life and planetary formation?" and "How does the solar system work?", by providing a wealth of invaluable, highly relevant scientific data which would contribute to providing the answers to those fundamental questions. Building on the successes of previous small-body missions such as Giotto, Stardust, and Deep Impact, and with the ongoing and anticipated wealth of results from Rosetta and Dawn, we believe that a visit to this newly-discovered class of objects promises to provide a step-change in our knowledge of the early solar system, and in turn that of our own planet.

References

- A'Hearn & Feldman (1992) *Icarus* 98, 54
 Anders & Owen (1977) *Science* 198, 453
 Bagnulo et al. (2010) *A&A* 514, A99
 Balsiger et al. (1995) *JGR* 100, 5827
 Baratta et al. (1994) *P&SS* 42, 759
 Barucci et al. (1998) *Icarus* 132, 388
 Beck et al. (2011) *A&A* 526, A85
 Bockelee-Morvan et al. (1998) *Icarus* 193, 147
 Boss (1998) *Ann. Rev. Earth Planet. Sci.* 26, 53.
 Brown et al. (2012) *P&SS* 60, 166
 Brownlee et al. (2006) *Science* 314, 1711
 Burchell & Kearsley (2009) *P&SS* 57, 1146
 Burchell et al. (2008) *M&PS* 43, 135
 Burchell et al. (2006) *Ann. Rev. Earth Planet. Sci.* 34, 385.
 Campins et al. (2010) *Nature* 464, 1320
 Capria et al. (2012) *A&A* 537, A71
 Ciarletti et al. (2010) *Proceedings of the IEEE*, Manuscript ID: 0023-SIP-2010-PIEEE.
 Cohen & Coker (2000) *Icarus* 145, 369
 Dauphas (2003) *Icarus* 165, 2, 326
 Davis et al. (2005) *ApJ* 620, 994
 De Val-Borro, et al. (2012) *A&A* 546, A4
 Drake (2005) *M&PS* 40, 519
 Eberhardt et al. (1995) *A&A* 302, 301
 Epstein et al. (1976) *Science* 263, 198
 Fanale & Savail (1989) *Icarus* 82, 97
 Genda & Ikoma (2008) *Icarus* 194, 42
 Genda, H., (2009) private communication, December 11, 2009.
 Gomes et al. (2005) *Nature* 435, 466
 Gradie & Tedesco (1982) *Science* 216, 1405.
 Grimm & McSween (1989) *Icarus* 82, 244
 Grimm & McSween (1993) *Science* 259, 653
 Hartogh et al. (2011) *Nature* 478, 218
 Heiken et al. (1991) *Journal of the British Astronomical Association* 101, 362
 Hiroi et al. (1996) *M&PS* 31, 321.
 Horner et al. (2007) *Earth, Moon and Planets* 100, 43.
 Hsieh et al. (2004) *AJ* 127, 2997
 Hsieh & Jewitt (2006) *Science* 312, 561
 Hsieh (2009) *A&A* 505, 1297
 Hsieh et al. (2009) *ApJ Letters* 694, L111
 Hsieh et al. (2010) *MNRAS* 403, 363
 Hsieh et al. (2011) *ApJL* 736, L18
 Hsieh et al. (2012) *AJ* 143, 104
 Hutsemékers et al. (2008) *A&A* 490, 31
 Ikoma & Genda (2006) *ApJ* 648, 696
 Izidoro et al. (2013) *ApJ* 767, 54
 Javoy et al. (2010) *EPSL* 293, 259
 Jewitt et al. (2010) *Nature* 467, 817
 Jewitt et al. (2009) *AJ* 137, 4313
 Jewitt (2012) *AJ* 143, 66
 Jewitt & Guilbert-Lepoutre (2012) *AJ* 143, 21
 Jones et al. 1990, *Icarus* 88, 172
 Jorda et al. 2008, ACM meeting, abs#8046
 Kaasalainen & Torppa (2001) *Icarus* 153, 24
 Kearsley et al. (2008) *M&PS* 43, 41
 King et al. (1992) *Science* 255, 1551
 Kofman et al. (2007) *Sp. Sci. Rev.* 128, 413
 Lebofsky (1980) *AJ* 85, 573
 Lebofsky et al. (1981) *Icarus* 48, 453
 Levison et al. (2009) *Nature* 460, 7253
 Licandro et al. (2011a) *A&A* 525, A34
 Licandro et al. (2011b) *A&A* 532, 65L
 Licandro et al. (2013) *A&A* 550, 17L
 Marzari et al. (2011) *Icarus* 214, 622
 McKeegan et al. (2006) *Science* 314, 1724
 Meier et al. (1998) *Science* 279, 842
 Mojzsis et al. (2001) *Nature* 409
 Morbidelli et al. (2000) *M&PS* 35, 1309
 Morbidelli et al. (2005) *Nature* 435, 462
 Moreno et al. (2011) *ApJ* 738, 16
 Moreno et al. (2012) *ApJ* 761, 12M
 Mousis et al. (2000) *Icarus* 148, 513
 Mousis & Alibert (2005) *MNRAS* 358, 188
 Muralidharan et al. (2008) *Icarus* 198, 400
 Nakamura et al. (2011) *Science* 333, 1113
 O'Brien et al. (2006) *Icarus* 184, 39
 Owen & Bar-Nun (1995) *Icarus* 116, 215
 Picardi et al. (2005) *Science* 310, 1925
 Prialnik & Rosenberg (2009) *MNRAS* 399, 79
 Raymond et al. (2004) *Icarus* 168, 1
 Rivkin et al. (2006) *Icarus* 185, 563
 Rivkin & Emery (2010) *Nature* 464, 1332
 Robert et al. (2000) *Sp. Sci. Rev.* 92, 201
 Robert (2001) *Science* 293, 1056
 Robert (2006) in *Meteorites and Early Solar System* (Lauretta & McSween Eds.), 341
 Roessler et al. (1992) *Annales Geophysicae* 10, 232
 Rubin et al. (2002) *M&PS*, 37, 125
 Sasselov & Lecar (2000) *ApJ* 528, 995
 Schorghofer (2008) *ApJ*, 682, 697
 Snodgrass et al. (2010) *Nature* 467, 814
 Stevenson et al. (2012) *ApJ* 759, 142
 Stimpfl et al. (2006) *J. Cryst. Growth* 294, 83
 Strazzulla et al. (1991) *Icarus* 91, 101
 Takir & Emery (2012) *Icarus* 219, 641
 Teolis et al. (2010) *Science* 330, 1813
 Thomas et al. (2005) *Nature* 437, 224
 Tsiganis et al. (2005) *Nature* 435, 459
 Tsou et al. (2003) *JGR E108* (E10), 8113
 Tuzzolino et al. (2003) *JGR E108* (E10), 8115

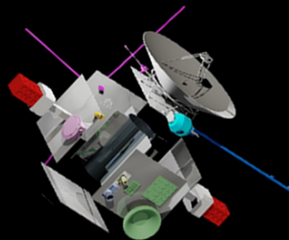
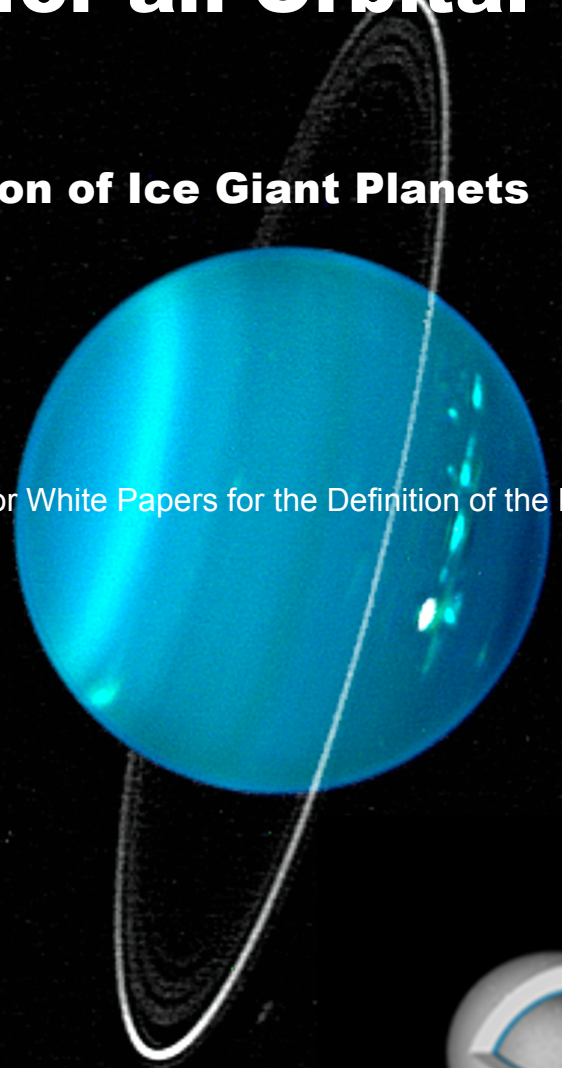
Mission to a Volatile-Rich Asteroid

- Tuzzolino et al. (2004) *Science* 304, 1776
Vernazza et al. (2005) *A&A* 436, 1113
Vilas, F., Jarvis, K. S., & Gaffey, M. J. 1994, *Icarus* 109, 274
Villanueva, G. L., et al. 2009, *AJ Letters* 690, L5
Wagoner et al. (1967) *ApJ* 148, 3
Walsh et al. (2011) *Nature* 475, 206
Weaver et al. (2008) in proceedings of Asteroids Comets and Meteors 2008
Wozniakiewicz et al. (2012) *M&PS* 47, 660
Zolensky et al. (1999) *M&PS Suppl.* 34, A124
Zolensky et al. (2000) *M&PS* 35, 9

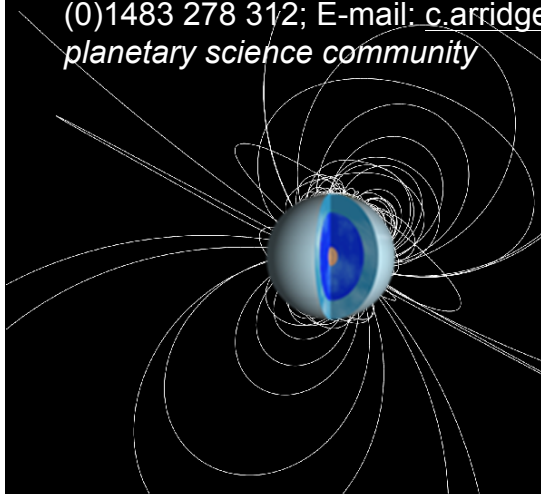
The Science Case for an Orbital Mission to Uranus

Exploring the Origins and Evolution of Ice Giant Planets

A paper submitted in response to the ESA 2013 Call for White Papers for the Definition of the L2 and L3 Missions in the ESA Science Programme



Spokesperson: Dr. Christopher S. Arridge, Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK. Tel: +44 (0)1483 204 150; Fax: +44 (0)1483 278 312; E-mail: c.arridge@ucl.ac.uk; *on behalf of 219 members of the international planetary science community*



List of Supporters and Contributors

This white paper is supported by 219 scientists world-wide. The full list of supporters is available online at <http://www.mssl.ucl.ac.uk/planetary/sites/uranus-whitepaper/>. The list below highlights the contributors to this white paper:

Austria: Andrea Maier (*Space Research Institute*); Armin Luntzer, Reinhard Tlustos (*U. Vienna*).

Belgium: Bertrand Bonfond (*U. Liège*).

France: Sebastien Charnoz (*CEA*); Christelle Briois (*CNRS*); Nicolas Rambaux (*IMCCE*); Mattieu Laneuville (*IPGP*); Nicolas André, Renaud Sallantin (*IRAP*); Thibault Cavalié (*LAB*); Sébastien Hess (*LATMOS*); Carine Briand, Daniel Gautier, Laurent Lamy (*LESIA*); Olivier Mousis (*Obs. de Besançon*); Bruno Christophe (*ONERA*); Gabriel Tobie (*U. Nantes*).

Germany: Achim Morschhauser (*DLR*); Jonathan Hillier, Mario Tieloff (*Heidelberg U.*); Fritz Neubauer (*U. Cologne*); Eberhard Grün, Harald Krüger, Peter Strub (*Max Planck Inst. Nuclear Physics*); Elias Roussos (*Max Planck Inst. Solar System Research*); Jürgen Blum (*Technical U. Braunschweig*); Konstantinos Konstantinidis (*U. der Bundeswehr München*); Frank Spahn (*U. Potsdam*); Nadine Nettelmann (*U. Rostock*); Emil Khalisi, Anna Mocker, Georg Moragas-Klosterbeyer, Frank Postberg, Rachel Soja, Ralf Srama, Veerle Sterken (*U. Stuttgart*); Christos Labrianidis (*UTESat-Spacecom GmbH*).

Greece: Ioannis Daglis (*U. Athens*).

Ireland: Colm Bracken, Susan McKenna-Lawlor (*National University of Ireland, Maynooth*).

Italy: Gianrico Filacchione, Davide Grassi, Davide Lucchesi, Anna Milillo, Roberto Peron, Federico Tosi, Diego Turrini (*INAF/IAPS*).

Japan: Hiroshi Kimura (*Center for Planetary Science*).

Spain: Marc Costa-Sitja (*ESAC*).

The Netherlands: Tatiana Bocanegra-Bahamon, Dominic Dirkx, Ingo Gerth (*Delft U. of Technology*); Jessica Agarwal (*ESTEC*).

United Kingdom: Craig Agnor (*QMUL*); Caitriona Jackman, Jane MacArthur, Tom Nordheim (*UCL*); Richard Ambrosi, Sarah Badman, Henrik Melin, Tom Stallard (*U. Leicester*); Richard Holme (*U. Liverpool*); Leigh Fletcher (*U. Oxford*).

United States of America: Don Banfield, Matt Hedman, Matthew Tiscareno (*Cornell*); Abi Rymer, Elizabeth Turtle (*JHU/APL*); Mihaly Horanyi, Sean Hsu, Sascha Kempf, Zoltan Sternovsky, Rob Wilson (*LASP, Colorado U.*); Shawn Brooks, Julie Castillo-Rogez, Mark Hofstadter (*NASA/JPL*); Kurt Retherford (*SwRI*); Julianne Moses (*Space Science Institute*); Christopher Russell (*UCLA*); George Hospodarsky (*U. Iowa*);

Fourteen of the twenty ESA Member and European Cooperating states are represented in the support for a Uranus orbiter mission. The table below lists the supporters by country, first listing the ESA member and European Cooperating states, then the rest of the world.

Country	Number of supporters
Austria	5
Belgium	5
Czech Republic	2
Finland	2
France	32

Germany	31
Greece	2
Hungary	2
Ireland	2
Italy	44
Poland	1
Spain	5

Sweden	3
The Netherlands	5
United Kingdom	29
Argentina	1
Japan	3
Russia	1
USA	44

Executive Summary

Giant planets account for more than 99% of the mass of the Sun’s planetary system, and helped to shape the conditions we see in the Solar System today. The Ice Giants (Uranus and Neptune) are fundamentally different from the Gas Giants (Jupiter and Saturn) in a number of ways and Uranus in particular is the most challenging to our understanding of planetary formation and evolution. A mission to the uranian system will open a new window on the origin and evolution of the Solar System and directly addresses two of the Cosmic Vision themes “What are the conditions for Planet Formation and the Emergence of Life?” and “How Does the Solar System Work?”. The fundamental processes occurring within the uranian system confirm that the exploration of Uranus is essential in meeting ESA’s Cosmic Vision goals.

The science case for a Uranus mission is arranged into three key themes: 1) Uranus as an Ice Giant Planet, 2) An Ice Giant Planetary System, and 3) Uranus’ Aeronomy, Aurorae and Highly Asymmetrical Magnetosphere. In addition, a mission to Uranus naturally provides a unique opportunity to study the outer heliosphere and fundamental gravitational physics and so we present a significant cruise phase science programme. The mission concept we propose consists of a Uranus orbiter combined with an atmospheric entry probe to provide crucial ground-truth measurements in the atmosphere of Uranus. The mission requires the development of radioisotope power sources that are already under development through ESA contracts, but which are expected to be available in good time before the next L-class launch opportunity. Otherwise the mission can be achieved with current technology.

This white paper has significant community support, reflected by (i) the 219 scientists across the world (170 in Europe from 14 of the 20 ESA member and European cooperating states) lending their support to this white paper; (ii) the key planetary objectives specified by white paper submissions to NASA’s Planetary and Heliophysical Decadal Surveys; and (iii) the wide support provided to the Uranus Pathfinder (Arridge et al., 2012) M-class mission proposal in 2010. Perhaps unsurprisingly, the level of community support is highest among early- and mid-career scientists. In September 2013 a Uranus-focused workshop will be held in France, demonstrating the world-wide interest in studies of Uranus. A European mission to a strange and distant world like Uranus provides a unique opportunity for public engagement, inspiring the next generation of European scientists. Its value to ESA and Europe should not be underestimated. The potential scientific return from the level of investment from ESA and the European community is without peer.

Uranus Mission Summary

Orbiter scientific payload	Focussed set of space qualified high (>5) TRL instruments, most with significant heritage on ESA missions: magnetometer, radio science, accelerometer, imaging spectroscopy in UV and near-IR, thermal infrared bolometer, narrow/wide angle visible cameras, radio and plasma wave science, dust detector, plasma/energetic particle detectors, Doppler imager.
Atmospheric Entry Probe	Probe to reach >5 bar depth with a mass of 312 kg. Instruments: nephelometer, ultrastable oscillator for radio science, accelerometer, and a mass spectrometer.
Mission profile	Launch from Kourou on Ariane 5 injecting into a polar science orbit. Nominal mission of 20 orbits with targeted flybys of all five major natural satellites.
Platform	Mars Express/Venus Express/Rosetta heritage platform with 4 m HGA and 100 W transmitter. AOCS capable of spin stabilisation and three-axis stabilisation.
Power	ESA radioactive power source (²⁴¹ Am) providing 400 W during prime mission
Operational lifetime	~16 years.
Launch and interplanetary cruise	Ariane 5 from Kourou. Estimate 10 – 15 year interplanetary cruise utilising a variety of gravity assists, consistent with previous mission designs.
Telemetry band	X / Ka
Data volume	160 Mbit per 8 hour downlink (5.6 kbps) over Ka-band
Orbit	Polar science orbit with multiple targeted flybys of major moons. Low periapsis (<1.1 R _U) orbits with Earth visibility required for studies of Uranus’ interior.

1 Science Case

Giant planets account for more than 99% of the mass of the Sun's planetary system, and helped to shape the conditions we see in the Solar System today. Our Solar System provides the only local laboratory in which we can perform studies that help us to understand the nature of planetary systems in general. Kepler observations that Uranus/Neptune class planets are a common class of planet (Fressin et al., 2013) make it all the more timely and compelling to better explore these fascinating systems. The Ice Giants are fundamentally different from the Gas Giants (Jupiter and Saturn) in a number of ways and Uranus in particular is the most challenging to our understanding of planetary formation and evolution, with its puzzling interior structure, unclear energy balance and internal energy transport mechanisms, and its high obliquity. Yet our exploration of the Ice Giants in our own Solar System remains incomplete, with several fundamental questions unanswered. Voyager 2 remains the only spacecraft to have returned data from the uranian environment and by the time of the L2/L3 launch, more than 40 years will have passed since new measurements were returned from Uranus.

A mission to Uranus will provide observations and measurements that are vital for understanding the origin and evolution of Uranus as an Ice Giant planet, answer the fundamental question of why some giant planets become icy and other so gas rich, and provide a missing link between our Solar System and planets around other stars. Observations of Uranus' rings and satellite system will also bring new perspective on the origin of giant planet systems and will help validate the models proposed for the origin and evolution of Jupiter's and Saturn's systems. A new planetary science mission to Uranus thus represents the quintessential aspects of ESA's Cosmic Vision.

The science case and main science questions are arranged into three themes: Uranus as an Ice Giant Planet, An Ice Giant Planetary System, and Uranus' Aeronomy, Aurorae and Highly Asymmetrical Magnetosphere. In addition we propose a significant cruise phase science programme.

1.1 Uranus as an Ice Giant Planet: The Interior and Atmosphere of Uranus

Table 1 lists the gross properties of Uranus. The bulk composition and internal structure of the Ice Giants reflect their different formation environments and evolutionary processes relative to the Gas Giants (e.g., Guillot, 2005), providing a window onto the early Solar System. Jupiter is an H/He planet with an ice and rock mass fraction of 4 – 12% as inferred from standard

interior models (Saumon and Guillot, 2004). Uranus and Neptune seem to consist mostly of ices and rocks, but current observations are only able to provide an upper limit of 85% on the ice and rock mass fraction (Fortney and Nettelmann, 2010). There is currently no interior model for Uranus that agrees with all the observations, representing a significant gap in our understanding of the Solar System. Understanding the internal structure of Uranus (the nearest Ice Giant) is indispensable for estimating the bulk composition of outer planets, in particular their ice:rock ratio. A Uranus orbiter mission will reveal the fundamental processes that shape the formation, evolution, dynamic circulation and chemistry of Ice Giant atmospheres.

Equatorial radius	25 559 km (=1 R _U)
Mass	14.5 M _E
Sidereal spin period	17h12m36s (±72 s)
Obliquity	97.77°
Semi-major axis	19.2 AU
Orbital period	84.3 Earth years
Dipole moment	50 M _E
Magnetic field	Highly complex, surface field up to 110000 nT
Dipole tilt	-59°
Natural satellites	27 (9 irregular)

Table 1: Physical and orbital parameters of Uranus.

The origin of Uranus' large obliquity is perhaps one of the most outstanding mysteries of our Solar System. A variety of explanations have been invoked, including a giant impact scenario which may also be implicated in Uranus' low luminosity and small heat flux, and tidal interactions (Boué and Laskar, 2010; Morbidelli et al. 2012). Examining the interior structure and composition of Uranus and its natural satellites, and studying the ring system may allow us to unravel the origin of this Solar System mystery.

Planets are warm inside and cool down as they age. Gravitational energy from material accretion was converted to intrinsic, thermal energy during formation and is steadily radiated away through their tenuous atmospheres as they age. Thermal evolution models probe the energy reservoir of a planet by predicting its intrinsic luminosity. Such models reproduce the observed luminosity of Jupiter and Neptune after 4.56 Gyrs of cooling, independent of detailed assumptions about their atmosphere, albedo, and solar irradiation. The same models, however underestimate it for Saturn and overestimate it for Uranus. Indeed, Uranus's atmosphere appears so cold (its intrinsic luminosity so low), that according to standard thermal evolution theory Uranus should be more than 3 billion years older than it is. However, the error bars on the Voyager-

determined energy balance are large enough to substantially reduce that discrepancy. In particular, as the observational error bars (Pearl et al., 1990) of the albedo and brightness-temperature data based effective temperature are significant, Uranus could as well cool down adiabatically, just as Neptune, if its real heat loss is close to the observed upper limit. This demonstrates the need for improved observational data for constraining Uranus' evolution and the derived structure and formation. The intrinsic luminosity of Uranus also has implications for understanding planetary dynamos and magnetic field generation. The unusual field properties suggest some fundamental difference between the dynamos of Uranus and Neptune and those of the other planets.

Uranus' atmosphere is unique in our solar system in that it receives a negligible flux of heat from the deep interior and experiences extremes of seasonal forcing due to the high 98° obliquity, with each pole spending 42 years in darkness. This unusual balance between internal and radiative heating means that Uranus' unique weather is governed principally by seasonal forcings. Furthermore, the substantial enrichment of some heavy elements (but perhaps not all, N being strongly depleted in the troposphere) and small envelopes of H₂-He in the Ice Giants and the cold atmospheric temperatures relative to the Gas Giants yield unique physiochemical conditions. Uranus therefore provides an extreme test of our understanding of planetary atmospheric dynamics; energy and material transport; seasonally varying chemistry and cloud microphysics; structure and vertical coupling throughout giant planet atmospheres. Even though Earth-based observations of Uranus (ISO, Spitzer, Herschel, ground-based) have improved dramatically in the decades since Voyager 2, many questions about this unexplored region of our Solar System remain unanswered. Certain spectral regions, particularly those longward of 20 μ m, are inaccessible from the ground. The overarching atmospheric science objective is to explore the fundamental differences in origin, meteorology and chemistry between the Ice and Gas Giants; to reveal the underlying mechanisms responsible for Uranus' unique conditions.

The temperature in Uranus' thermosphere is several hundred degrees hotter than can be explained by solar heating (as is also found for Saturn and Jupiter) and remains a fundamental problem in our understanding of giant planet upper atmospheres in general. Moreover, this temperature is strongly correlated with season such that at Solstice, the upper atmosphere is more than 200 K hotter than at Equinox. The exosphere and co-located thermosphere and ionosphere form a crucial transition region between interplanetary space and the planet itself.

Powerful currents, generated by electric fields imposed by the magnetosphere of magnetised planets, may result in large energy inputs to the upper atmosphere due to Joule heating and ion drag; the energy from these sources may be tens to hundreds of times greater than that due to the absorption of solar (EUV) radiation. The unique orientations of Uranus' magnetic dipole and spin axis combined with strong seasonal driving produce a highly time-dependent and complex interaction between the solar wind, magnetosphere, ionosphere and thermosphere. This system provides a unique opportunity to understand how insolation and particle precipitation from the solar wind magnetosphere contribute to the energy balance in the upper atmosphere.

The composition of Uranus contains clues to the conditions in the protosolar cloud and the locations in which it formed. For instance, a subsolar C:O ratio could indicate formation at a distance where water (but not CH₄) was frozen. The common picture of gaseous planet formation by first forming a 10 M_E core and then accreting a gaseous envelope is challenged by state-of-the-art interior models, which instead predict rock core masses below 5 M_E (Saumon and Guillot, 2004; Fortney and Nettelmann, 2010). Uranus inclination and low heat loss may point to another catastrophic event and provides additional important constraints for planetary system formation theory. New observations of Ice Giants are therefore crucial in order to resolve this and achieve Cosmic Vision goals on the formation of planets.

1.1.1 What is the internal structure and composition of Uranus?

At present there is no Uranus interior model that is consistent with all of the physical constraints, such as Uranus' gravity field, luminosity, magnetic field, and realistic ice:rock ratio. Figure 1 illustrates a model that is consistent with the gravity and magnetic field data but not with the luminosity of Uranus. Uranus and Neptune are known to have substantial elemental enrichments in carbon and deuterium (Owen and Encrenaz, 2006; Feuchtgruber et al., 2013), but abundances of other simple elements (N, S and O), their isotopic ratios (¹²C/¹³C, ¹⁴N/¹⁵N, ¹⁶O/¹⁷O) and the noble gases (He, Ne, Ar, Xe, Kr) have never been adequately constrained. Nevertheless, Uranus' bulk atmospheric composition provides a key diagnostic of planetary formation models.

To develop improved models of Uranus' interior better compositional data must be obtained (Helled et al., 2010). The mass of the core also places constraints on planetary formation models. If H/He is mixed into the deep interior with only a small central core this would suggest gas accretion onto a low-mass proto-planetary

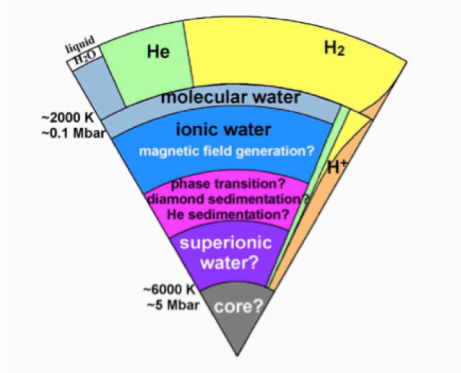


Figure 1: Illustration of an interior model which satisfies some but not all of the observational constraints.

core, or efficient vertical mixing, or inclusion of disk-gas into the building planetesimals, rather than accretion onto a large ice-rock core of $\sim 10 M_E$. Furthermore, the predicted large size of Uranus' core relative to the H_2 -He envelope may make Uranus our best opportunity for studying the elemental composition and thermochemistry of the outer solar nebula at the earliest stages of planetary formation. Measurements of Uranus' bulk atmospheric composition, luminosity, magnetic and gravity fields, and normal-mode oscillations will place new constraints on Uranus' interior and on the origins and evolution of Uranus. The gravity field can be measured both by radio science and by observing the precession of Uranus' ten dense narrow elliptical rings (Jacobson et al., 1992; Jacobson, 1998, 2007). Magnetic field measurements will be used to assess the structure of the dynamo region. Measurement of noble gas abundances and isotopic ratios can be achieved with a shallow (1 bar) entry probe (some isotopic ratios can be determined by remote sensing). A deep atmospheric entry probe will enable us to measure if the S/N ratio is enhanced above solar abundance. Giant-planet seismology, building upon the mature fields of helio- and astroseismology, will revolutionise our ability to probe the interior structure and atmospheric dynamics of giant planets.

Improved knowledge of the composition and interior structure of Uranus will also provide deeper insight into the processes that remixed material in the protoplanetary disk, caused for example by the formation of Jupiter (Safronov, 1972; Turrini et al., 2011) or due to extensive primordial migration of the giant planets (Walsh et al., 2011).

1.1.2 Why does Uranus emit very little heat?

Voyager measurements suggest that Uranus' evolution produced a planet with negligible self-luminosity, smaller than any other planet in our Solar System (Pearl et al., 1990). Combined with the sluggish appearance of

the atmosphere as viewed by Voyager, this suggests that the interior of Uranus is either (a) not fully convective or that (b) it suffered an early loss of internal heat. Case (b) would suggest that the interior is colder than in the adiabatic case, with crystalline water deep inside (Hubbard et al., 1995). This points to a catastrophic event in Uranus' early history that shocked the matter and led to a rapid energy loss. In case (a) we would expect the interior to be warmer, with water plasma implying large-scale inhomogeneities, possibly caused by immiscibility of abundant constituents such as helium and carbon or upward mixing of core material, that inhibit efficient heat transport. However, during the last decade ground-based observations have revealed the appearance of prominent cloud features suggesting localised convective regions of adiabatic thermal gradients in the deep troposphere. Vertical transport of energy and material seems to occur only in localised regions on this enigmatic planet. In fact, the inferred size of a non-convective internal region depends sensitively on the imposed intrinsic heat flux value: a mostly stable interior is predicted if the heat flux is close to zero, but a fully convective interior is possible, as for Neptune, should the upper limit of the observed heat flux value prove true.

In order to better constrain Uranus' internal heat flux, which was derived by Voyager from the measured albedo and brightness temperatures, tighter observational constraints of these quantities are necessary. These inferences come from a single measurement from the Voyager flyby, at a single point in Uranus' seasonal cycle. Indeed, ground-based observations of the uranian atmosphere have revealed far more dynamic activity during the present season, just past the northern spring equinox in 2007. The appearance of convective cloud structures in localised regions (typically mid-latitudes) suggests convective regions of adiabatic thermal gradients in the deep troposphere. Thus the balance between Uranus' emission and absorption may be seasonally variable, and new global measurements of reflected solar and emitted IR radiation are required to assess the presence or absence of an internal heat source, and its importance as driving mechanisms for Uranus' meteorological activity. Atmospheric properties and profiles, measured by an atmospheric entry probe using a combination of radio science, an on-board accelerometer and a nephelometer, may also shed light on heat transport in the atmosphere.

1.1.3 What is the configuration and origin of Uranus' highly asymmetric magnetic field?

Understanding the configuration of Uranus' internal magnetic field is essential for correctly interpreting the

configuration of the magnetosphere, its interaction with the rings and moons, and for understanding how dynamo processes in the interior of Uranus generate the field. In contrast to the magnetic fields of Earth, Mercury, Jupiter and Saturn, which are dominated by a dipole nearly co-aligned with the rotation axis, those of Uranus and Neptune are characterised by a large tilt between the dipole and spin axes with strong quadrupole and octupole contributions to the internal magnetic field. The magnetic field data from Voyager 2 are sufficient to crudely constrain the internal field of Uranus but more complex and (currently) poorly constrained models are required to fit the data (Holme and Bloxham, 1996). At the planetary surface the magnetic dipole, quadrupole and octupole components of the total internal field are of comparable strength, but at the top of the dynamo region ($\sim 0.75 R_U$) the latter two dominate. Figure 2 illustrates the highly asymmetrical nature of Uranus' internal magnetic field.

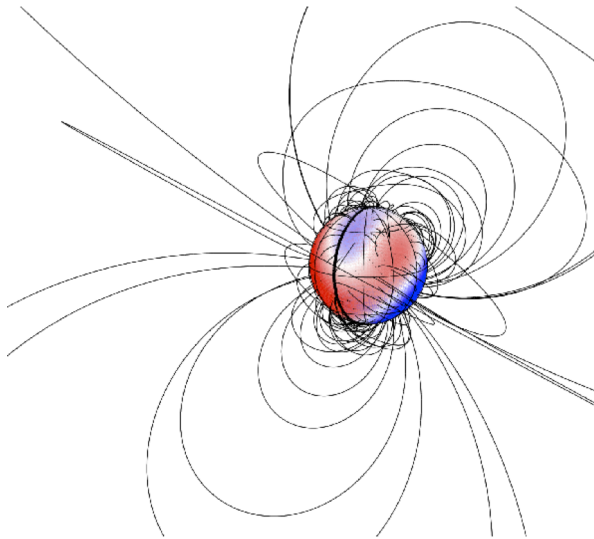


Figure 2: Configuration of Uranus' internal magnetic field (colours indicate the magnitude and sign of the radial field at the surface, bold black curve indicates zero longitude).

A variety of competing numerical dynamo models (e.g., Stanley and Bloxham, 2004, 2006; Soderlund et al., 2013) have been developed which can explain these fields but new magnetic field measurements are required to allow us to determine which is the closest to reality. The field is also expected have undergone secular change since the Voyager 2 epoch (Christensen and Tilgner, 2004). Making magnetic field measurements at a variety of planetocentric latitudes and longitudes will provide a wealth of data from which to test these competing models. This will lead to significant changes in our understanding of field generation in Ice Giant planets and of planetary

magnetic field generation in general. Models of the internal field can also be greatly improved by the use of auroral images which provide additional high-latitude constraints. Herbert (2009) combined the Voyager observations of the internal field with the locus of the UV auroral oval to derive such a higher order model. Better-quality images of auroral emissions than are possible from Earth (e.g., Lamy et al., 2012) are paramount for improving the accuracy of the planetary field model.

1.1.4 What is the rotation rate of Uranus' interior?

A correct interpretation of the internal structure of Uranus relies on an accurate knowledge of the internal rotation rate of the planet (Nettelmann et al., 2013). Modelling of Uranus' internal magnetic field, and observations of radio emissions (Uranian Kilometric Radiation, UKR) and atmospheric motions all provide independent estimates of the rotation rate of the planet, although not always from the same region of the planet. Analyses of Voyager 2 data have yielded three estimates of the rotation rate of Uranus, between 17 hours 12 minutes 36 seconds (± 72 seconds) (Herbert, 2009) and 17 hours 17 minutes 24 seconds (± 36 seconds) (e.g., Ness et al., 1986). New measurements of Uranus' magnetic field and UKR will enable us to significantly improve the accuracy on the determination of the planetary period (to a few parts in 10^{-5}), and check if second order effects (e.g., Saturn displays different radio periods in both magnetic hemispheres, each varying with time) are present.

1.1.5 How is Uranus' weather structure and composition influenced by its unique seasons?

The potential absence of an internal heat source renders Uranus' weather unique among the giant planets. Neptune, with its powerful self-luminosity, provides an important counter-example of a convectively-active weather layer. The extreme 98° obliquity of Uranus subjects the atmosphere to extremes of seasonal forcing, with each pole spending decades in darkness.

Despite the bland visible appearance of Uranus from Voyager, recent ground-based observations have shown the planet to be more dynamically active than previously thought (figure 3). Large-scale atmospheric motions are known to be present, as ground-based microwave observations reveal the polar regions of the deep troposphere to be depleted in absorbers relative to the equator (bright poles in VLA 1.3-cm images, figure 4), while IR measurements see the same pattern at higher altitudes in the CH_4 distribution (Karkoschka and Tomasko, 2009).

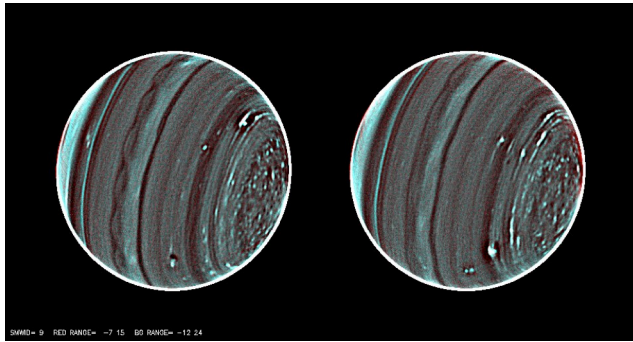


Figure 3: Images from the Keck telescope in 2012 revealing a wave around the equator, discrete clouds at mid latitudes, and a mottled chaotic appearance at the poles (Sromovsky et al., 2012)..

Seasonal changes in clouds and dynamics have also been observed: in 1986, the sunlit South Pole appeared bright due to a polar ‘cap’ of stratospheric aerosols. The bright South Pole diminished over the ensuing years, and became a faint polar band of brighter material, while a new collar of bright material became visible in the northern springtime hemisphere. High resolution ground-based observations in 2012 (figure 3) reveal what may be convective clouds of CH_4 , which may eventually form a polar hood as was seen in the southern polar regions during the Voyager flyby (Sromovsky, Fry, Hammel, de Pater, 2012-2013). All of these are indicative of the meridional circulation, which on this highly seasonally driven planet is likely to be unique, but instructive about how planets work under more general obliquity/insolation conditions. The long temporal baseline of high spatial resolution atmospheric observations will allow us to study the nature, frequency, distribution and morphology of discrete cloud activity (e.g., storms, vortices). In particular, we aim to understand the origin, lifecycle and drift rates of Uranus’ dark spots and associated bright clouds (large anticyclonic vortices, e.g., Hammel et al., 2006), for a direct comparison with the lifecycles observed on Neptune. Finally, the relative importance of wave activity versus moist convection in vertical mixing could be uniquely tested on Uranus, given the anticipated low levels of convective activity.

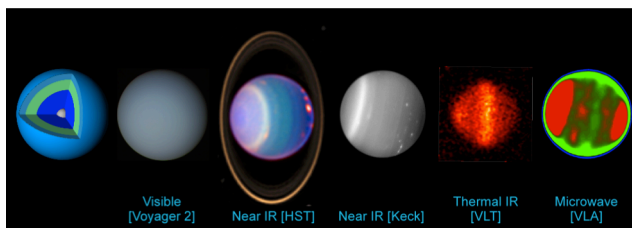


Figure 4: Images of Uranus in a variety of wavelengths compared with an interior model.

1.1.6 What processes shape atmospheric chemistry and cloud formation on an Ice Giant?

Analysis of reflected sunlight observations has attempted to identify the composition of the clouds, with suggestions that the bright white features are ices of CH_4 , overlying a putative cloud of NH_3 or $(\text{NH}_4)\text{SH}$, with a deep cloud of water hypothesised at much deeper levels. Below the clouds, the atmospheric composition is poorly known. The altitude of the deep H_2O condensation cloud is poorly understood because the bulk water abundance may be enhanced by 10-30 times the solar abundance (de Pater and Lissauer, 2010). The H_2O cloud may exist over extended pressure ranges beneath 50-80 bar, and may even merge with a region of super-critical H_2O in Uranus’ interior. It is not clear what chemical gradients are responsible for the emergence of dark spots (anti-cyclones) and associated bright orographic clouds. Above the clouds, the Infrared Space Observatory (ISO, 1995-1998) and Spitzer Space Telescope (2003-Present) showed that stratospheric chemistry initiated by the ultraviolet destruction of CH_4 powers a rich photochemistry, resulting in a soup of hydrocarbons in the upper atmosphere. This hydrocarbon chemistry differs from the other giant planets, as the sluggish vertical mixing means that CH_4 is not transported to such high altitudes, so that hydrocarbon photochemistry operates in a very different regime (i.e., higher pressures) than on the other giants. Furthermore, ISO and Herschel (2009-2013) observed oxygenated species in the high atmosphere, potentially due to infalling dust and comets. It is important to search for previously unidentified or unmapped stratospheric species (CO , HCN , CO_2 , etc.) such as those related to coupling between the neutral atmosphere and the uranian ring/satellite system.

Remote sounding observations are required to place constraints on Uranus’ bulk inventory, vertical distribution, composition, and optical properties of Uranus’ clouds and hazes. A deep (>5 bar) atmospheric entry probe will enable the measurement of bulk CH_4 and H_2S abundances.

1.1.7 What Processes Govern Upper Atmospheric Structure?

The temperature in Uranus’ upper atmosphere (thermosphere) is several hundred degrees hotter than can be explained by solar heating alone (as is also found for Saturn and Jupiter) and remains a fundamental problem in our understanding of giant planet upper atmospheres in general. Moreover, this temperature is strongly correlated with season such that at Solstice, the upper atmosphere in the illuminated hemisphere is more than 200 K hotter than at Equinox. It seems likely that a key component of the required additional heating is

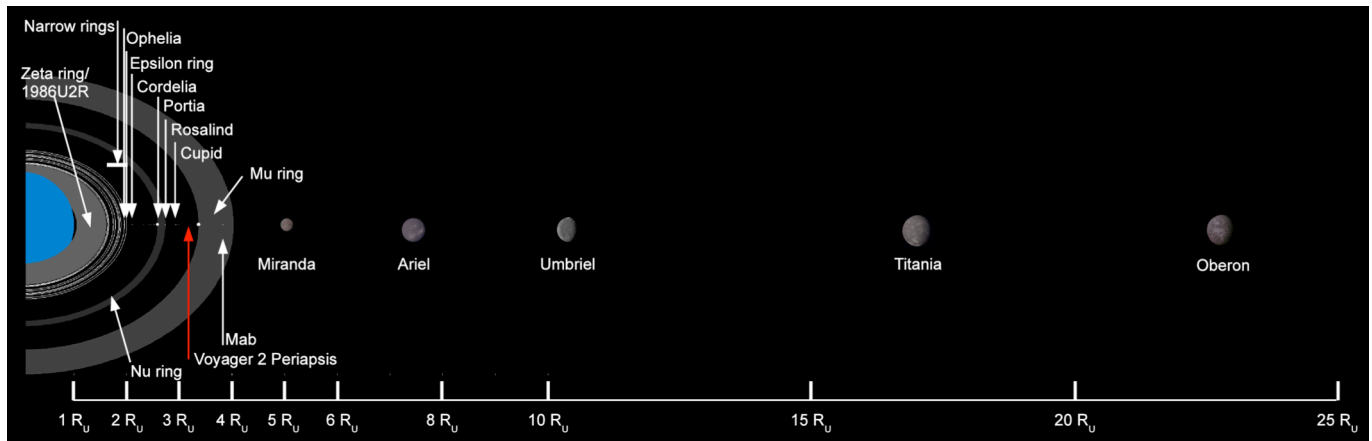


Figure 5: Illustration of Uranus' system of natural satellites and rings.

driven by charged particle precipitation and/or the way in which varying magnetospheric configurations couple with the upper atmosphere to produce time-variable fields and currents. Mapping temperatures, electron densities, and the distributions of ions and molecules in the ionosphere and thermosphere using UV and IR remote sensing will permit an unravelling of the thermospheric heating problem and will provide evidence for auroral activity in response to varying solar activity.

1.2 An Ice Giant Planetary System: Rings and Natural Satellites

Uranus has a rich planetary system of both dusty and dense narrow rings, and regular and irregular natural satellites. This unique example of a planetary system holds an important key to help us unravel the origin and evolution of the Solar System. Ground-based observations have found changes in the rings and satellites since the Voyager 2 flyby, indicating that fundamental instabilities in the coupled ring-moon system are of clear importance for understanding the evolution of planetary systems (de Pater et al., 2007). Figure 5 illustrates some of the main features of rings and natural satellites.

The study of the moons and rings of Uranus – in particular their composition and dynamical stability, their subsurface and deep interior structure, and their geological history and evolution and how that relates to their formation – are important parts of the Cosmic Vision goal for understanding the how the Solar System works. The possibility that Uranus' irregular satellites are captured Centaurs or comets can also contribute to understanding small primitive bodies and may provide lessons for our understanding of the origin of life in the Solar System, particularly since objects exposed to the solar wind are subjected to very different space weathering processes than those protected from the solar wind.

Little is known about the composition of the rings, partly because Voyager could not detect them in the near infrared, and understanding this composition would provide significant constraints on planetary evolution models. High spatial-resolution imaging of the narrow rings is needed to unravel the dynamics of their confinement and to confirm theories of self-maintenance and of shepherding by moons, which are relevant to other disk systems including protoplanetary disks. Also of interest are the rings' interaction with Uranus' extended exosphere and their accretion/disruption interplay with the nearby retinue of small moons. The ring system has also changed significantly since the Voyager flyby in ways we do not understand (Showalter and Lissauer, 2006) and new rings and satellite components have been discovered. These need to be characterised at close range in order to understand how their rapid evolution fits into various paradigms of Solar System evolution. Voyager's single high-phase image of the rings revealed a plethora of otherwise unknown dust structures (figure 6); more details of their structure and a first understanding of their evolution would be immeasurably valuable.

The five largest moons of Uranus (Miranda, Ariel, Umbriel, Titania, Oberon – see figure 7) are comparable in sizes and orbital configurations to the medium-sized moons of Saturn. They are, however, characterised by larger mean densities, about 1.5 g cm^{-3} on average, and by different insolation patterns, with their poles directed towards the Sun during solstice, owing to the large axial tilt of the planet. The moons are inside Uranus' magnetosphere; hence space weathering modifies surface properties and causes particle to be ejected from the surface. The observations performed during the flyby of Voyager 2 revealed surprising amounts of geological activity on these moons, possibly involving cryovolcanic processes. Miranda exhibits striking structural geology, despite its small size (472 km in diameter), with ridges and grooves that may be the result of internal differentiation processes (Janes and

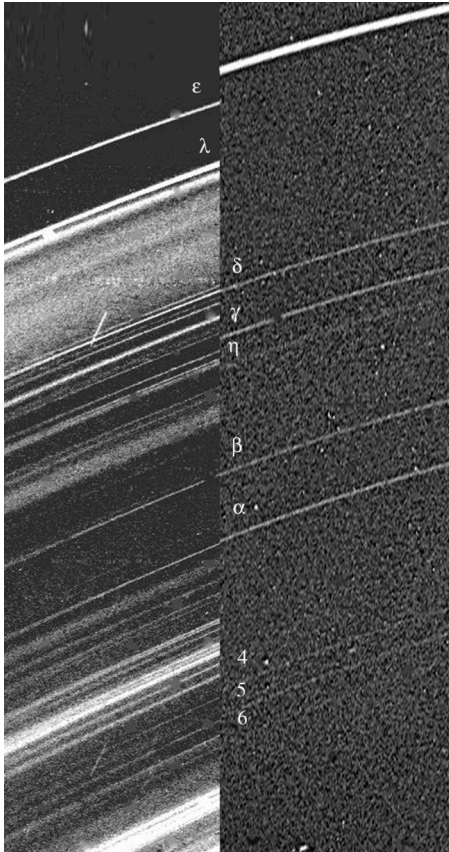


Figure 6: Composite image of the ring system in forward-scattered (left) and back-scattered (right) light. Credit: NASA/JPL.



Figure 7: Voyager 2 images of the largest natural satellites to scale. Credit: Paul Schenk.

Melosh, 1988) or the surface expression of large-scale upwelling plumes (e.g. Pappalardo et al., 1997). Similar internal processes possibly occurred on Enceladus in the saturnian system, before its intense surface activity and cryovolcanic plumes developed. Observations of Miranda thus provide a unique opportunity to understand how small moons can become so active (Castillo-Rogez and Lunine 2012). Moreover, the convex floors of Ariel's graben may provide the only evidence for widespread cryovolcanism in the form of viscous extrusive cryolava flows (Croft and Soderblom, 1991; Schenk, 1991), a process that has been elusive in the Solar System, with only a few small examples documented elsewhere to date, for example, Sippar Sulcus on Ganymede (Schenk and Moore, 1995) and Sotra Patera on Titan (Lopes et al., 2013). However, only very limited observations were possible during

Voyager 2's brief encounter, at which time only the southern hemispheres of the satellites were illuminated. The diversity of the medium-sized icy satellites at Uranus demonstrates the complex and varied histories of this class of object.

1.2.1 What is the composition of the uranian rings?

The composition of the uranian rings is almost entirely unknown, as Voyager 2 did not carry an infrared spectrometer capable of detecting the rings. However, it is clear from their low albedo that at least the surfaces of the ring particles are very different from those in Saturn's rings, and must have a significant non-water-ice component. The particle-size distribution of Uranus' main rings is also mysterious, with a surprising lack of cm-size particles detected by the Voyager 2 radio occultation (French et al., 1991). The composition of outer Solar System bodies is known to be diverse, likely reflecting a diverse array of processes shaping their surfaces. A Uranus orbiter will enable high-resolution near-infrared observations of the rings and small moons which will constitute a significant advance in our understanding of the evolution of the uranian system and may shed light on that of the outer Solar System in general. Mapping the spatial variations of both composition and particle size will clarify phenomena such as pollution and material transport within the system. Stellar, solar and radio occultations will enable the determination of the ice-fraction and size distribution of ring particles. A dust detector can directly determine from *in-situ* measurements the number densities as well as the speed and size-distributions of dusty ring material. Moreover, a chemical analyzer subsystem can provide unique information on the composition of these grains, bearing the possibility to constrain isotopic ratios of the constituents (Briois et al., 2013). Because larger ring particles and the uranian satellites are the sources of the dust, dust measurements give direct information on the composition of these bodies.

1.2.2 How do dense rings behave dynamically?

The main rings are eccentric and inclined and generally bounded by sharp edges (see reviews by Elliot and Nicholson, 1984; French et al., 1991). Although theories exist regarding these characteristics, including resonant interactions, "shepherding" by nearby satellites, and self-maintenance, the mechanisms are far from understood. Our understanding of these mechanisms is highly relevant to other disc systems, including protoplanetary and debris discs. Existing data give preliminary hints that self-gravity wakes and spiral density waves, which are important diagnostics as well as driving phenomena in Saturn's rings (see, e.g., Cuzzi

et al., 2010), also exist in at least some parts of Uranus’ rings, but much more detailed observation is needed to characterise them.

The rings of Uranus are the best natural laboratory for investigating the dynamics of dense narrow rings, an important complement to the dense broad disk exemplified by Saturn’s rings, and diffusive rings at Jupiter and Neptune (Tiscareno, 2013). These observations will undoubtedly reveal many new structures and periodicities, and possibly new moons that play important roles in ring confinement. Rings can also shed light on the planet’s gravitational and magnetic fields as well as the influx of interplanetary meteoroids (e.g., Hedman and Nicholson, in press). High-resolution images of the rings from a number of orbits and phase angles is needed in order to unravel their dynamics.

1.2.3 How do Uranus’ dusty rings work?

The Cassini mission has taught us that dusty rings are shaped by solar radiation forces, which depend on particle properties (size, albedo, etc.), as well as by the gravitational influence of satellites. Thus, a study of the dynamical structure of dusty rings will unveil much about the particles’ currently unknown material properties.

The post-Voyager discovery of the ν ring is especially intriguing, as this dusty ring lies between the orbits of two closely-packed satellites, but does not itself have any apparent source (Showalter and Lissauer, 2006). It is quite possible that the ν ring is the remains of a moon that was disrupted by a collision fairly recently. The innermost dusty ζ ring appears to have moved several thousand km outward between the Voyager 2 flyby and recent Earth-based observations (de Pater et al., 2007), but this changing ring has not been studied closely. Finally, Voyager’s single high-phase image of the rings revealed a plethora of otherwise unknown dust structures (Murray and Thompson, 1990). The bright bands and gaps in this dusty region are difficult to reconcile with conventional theories. High-resolution images of these dusty rings will allow us to determine their structure and evolution. Detailed observations may reveal one or more large source objects for this dusty region with possible evidence of accretion among them. In-situ detection with a dust detector, together with radio and plasma wave observations, allow a direct measurement of the local dust density. A dust detector can also provide information on the size-distribution and the distribution of orbital elements of the grains, as well as on their charging state, which might be key to understand the individual and collective dynamics of micron-sized particles.

1.2.4 How do the rings and inner satellites interact?

The inner moons of Uranus contain the most densely-packed known satellite system, with nine objects on orbits ranging from 59100 to 76400 km from the planet’s centre. This densely-packed system appears to be subject to mutual collisions on timescales as short as $\sim 10^6$ yr (Duncan and Lissauer, 1997; Showalter and Lissauer, 2006; French and Showalter, 2012), and several moons show measurable orbital changes within a decade or less, raising important questions regarding the origin, evolution, and long-term stability of the Uranus system. Lying immediately exterior to Uranus’ main ring system, but outside the “Roche limit” so that collisional products are able to re-accrete into new moons, these uranian inner satellites both interact with the rings (as well as with each other) and comprise a parallel system, a natural laboratory in which the effects of collisional disruption and re-accretion can be studied. The moon Mab lies at the centre of the μ ring, which shares with Saturn’s E ring the unusual characteristic of a blue colour likely due to a preponderance of monodisperse small particles (de Pater et al., 2006). However, while Enceladus creates the E ring by means of a fine spray of water crystals escaping from geysers, Mab seems much too small (~ 50 km across) to plausibly sustain any internal activity; it is, however, important to note that the same was formerly said of Enceladus. Mab also exhibits large unexplained deviations in its orbit (Showalter et al., 2008). Close observations of the surface of Mab, as well as its orbit and its interaction with the μ ring, are certain to yield significant discoveries on the evolution of coupled ring-satellite systems. Astrometric imaging of the uranian inner moons would significantly contribute to understanding this system, identifying resonant and chaotic interactions that can explain its current workings and past history.

1.2.5 What is the origin of the rings/satellite system?

The close packing of Uranus’ small moons and its ring system has suggested that there could be a genetic link between the two. Colwell and Esposito (1993) have suggested that Uranus’ rings may be the debris of moons destroyed by the meteoroid bombardment over the age of the Solar System. The giant impact theory for Uranus’ large obliquity also provides a mechanism for producing the rings from a disruption of the original satellite system (Coradini et al., 2010). More recently it has been suggested that tides themselves may destroy moons and create the rings (Leinhardt et al., 2012). These scenarios are similar to recent suggestions that Saturn’s, Uranus’ and Neptune’s satellites systems may result from ring evolution (Crida and Charnoz 2012). These scenarios

would imply the existence of a cycle of material between rings and moons. Since Uranus' ring/moon system evolves on timescales as short as decades, *in situ* tracking of this evolution would be a formidable opportunity to study this cycle, which may be at work also for Neptune and Saturn, but on longer time-scales for these systems. By leading a comparative study of spectral characteristics of the rings and moons, we may unveil the origin of both the satellites and rings by inferring whether they are made of the same material or not.

1.2.6 What is the composition of the uranian moons?

The albedos of the five major satellites of Uranus, varying between 0.21 and 0.39, are considerably lower than those of Saturn's moons (except Phoebe and the dark hemisphere of Iapetus). This reveals that water ice, which dominates their surfaces, is mixed in varying proportions to other non-ice, visually dark material.

Given the absence of a near infrared spectrometer in the payload of Voyager 2, no detailed information is available on the surface chemistry of the icy moons. Just to give a few examples, there is no indication about the chemistry of the structural provinces identified on the surfaces of Titania and Oberon, exhibiting different albedos and different crater density that reveal different ages. Similarly unknown is the nature of dark material (perhaps rich in organics) that fills the floors major impact craters on Oberon, as well as the composition of the annulus of bright material that is enclosed in the large crater Wunda on Umbriel. The chemical nature of the flows of viscous material observed on Ariel and Titania is also unknown, and we lack a certain indication of the presence of ammonia hydrate on the surface of Miranda, which has been suggested on the basis of telescopic observations (Bauer et al., 2002).

By using an imaging spectrometer in the near infrared range from 1 μm to at least 5 μm , it will be possible to unveil the surface composition of the moons by identifying and mapping various chemical species (with particular emphasis on non-water-ice materials, including volatiles and organics). This will ultimately enable an unprecedented correlation of surface composition with geologic units at various spatial scales. A spatially resolved chemical mapping will also help separating the relative contributions of endogenic subsurface chemistry and exogenic magnetosphere-driven radiolysis across the moons, and assess the role of processes that exchanged material between the surface and subsurface.

1.2.7 What is the origin of Uranus' moons and how have they evolved?

As in the jovian and saturnian systems, tidal and magnetospheric interactions are likely to have played key roles in the evolution of the uranian satellite system. For instance, intense tidal heating during sporadic passages through resonances is expected to have induced internal melting in some of the icy moons (Tittlemore and Wisdom, 1990; Tittlemore, 1990). One such tidally induced melting event may have triggered the geological activity that led to the late resurfacing of Ariel. The two largest moons, Titania and Oberon, with diameters exceeding 1500 km, might still harbour liquid water oceans between their outer ice shells and inner rocky cores, remnants of past melting events (Hussmann et al. 2006).

The surfaces of the five major satellites of Uranus exhibit extreme geologic diversity; however, understanding of their geologic evolution and tectonic processes has suffered greatly from incomplete Voyager image coverage (imaging restricted to the southern hemispheres) and only medium to low image resolutions (order of several kilometres per pixel, except for part of Miranda) which only allow characterization of the largest geologic units in the areas that could be imaged by Voyager (e.g., Croft and Soderblom, 1991). The crater size-frequency distributions of the five satellites, used as a tool for dating surface features and for constraining impactor origin, are known only for the southern hemispheres and crater sizes larger than a few kilometres (e.g. Plescia, 1987). There are also still large uncertainties in the bulk composition of the moons (e.g. Hussmann et al., 2006), which provide fundamental constraints on their origins.

High-resolution images of the satellite surfaces, which will provide key information on the ages and compositions of the surfaces and will constrain the dynamical and geologic histories that led to the observed diversity. For example, Miranda and Ariel exhibit evidence of significant endogenic geological activity. High-resolution surface mapping will enable us to determine the degree to which tectonic and cryovolcanic activity has occurred, permitting characterisation of the role played by tidal dissipation and understanding whether uranian moons have experienced internal activity similar to that at Enceladus. Mapping of the moons will help constrain the nature and timescale of this activity, and characterizing the environment in their vicinity may reveal outgassing if, as at Enceladus, activity is continuing. Collisional activity amongst the irregular satellites can produce contamination of the regular satellite surfaces with material from the irregular satellites via dust transport

(Schubert et al., 2010). High-resolution imagery might reveal evidence of such processes.

Accurate astrometric measurements can also be used to quantify the influence of tidal interactions in the system at present, providing fundamental constraints on the dissipation factor of Uranus (Lainey et al., 2008). Gravimetric and magnetic measurements, combined with global shape data, will greatly improve the models of the satellites' interiors, bringing fundamental constraints on their bulk composition (density) and evolution (mean moment of inertia). Understanding the composition (particularly the ice-to-rock ratio) and the internal structure of the natural satellites will also enable us to understand if Uranus' natural satellite system was the original population of bodies that formed around the planet, or if they were subsequently disrupted, potentially via a giant impact that might have produce Uranus' large obliquity (Coradini et al., 2010).

Crater statistics will be crucial in determining the satellites' geological histories as well as projectile flux in the outer Solar System. Near- and mid-infrared spectroscopy will enable us to understand the surface composition of the moons yielding further information on their origin and evolution. Occultations will enable us to probe any tenuous atmospheres that may be present and UV spectroscopy may then lead constraints on their chemistry, with implications for the subsurface. The dayside magnetopause lies at a distance of $18 R_U$ and, therefore, the major moons are embedded within the magnetosphere. This implies that their water-ice surfaces are eroded by magnetospheric charged particles in addition to photons and micro-meteoroids. Measuring the properties of the charged particles that these moons can encounter and the energetic neutral particles released after the ions impact the surface will constrain the role of plasma bombardment on surface evolution. These data will constitute strong constraints to allow us to understand how satellite systems form and evolve around Ice Giants. The composition of the moons will represent an essential data point in understanding the nature and origins of organic and volatile material in the outer Solar System.

Recent models of icy satellite interiors suggest the larger uranian satellites, Titania and Oberon, may contain subsurface oceans (Husmann et al., 2006) and Miranda may be subject to recent or even ongoing activity (Castillo-Rogez and Turtle 2012). The magnetic field induced in Europa's subsurface ocean was readily detectable by Galileo (e.g., Khurana et al., 1998) and any such signatures at Uranus are expected to be strong due to Uranus' asymmetrical field.

Remote observations of Uranus' irregular satellites can be used to search for potential genetic relationships with

the irregular satellites found in other giant planet systems and thus understand the evolution of Solar System minor bodies and giant planet natural satellites.

1.3 Uranus' Aeronomy, Aurorae, and Highly Asymmetrical Magnetosphere

The configuration of all the planetary magnetospheres in the Solar System is determined by the relative orientations of the planet's spin axis, its magnetic dipole axis, and the solar wind flow. In the general case, the angle between the magnetic dipole axis and the solar wind flow is a time-dependent quantity and varies on both diurnal and seasonal timescales. Uranus presents a particularly interesting and poorly understood case because this angle not only varies seasonally but because of Uranus' large obliquity the extent of diurnal oscillation varies with season. At solstice this angle does not vary with time and Uranus' magnetic dipole simply rotates around the solar wind flow. This is a magnetospheric configuration not found anywhere else in the Solar System. Figure 8 illustrates the configuration of Uranus' magnetosphere near equinox, as sampled by Voyager 2.

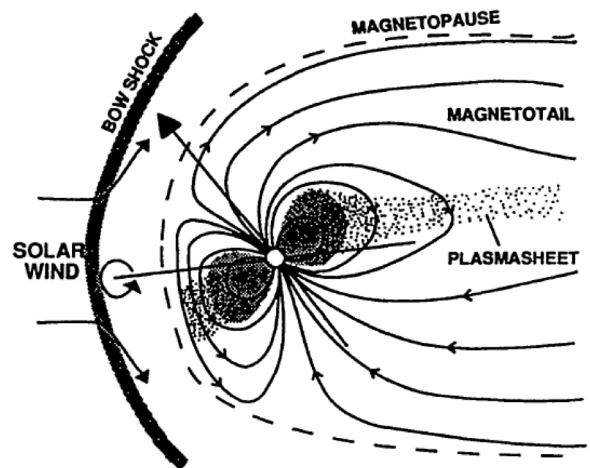


Figure 8: Schematic of the uranian magnetosphere near equinox (Bagenal, 1992).

Because of this unique extreme orientation, its magnetosphere is expected to vary from a pole-on to orthogonal configuration during a uranian year and to change from an “open” (connected to the solar wind) to a “closed” configuration during a uranian day. Such a rapidly reconfiguring magnetosphere with a highly asymmetric internal magnetic field (section 2.1.3) at its core provides a challenge for our theories of how magnetospheres work and will bring new insight in fundamental and universal magnetospheric processes. Uranus also presents a special case because of its distant location in the heliosphere where the properties of the

solar wind are very different to the near-Earth environment. This provides opportunities to investigate fundamental processes such as magnetic reconnection under a different parameter regime. Furthermore, in order to further our understanding of how life and the platforms for life exist in the wide variety of magnetic environments in the Universe it is vital that we make comprehensive measurements in the widest possible variety of environments. These aspects make a study of Uranus' magnetosphere a very important objective for understanding how the Solar System works and for achieving ESA's Cosmic Vision goals. These are not only relevant for the important question of understanding how asymmetric Ice Giant magnetospheres work, but are also highly relevant in providing "ground-truth" for understanding exoplanet magnetospheres.

Along with the planetary magnetic field, the ionosphere of Uranus is the internal core of the magnetosphere. Models indicate that Uranus' ionosphere is dominated by H^+ at higher altitudes and H_3^+ lower down (Capone et al., 1977; Chandler and Waite, 1986; Majeed et al., 2004), produced by either energetic particle precipitation or solar ultraviolet (UV) radiation. Recent analysis of observations of H_3^+ emissions from Uranus spanning almost 20 years (Melin et al., 2011), have revealed a phenomenon that is not seen at the other Gas Giants in our Solar System. The temperature is strongly correlated with season, such that at solstice, the upper atmosphere is more than 200 K hotter than at equinox. It seems likely that a key component of the required additional heating is driven by particle precipitation and/or the way in which varying magnetospheric configurations couple with the upper atmosphere.

Auroral emissions are also generated at kilometric (radio) wavelengths (1-1000 kHz), which cannot be observed from Earth or distant observers. As at other planets, UKR is thought to be generated by the Cyclotron Maser Instability. However, UKR appears to be more complex than similar radio emissions at Earth, Saturn or Jupiter and only comparable to Neptune's ones. Understanding the circumstances under which these peculiar radio emissions are generated is of prime

importance for the ground-based radio detection of exoplanets with a magnetic field (essential to the development of life), particularly those with highly inclined magnetic axis with respect to the stellar flow.

1.3.1 What is the overall configuration of the uranian magnetosphere?

Our understanding of the uranian magnetosphere is currently essentially limited to data from the Voyager 2 flyby which provided a single snapshot where the angle of attack between the solar wind axis and the magnetic dipole axis varied between 68° and 52° , to some extent similar to the Earth's magnetosphere. However, the near alignment of the rotation axis with the planet-Sun line during solstice means that plasma motions produced by the rotation of the planet and by the solar wind were effectively decoupled (Vasyliunas, 1986). Therefore, in contrast with Jupiter and Saturn, solar wind plasma may be able to penetrate deep within the magnetosphere despite the planet being a fast oblique rotator. This may result in short residence times for magnetospheric plasma produced deep within the magnetosphere and may limit amount of plasma trapping inside the magnetosphere and consequently the amount of charged particle acceleration. Proton and electron radiation belts (with energies up to tens of MeV) albeit slightly less intense than those at Saturn were also observed in the inner magnetosphere of Uranus (Cheng et al., 1991) but their diurnal and seasonal variability is largely unknown.

The significant asymmetries in the magnetosphere results in large-scale diurnal reconfigurations of the system on timescales of hours resulted in a twisted magnetotail topology (Tóth et al., 2004). The main plasma sources, transport modes and loss processes in the uranian magnetosphere, and the modes of interaction (pick-up, sputtering, and charge exchange) between the magnetospheric plasma and the rings and moons of Uranus are also largely unknown. The configuration and dynamics of the uranian magnetosphere at equinox are entirely unknown and it is not clear if this will result in a fairly quiescent magnetosphere such as Neptune, or a more rotationally dominated magnetosphere like Jupiter or Saturn.

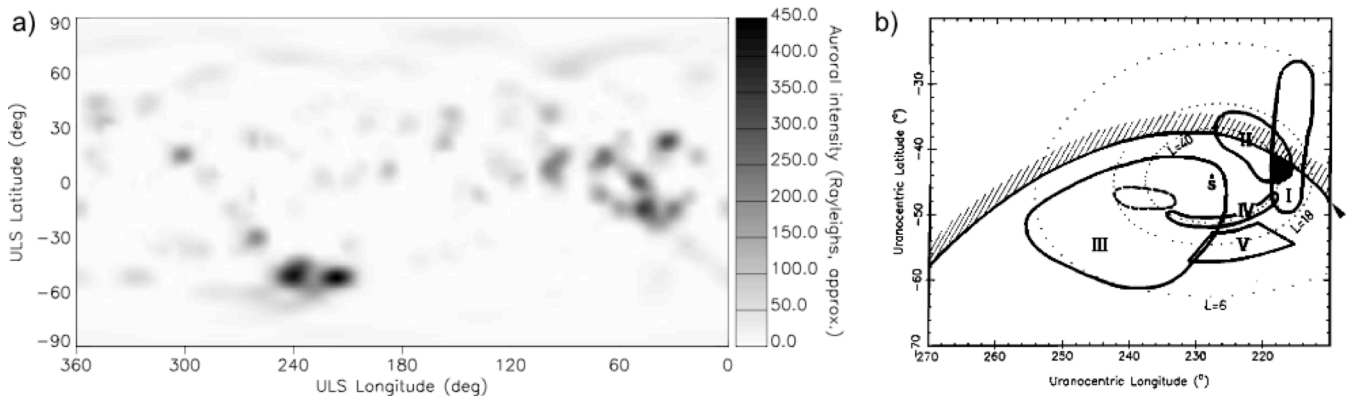


Figure 9: H₂ band emission map of Uranus' aurorae from Voyager 2 (Herbert, 2009) (left). Source regions for the most intense UKR component (Zarka and Lecacheux, 2007).

1.3.2 What are the characteristics and origins of the uranian aurorae

Aurorae are the most striking diagnosis of the magnetosphere dynamics, as they can be traced back to the currents generated by the magnetospheric interactions. Several kinds of interactions have been characterised at Earth, Jupiter and Saturn, but the Uranus optical and radio aurorae, as they are known from Voyager 2 observations seem to indicate new kinds of interactions. The charged particles responsible for both optical and radio auroral emissions and their source regions are also unknown. A study of the uranian auroral regions can also lead to information on the thermosphere due to atmospheric sputtering produced by auroral particle precipitation. Such sputtered particles can be monitored by a neutral particle detector.

There has only been one spatially resolved observation of the UV aurora of Uranus (Herbert, 2009), using a mosaic of Voyager 2 UV observations mapping emission from H Lyman- α and EUV H₂ band emission (Figure 9, left). The emission appeared patchy and was generally centred on the magnetic poles, with the emission being the brightest about midnight magnetic local time. There have been subsequent attempts to observe the aurora in both the FUV using the Hubble Space Telescope (HST) (Ballester et al., 1998) and in the IR using ground-based telescopes (e.g., Trafton et al., 1999). Uranus' aurorae was recently redetected in the UV using HST (Lamy et al., 2012) and revealed a radically different set of auroral processes controlled by the interaction between the magnetosphere and the solar wind, and raising important questions on the generation of planetary auroral emissions and possible secular drift of Uranus' intrinsic magnetic field.

The UKR components, which indicate different active regions in the magnetosphere, divide into two categories: (i) “bursty” (<10 min) emissions comparable

to that at Earth and Gas Giants, and (ii) “smooth emissions” time-stationary emissions (lasting for hours) that are specific to Ice Giants. These latter components require a continuous source of free energy that has not yet been identified and is apparently maintained in a highly variable magnetosphere (Figure 9, right). New radio observations with a modern instrumentation will provide wave properties that were inaccessible to Voyager 2, such as the wave direction and polarisation. Continuous remote observations of UKR and in situ measurements within their various source regions will provide essential information to understand the origin and characteristics of the variety of known uranian radio components and check for new ones.

Recent calculations show that new ground-based radio telescopes could detect radio emissions from hot Jupiters (Zarka et al., 2007). Unlike our Solar System, eccentric and complex orbital characteristics appear to be common in other planetary systems, so that the understanding of radio emission produced by Uranus could have profound importance in interpreting future radio detections of exoplanets.

1.3.3 How does magnetosphere-ionosphere-Solar Wind coupling work at Ice Giants?

The uranian magnetosphere interacts with a fast magnetosonic Mach number and high-beta solar wind, which is an important plasma regime in which to understand magnetic reconnection. Evidence of dynamics, similar to Earth-like substorm activity but possibly internally-driven, was also reported at Uranus by Mauk et al. (1987) which indicate that important energy sources need to be quantified, including the energy input from the solar wind. We do not know how the solar wind-magnetosphere interaction is interrupted and modulated by the diurnally changing geometry. Understanding how the aurorae of Uranus respond to changes in the solar wind is essential to understanding the Solar Wind interaction with giant planets more

generally. While these responses are well studied for the Earth the situation for the outer planets is less well understood, due to the lack of dedicated deep space solar wind monitors. Recent theoretical work (Cowley, in press) has argued for distinct differences in magnetotail processes between equinox and solstice thus providing a framework for the interpretation of new auroral images and demonstrating the need for new in situ measurements. The magnetosphere of Uranus was observed to be the site of intense plasma-wave activity with remarkably intense whistler mode emissions (Kurth et al., 1991). The role of wave-particle interactions for the magnetosphere-ionosphere coupling and the generation of Uranus' auroral emissions, as well as for the overall energy budget of the magnetosphere requires further consideration.

1.4 Cruise phase science in the outer heliosphere

A mission to Uranus naturally involves a relatively long duration interplanetary transfer. However, this presents an opportunity to undertake studies of the outer heliosphere, minor Solar System bodies, and fundamental gravitational physics.

1.4.1 Physics of the interplanetary medium

The structure of the heliosphere originates in the structure of the solar magnetic field and is strongly modified by the solar corona. There are a range of important questions on how this structure is further modified and processed in the heliosphere and goes on to modulate cosmic ray flux in the inner heliosphere, on the generation of turbulence, and how minor bodies interact with the heliosphere. One of the major issues of the physics of interplanetary medium is to understand the mechanisms of energy dissipation. Injected with large spatial scales by the Sun, the energy is transferred to smaller scales (ion/electron), where it is dissipated as heat. Measurements made by the Voyager probes have revealed variations of the exponents of the power law of certain parameters (eg, speed, magnetic field, density) with distance from the Sun, suggesting regime change in the process of energy transfer (Burlaga et al. 1997). Few observations of the heliospheric environment beyond 10 AU have been made since Pioneer 10 and 11, Voyagers 1 and 2, and New Horizons with very few observations made at solar maximum. Energetic particle observations during cruise will facilitate further study of the interaction between the outer heliosphere and interstellar medium, as carried out by Cassini at 9.5 AU and IBEX at 1 AU.

1.4.2 Fundamental physics and departures from General Relativity

General Relativity is in good agreement with most experimental tests of gravitation. But experimental tests leave open windows for deviations from this theory at short (Antoniadis et al., 2011) or long (Reynaud and Jaekel, 2005) distances. General Relativity is also challenged by observations at galactic and cosmic scales. The rotation curves of galaxies and the relation between redshifts and luminosities of supernovae deviate from the predictions of the theory. These anomalies are interpreted as revealing the presence of so-called “dark matter” and “dark energy”. Their nature remains unknown and, despite their prevalence in the energy content, they have not been detected up to now by other means than gravitational measurements.

Given the immense challenge posed by these large scale observations, in a context dominated by the quest for the nature of dark matter and dark energy, it is important to explore every possible explanation including the hypothesis that General Relativity is not the correct description of gravity at large scales (Aguirre et al., 2001; Nojiri and Odintsov, 2007). Testing gravity at the largest scales reachable by therefore essential to bridge the gap between experiments in the Solar System and astrophysical or cosmological observations (Turyshev, 2008). Combining radio-science and acceleration measurements not only improves the precision and quality of spacecraft navigation but also allows us to remove, as fully as possible, the systematic effects of non-gravitational forces acting on the spacecraft (Iafolla et al., 2010). These scientific goals are intimately connected since gravitation is directly connected to planetary ephemeris (Fienga et al., 2010) as well as to the origins of the solar system (Blanc et al., 2005).

2 Strawman Mission Concept

The primary trade space in mission options is between an orbiter and a flyby mission. Some goals can be partially satisfied with a flyby mission but to full answer the questions laid out in section 1 requires an orbiting platform to make repeated observations of Uranus and its planetary system. There exists an additional trade space between enhanced remote sensing instrumentation and an entry probe. But some science questions (1.1.1/1.1.5/ 1.1.6) can only be answered with an atmospheric entry probe to a >5 bar depth. For the purposes of this Uranus white paper the outline mission concept consists of an orbiter in a polar science orbit with an atmospheric entry probe.

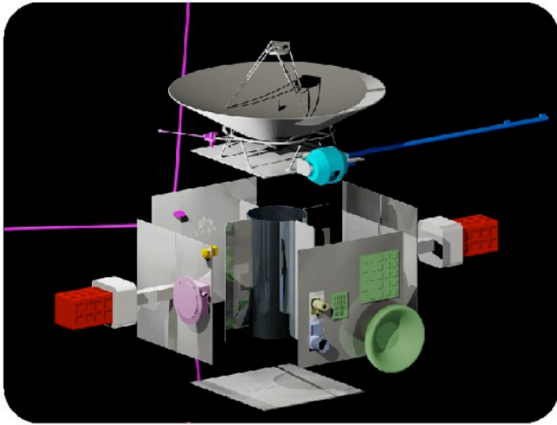


Figure 10: Illustration of the Uranus Pathfinder spacecraft. Credit: P. Dunn/C.S. Arridge.

Key items in the mission concept summary are presented in the table in the executive summary, and various items discussed in detail below, including critical items and technology development requirements. Figure 10 illustrates the strawman configuration for Uranus Pathfinder spacecraft as proposed for ESA M3 call in 2010 (Arridge et al., 2012) and which is a reasonable strawman concept for an L-class Uranus orbiter. The adoption of an L-class design over an M-class mission permits the inclusion of an atmospheric entry probe, an expanded suite of instruments, and additional fuel which greatly increases the science return. Table 2 illustrates the strawman instrument suite, composed of high TRL instruments with excellent European heritage.

2.1 Interplanetary transfers and orbital entry

Interplanetary transfers to Uranus have been studied in a number of mission analyses (Arridge et al., 2012; Hubbard et al., 2010) and demonstrate the feasibility of a mission to Uranus with current technology and including an interplanetary transfer between 10 and 16 years. The range of acceptable periapsis latitudes and radial distances at Uranus orbit insertion are limited due to the largely unknown ring plane hazards. This can be mitigated with a high latitude periapsis and orbit insertion manoeuvre followed by a ring plane crossing beyond 52000 km, inside of which are the main ring plane hazards. Although aerocapture is a natural technology to use at orbit insertion, the atmosphere of Uranus is poorly understood and aerocapture is low TRL technology, thus representing a high risk option.

Uranus' large obliquity naturally results in a polar science orbit which is ideal for studies of Uranus' interior, atmosphere and magnetic field that are required to meet the goals in section 1.

2.2 Atmospheric entry probe

An atmospheric entry probe for Uranus has been studied by the ESA Concurrent Design Facility, which led to a 312 kg entry probe (including 20% system margin) using a dedicated carrier platform. The mission concept we outline would involve using the Uranus orbiter as a carrier and communications relay. The instrumentation for such an entry probe is all available within Europe and is high TRL. The key technology development requirement is the thermal protection system for the entry probe.

2.3 Critical issues

2.3.1 Electrical Power

The key technology development requirement for a mission to Uranus is the provision of sufficient electrical power at 19.2 AU. Scaling ESA's Rosetta mission solar arrays out to Uranus we estimate that providing 400 W_e at Uranus would require 800 m² solar arrays producing system level issues associated with a large launch mass and spacecraft moment of inertia. At present a nuclear (radioisotope) power source (RPS) is the only viable alternative. ²⁴¹Am is the isotope that has been selected for ESA RPS devices that are currently in the developmental stage (see Arridge et al. (2012) for a discussion of issues relating to the use of ²⁴¹Am). To provide target electrical power of 400 W_e at Uranus after 14 years flight time would require a total RPS system mass of 200 kg based on an RTG design (with a specific power of 2.0 W/kg, compared with 2.3 W/kg for a NASA MMRTG using ²³⁸Pu). Although the development of such technology presents a schedule and cost risk, this is currently under development via ESA contracts and should be available and high TRL before the proposed L2/L3 launch windows.

2.3.2 Thermal control

Thermal control is an important driver for every mission. Extreme differences in thermal environment between the inner heliosphere and Uranus, and due to the continuous supply of thermal energy from RPS units present the most important issues. Such thermal control issues can be adequately managed by modifying existing designs from Rosetta and Mars/Venus Express. We have studied thermal control for a Uranus mission using ThermXL, based on a spacecraft of a similar size to Mars Express and including heat dissipation from the RPS, and estimate that electrical heaters consuming around 50 W will be sufficient to maintain an internal spacecraft temperature of -30° against losses to space. Waste electrical power from the RPS can be dissipated via externally- or internally-mounted shunt resistors and spot heating might be provided by radioactive heating

Instrument	Heritage
	Orbiter
Magnetometer	Cassini/MAG Solar Orbiter
Plasma and Particle Package	Rosetta/RPC-IES New Horizons/PEPPSI
Radio and Plasma Wave Experiment	Cassini/RPWS Bepi-Colombo/MMO/PWI
Microwave radiometer	Juno/MWR
Thermal Infrared Bolometer	LRO/Diviner BepiColombo (detectors)
Visual and Near-Infrared Mapping Spectrometer	New Horizons/RALPH Rosetta/VIRTIS Dawn/VIR
Ultraviolet Imaging Spectrometer	BepiColombo/PHEBUS Mars Express/SPICAM-UV
Visible Camera	Mars Express/SRC New Horizons/LORRI
Radio Science Experiment	Venus Express/VeRa Rosetta/RSI
Accelerometer	CHAMP/STAR
Dust detector	Cassini/CDA
	Probe
Mass spectrometer	Huygens/GCMS Galileo/GPMS
Nephelometer	Galileo/NEP
Radio science	Huygens/DWE
Accelerometer	Huygens/HASI

Table 2: *Strawman scientific payload.*

units based on ^{241}Am , thus lessening the demands for electrical heating.

2.3.3 Telemetry rates

To answer the questions in section 1 requires significant volumes of data to be returned over ≈ 20.9 AU. Downlink transmissions over Ka-band to ESA's Cebreros station, using a 4m (3m) high gain antenna on an orbiter with a pointing accuracy of 0.05° (comparable to the Cassini orbiter) will achieve a downlink rate of 5.6 (3.6) kbps, equivalent to 160 (100) Mbit per 8 hour downlink. These data volumes should be sufficient to achieve the essential science goals.

2.3.4 Long cruise phase duration

To reduce cruise phase costs a Uranus mission can employ hibernation modes (similar to those used on New Horizons and Rosetta) to minimise operations costs and deep space antenna usage. A cruise phase science programme, as outlined in section 1, will periodically enable the platform and science instruments to be utilised and tested. In addition, special hibernation modes would permit some instruments to collect low-

rate cruise phase science data. The use of high TRL technology and minimising the cruise phase operations will reduce demands on spacecraft platform components, will reduce the mission cost-at-completion, and lessen demands on the electrical power system.

2.4 International cooperation

Such a large and significant interplanetary mission would naturally benefit from collaboration with other space agencies. The List of Supporters shows broad support within NASA and JAXA, and within Europe. Uranus has been named a priority by NASA as recommended by the NRC Planetary Decadal Survey. In the context of international cooperation, a partner agency may provide an atmospheric entry probe, provide instruments for the orbiter/entry probe thus lessening the demand on ESA member states, or may provide a launch vehicle.

3 Acronyms

AOCS	Attitude and Orbit Control System
FUV	Far Ultraviolet
EUV	Extreme Ultraviolet
HGA	High Gain Antenna
HST	Hubble Space Telescope
IBEX	Interstellar Boundaries Explorer
IR	Infrared
ISO	Infrared Space Observatory
MMRTG	Multimission Radioisotope Thermoelectric Generator
NRC	National Research Council
RPS	Radioisotope Power Source
RTG	Radioisotope Thermoelectric Generator
TRL	Technology Readiness Level
UKR	Uranian Kilometric Radiation
UV	Ultraviolet
VLA	Very Large Array
VLT	Very Large Telescope

4 References

- Aguirre et al. (2001) *Classical Quant. Grav.* **18**, 223.
- Antoniadis et al. (2001) *C. R. Phys.* **12**(8), pp. 755-778.
- Arridge et al. (2012) *Exp. Astron.* **33**(2-3), pp. 753-791.
- Bagenal (1992) *Annu. Rev. Earth Planet. Sci.* **20**, pp. 289-328.
- Ballester et al. (1998) *Proc. Conf. Ultraviolet Astrophysics, beyond the IUE Final Archive*, Sevilla, Spain, 21.
- Bauer et al. (2002) *Icarus* **158**, pp. 178-190.
- Blanc et al. (2005) Tracing the origins of the Solar System, In: Favata, F., Sanz-Forcada, J., Giménez, A., Battrick, B. (eds.) 39th ESLAB Symposium on Trends in Space Science and Cosmic Vision 2020, ESA Special Publication, **588**, p. 213.
- Boué and Laskar (2010) *Ap. J.*, **712**, L44-L47.
- Briois et al. (2013) *LPI* **1719**, 2888.
- Burlaga et al. (1997) *JGR* **102**, pp. 4661-4672.
- Capone et al. (1977) *ApJ* **215**, 977-983.
- Castillo-Rogez and Turtle (2012) *DPS* #44, #104.02.
- Castillo-Rogez and Lunine (2012) 43rd LPSC, abstract #1707.
- Chandler and Waite (1986) *GRL* **13**, 6-9.
- Cheng et al. (1991) Energetic particles at Uranus, in *Uranus*, U. Arizona Press.
- Christensen & Tilgner (2004) *Nature* **429**, 169-171.
- Colwell and Esposito (1993) *JGR* **98**(E4), pp. 7387-7401.
- Coradini et al. (2010) *Space Sci. Rev.* **153**, pp. 411-429.
- Cowley et al. (in press) *J. Geophys. Res.*
- Crida and Charnoz (2012) *Science* **338**(6111), pp. 1196.
- Croft & Soderblom (1991) Geology of the Uranian satellites, in *Uranus*, U. Arizona Press.
- Cuzzi et al. (2010) *Science* **327**, 1470-1475.
- Duncan & Lissauer (1997) *Icarus* **125**, 1-12.
- Elliot & Nicholson (1984) The rings of Uranus, in *Planetary Rings*, 25-72, U. Arizona Press.
- Feuchtgruber et al. (2013) *A&A* **551**, A126.
- Fienga et al. (2010) Gravity tests with INPOP planetary ephemerides, In: Klioner, S.A., Seidelmann, P.K., Soffel, M.H. (eds.) *IAU Symposium*, **261**, pp. 159-169, doi:10.1017/S1743921309990330 (arXiv:0906.3962).
- Fortney & Nettelmann (2010) *SSR* **152**, 423-447.
- French et al. (1991) Dynamics and structure of the uranian rings, in *Uranus*, 327-409, U. Arizona Press.
- French and Showalter (2012) *Icarus* **220**(2), pp.911-921.
- Fressin et al. (2013) *ApJ* **766**, p. 81.
- Guillot (2005) *AREPS* **33**, 493-530.
- Hammel et al. (2006) *ApJ* **644**(2), pp.1326-1333.
- Hedman and Nicholson (in press) *Icarus* (arXiv:1304.3735).
- Helled et al. (2010) *Icarus* **210**(1), 446-454.
- Herbert (2009) *JGR* **114**(A11), A11206.
- Holme and Bloxham (1996) *JGR* **101**, 2177-2200.
- Hubbard et al. (1995) The Interior of Neptune, In: Hubbard, Podolak and Stevenson, *Neptune and Triton*, pp. 109-138. U. Arizona Press.
- Hubbard et al. (2010) Ice Giants Decadal Study, http://sites.nationalacademies.org/SSB/SSB_059331, accessed 23 May 2010.
- Hussmann et al. (2006) *Icarus* **185**(1), 258-273.
- Iafolla et al. (2010), *Plan. Space Sci.* **58**, 300-308.
- Jacobson et al. (1992) *ApJ* **103**, 2068-2078.
- Jacobson (1998) *ApJ* **115**, 1195-1199.
- Jacobson (2007) *DPS Abstracts*, **39**, 23.06.
- Janes & Melosh (1988) *JGR* **93**, 3127-3143.
- Karkoschka & Tomasko (2009) *JGR* **202**(1) 287-309.
- Khurana et al. (1998) *Nature* **395**(6704) pp. 777-780.
- Kurth et al. (1991) Wave-particle interactions in the magnetosphere of Uranus, in *Uranus*, U. Arizona Press.
- Lainey (2008) *Planet. Space Sci.* **56**(14), pp. 1766-1772.
- Lamy et al. (2012) *GRL* **39**, L07105.
- Leinhardt et al. (2012) *MNRAS* **424**(2), pp. 1419-1431.
- Lopes et al. (2013) *JGR* **118**, pp. 416-435.
- Majeed et al. (2004) *Adv. Sp. Res.* **33**(2), pp.197-211.
- Melin et al. (2011) *Astrophys. J.*, **729**, 134.
- Mauk et al. (1987) *JGR* **92**, 15283-15308.
- Morbidelli et al. (2012) *Icarus* **219**, p. 737.
- Murray and Thompson (1990) *Nature* **348**, 499-502.
- Ness et al. (1986) The magnetic field and magnetospheric configuration of Uranus, in *Uranus*, U. Arizona Press.

- Nettelmann et al. (2013) PSS **77**, 143-151.
- Nojiri and Odintsov (2007) Int. J. Geom. Methods Mod. Phys. **4**(1), 115–145.
- Owen & Encrenaz (2006) PSS **54**, 118-1196.
- Pappalardo et al. (1997) JGR **102**(E6), 13369-13380.
- de Pater & Lissauer (2010) Planetary Sciences. CUP.
- de Pater et al. (2006) Icarus **180**(1), pp.186-200.
- de Pater et al. (2007) Science **317**(5846), 1888.
- Pearl et al. (1990) Icarus **84**(1), 12-28.
- Plescia (1987) JGR **92**, 14918-14932.
- Reynaud and Jaekel (2005) Int. J. Mod. Phys. A **20**, 2294.
- Safronov (1972) Evolution of the protoplanetary cloud and formation of the earth and planets. Translated from Russian. Jerusalem (Israel): Israel Program for Scientific Translations, Keter Publishing House, p. 212.
- Saumon & Guillot (2004) ApJ **609**(2) 1170-1180.
- Schenk (1991) JGR **96**, pp.1887-1906.
- Schenk and Moore (1995) JGR **100**, pp. 19009-19022.
- Schubert et al. (2010) Space Sci. Rev. **153**, pp. 447-484.
- Showalter et al. (2008) DPS Abstracts 40, 24.07
- Showalter and Lissauer (2006) Science **311**(5763) pp. 973-977.
- Soderlund et al. (2013) Icarus **224**, pp.97-113.
- Sromovsky et al. (2012),
<http://www.news.wisc.edu/newsphotos/uranus2012.html>, accessed 24 May 2013.
- Stanley & Bloxham (2004) Nature **428**, 151-153.
- Stanley & Bloxham (2006) Icarus **184**, 556-572.
- Tiscareno, M.S. (2013) Planetary Rings, in Planets, Stars and Stellar Systems, by Oswalt, French and Kalas, Springer (arXiv:1112.3305).
- Tittemore & Wisdom (1990) Icarus **85**, 394-443.
- Tittemore (1990) Icarus **87**, 110-139.
- Tóth et al. (2004) JGR **109**, A11210.
- Trafton et al. (1999) ApJ **524**(2), 1059-1083.
- Turrini et al. (2011) MNRAS, **413**, 2439.
- Turyshev (2008) Annu. Rev. Nucl. Part. Sci. **58**, 207-248.
- Vasyliunas (1986) GRL **13**(7), 621-623.
- Walsh et al. (2011) Nature, **475**, 206.
- Zarka (2007) PSS **55**(5), 598-617.

Master

A Mission to Return a Sample From Mars to Earth

Submitted by

Monica M. Grady

The Open University
Milton Keynes MK7 6AA

Monica.Grady@open.ac.uk

+44 1908 659 251

On behalf of the Master Consortium

24th May 2013

List of Contributors

Given Name	Surname	Institution	Country	E-mail
Francesca	Altieri	IAPS/INAF	Italy	francesca.altieri@iaps.inaf.it
Ed	Chester	University of Leicester	UK	ec65@leicester.ac.uk
Lewis	Dartnell	UCL	UK	l.dartnell@ucl.ac.uk
Yang	Gao	University of Surrey	UK	yang.gao@surrey.ac.uk
Peter	Grindrod	UCL	UK	p.grindrod@ucl.ac.uk
Sanjeev	Gupta	Imperial	UK	s.gupta@imperial.ac.uk
Patrick	Harkness	Univ Glasgow	UK	Patrick.Hartness@glasgow.ac.uk
Katie	Joy	Univ Manchester	UK	katherine.joy@manchester.ac.uk
Elliot	Sefton-Nash	UCLA	USA	esn@ucla.edu
Caroline	Smith	NHM	UK	caroline.smith@nhm.ac.uk
Stephanie	Werner	Univ Oslo	Norway	stephanie.werner@geo.uio.no
Ian	Wright	Open University	UK	ian.wright@open.ac.uk
Susanne	Schwenzer	Open University	UK	susanne.schwenzer@open.ac.uk
Frances	Westall	CNRS	France	westall@cnrs-orleans.fr

A list of more than 100 supporters can be found at

https://docs.google.com/spreadsheet/ccc?key=0AiUDByeS_FwdG1XcFBPa2VEWWp3OW5IMHBXNWhueEE#gid=0

They come from: Austria, Belgium, Finland, France, Germany, Italy, Morocco, Netherlands, Norway, Spain, Sweden, Switzerland and UK

Executive Summary

We present a concept, *Master* (Mars to Earth) for a mission to return samples of rock and dust from Mars. The mission duration is 3 years, including 1 year in martian orbit. Material will be acquired by a grab technique, and returned in a gas-tight canister to the Earth. The canister will also contain a sample of ambient martian atmosphere, collected concurrently with the solid material. It is anticipated that ~ 150 g of material, with a grain size of ~ 0.02 to 2 cm will be recovered. Following a 6 month preliminary examination, during which rigorous planetary protection procedures will be followed, material will be distributed to the wider scientific community for detailed analysis. The mission will be scientifically exciting as well as technologically challenging, and will arouse widespread public interest.

Introduction

Mars is a small rocky planet, with a diameter about half that of Earth. It has a thin atmosphere (~ 6 mbar), mostly of carbon dioxide. Because it is readily visible (by telescope) in the night sky, Mars has been an object of fascination and scientific study for almost 400 years. In the modern era of space exploration, Mariner 9 (in 1971) was the first mission to orbit Mars and send back tantalising images of volcanoes (Figure 1a) and features that looked to have been carved by rivers. Since then, a series of spacecraft have orbited the planet and returned numerous images of channels, gullies, impact craters and volcanoes, indicating the active thermal (Figure 1b) and hydrological (Figure 1c) history that Mars has experienced.

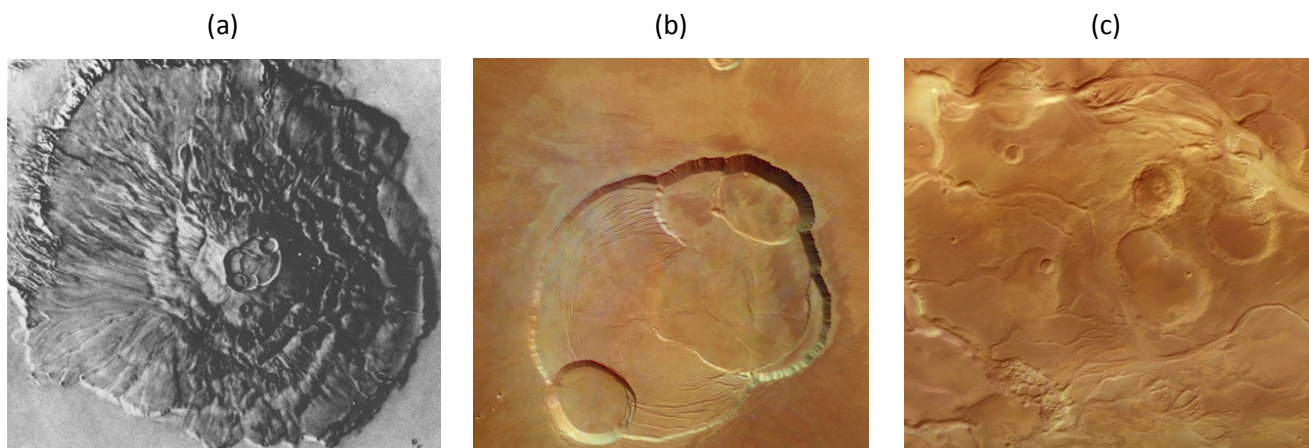


Figure 1: Images of the martian surface from orbit. (a) the first picture of Olympus Mons (Mariner 9, 1971); Complex crater at the summit of Olympus Mons (Mars Express, 2004); (c) Fluvial channels at Mangala Valles (Mars Express, 2004); Image (a), copyright NASA; (b), (c), copyright ESA.

Complementing images taken from orbit are those recorded by several landing craft. The Viking landers pictured frost coating the surfaces of boulders strewn across a dust-covered plain (Figure 2a). The angular nature of the boulders and the ubiquity of the dust implied that erosional processes were still active on Mars' surface, even if the weathering taking place was through frost-action and ablation by wind-blown dust, rather than by water or ice. The Spirit and Opportunity rovers were the first to take close-up images of individual martian rocks, showing the products of hydrous activity (Christensen and Ruff, 2004; Figure 2b). The most recent mission, NASA's Curiosity rover, is currently returning high resolution images of its landing site in Gale Crater, including the first detailed pictures of what are undoubtedly a sedimentary conglomerate – a rock composed of a mixture of water-worn pebbles of different sizes and compositions held together by a matrix of finer-grained material (Williams et al., 2013; Figure 2c).

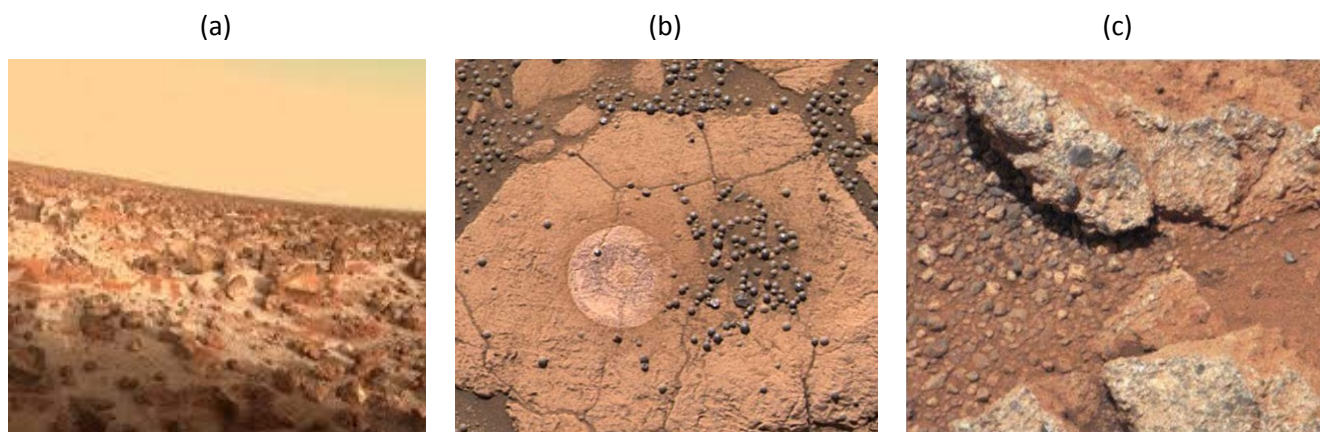


Figure 2: Images of the martian surface from the ground. (a) Frost of water ice coating the ground at Utopia Planitia (Viking 2, 1979); (b) Close-up of haematite spherules at Meridiani Planum (Opportunity, 2004); (c) Outcrop of conglomerate at the Link site in Gale Crater (Curiosity, 2012). Images copyright NASA.

One of the problems of tracing an evolutionary history for Mars is the lack of absolute ages for different lithological units. Crater counting over the landscape is the method by which relative ages are fixed (Tanaka, 1986), thus allowing dimension of martian history into the various epochs (Noachian, Hesperian, Amazonian), but the absolute time at which one epoch merged into another is unknown. Comparison with the lunar cratering record (calibrated with dates measured on Apollo samples) allows broad timings to be fixed (Hartmann and Neukum, 2001). The approximate age ranges of the different martian epochs are given in Table 1. One of the most important objectives of the **Master** mission is to assign absolute ages to rocks returned from Mars, so that we have a properly calibrated chronology for the planet – something that currently only exists for the Earth and the Moon.

Table 1: Martian Epochs, based on crater counts in the Tharsis region (Tanaka, 1986; Hartmann & Neukum, 2001; Head et al., 2001)

Epoch	Age (Ga before present)
Noachian	4.65-3.7
Late Noachian – Early Hesperian	3.8-3.6
Early Hesperian	3.7-3.6
Late Hesperian – early Amazonian	3.6-2.1
Middle – Late Amazonian	2.1-0

Images of the landscape are matched by spectroscopic data, which have been used to map the distribution of minerals over the surface (Figure 3). What has been particularly exciting is the discovery of clay minerals – species produced by the alteration of primary basaltic rock by water (Mustard et al., 2008). Several different generations of clay minerals can be identified, and from such data we have been able to build up a better appreciation of Mars' history (Bishop et al., 2008). Water is not stable on Mars' surface today, because the atmospheric pressure is too low. However, the existence of morphological features apparently produced by flowing water, accompanied by secondary minerals indicate that at times in its past, Mars has had a very active fluvial history. This, in turn, implies that Mars must have had a much thicker atmosphere in its past, one that

allows liquid water to be stable (Clifford and Parker, 2001). The surface temperature when water was present must also have been higher than it is now (although it still would not have been warm). The conventional interpretation was that an early thick atmosphere (up to several bars) of CO₂ allowed liquid water to flow on Mars (e.g., Pollack et al., 1987). The inference that carbon should occur as carbonates in the martian crust and soils is confirmed to limited extent, and carbonates have indeed been identified by emission spectroscopy as present in the dust (Bandfield et al., 2003), although no massive carbonate deposits have been identified at the surface (Bibring et al., 2005; Orofino et al., 2009). The presence of sulphur at the martian surface (Baird et al., 1976), the subsequent identification of sulphates *in situ* (Squyres et al., 2004) and the evidence for the action of brines (Bridges et al., 2001) seems to suggest highly acidic conditions on early Mars (Hurowitz et al., 2006), in which case massive carbonate deposits might never have formed. Orbital imagery has shown features on Mars' surface that appear to have been carved by fluid (presumed to be water and/or ice), and secondary minerals produced by water have been identified in soils (Poulet et al., 2005).

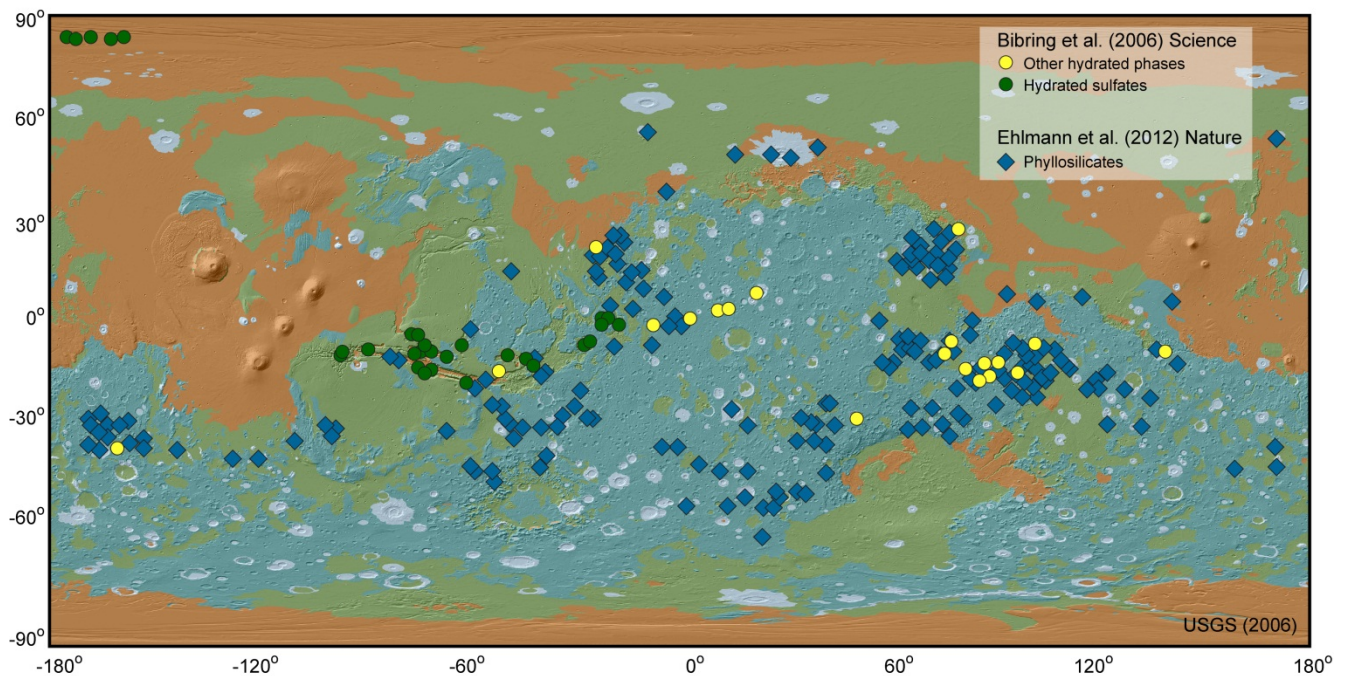


Figure 3: Distribution of minerals produced by hydrous alteration across the surface of Mars, as recorded by the OMEGA instrument on Mars Express (Bibring et al., 2006) and CRISM on Mars Reconnaissance Orbiter (Ehlmann et al., 2012).

Although recognition of a fluvial history for Mars has been accepted for several decades, it is only with the advent of high resolution imaging and spectroscopy from orbit that secondary products from the action of water have been mapped (Figure 3). Clay minerals and sulphates and, to a lesser extent, carbonates have been found across the surface, allowing a mineral stratigraphy to be established (Bibring et al., 2006). The secondary alteration history of Mars can be related to Mars' chronology, as defined by crater counting (Figure 4).

Part of the fascination with Mars, and the continued thrust towards its exploration, is the potential that the planet has to host life. Mars is the most Earth-like planet in the Solar System, and it formed at the same time as the Earth, and from similar materials. It is likely that 4.5 Gyr ago, Mars was very like the Earth. Both Earth and Mars experienced bombardment by asteroids and comets, maintaining a molten surface and heated atmosphere. However, once the bombardment decreased, and the planets cooled and differentiated, their evolutionary pathways diverged almost immediately (Zahnle, 2006). Where Earth retained an atmosphere and acquired a hydrosphere, Mars very quickly lost the bulk of its atmosphere, and thus had no extensive surface hydrosphere. The tectonic history of the planets also differs. Earth's crust is broken into plates, the movement of which has led to regular overturn and recycling of crustal material, such that no trace of the first generation of differentiated crustal material remains. In contrast, Mars' surface seems to have been almost static: evidence for a limited extent of plate tectonics came from magnetic measurements recorded by Mars Global

Surveyor (Connerney et al., 1999). The most active phase of Mars' tectonic history, presumably including incipient plate tectonics, had ceased by ~ 3.5 Gyr ago (Nimmo and Tanaka, 2005; Solomon et al., 2005), so Mars maintains a more complete record of its history, including that of the first billion years, which is all either missing from Earth's record or has been severely metamorphosed by subsequent processing.

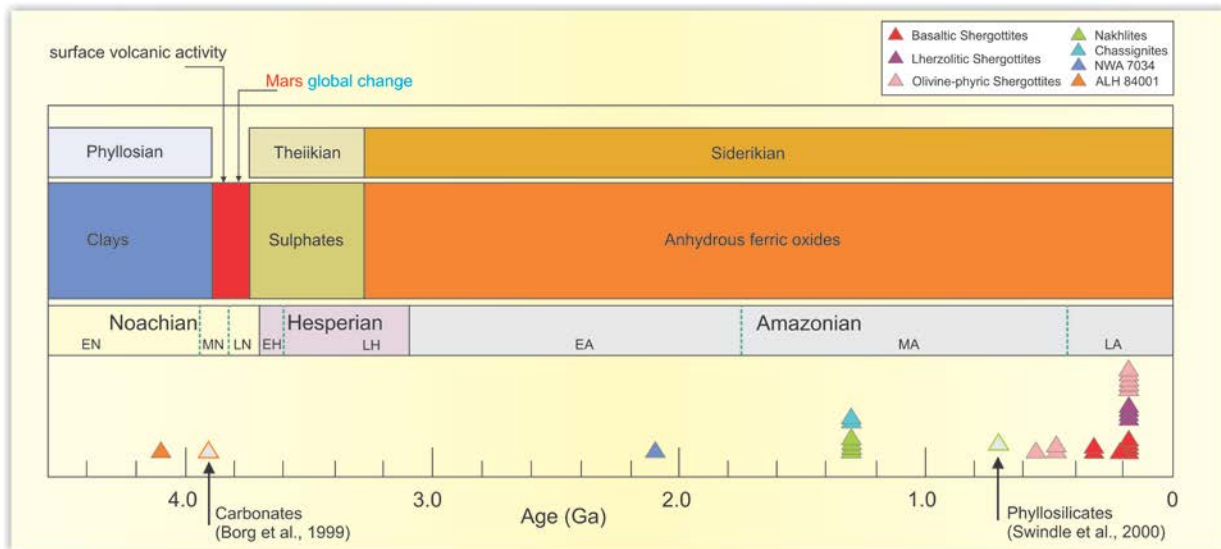


Figure 4: Relationship of the crystallisation history of martian meteorites to the chronology and stratigraphy of Mars. Lowest band: the three epochs of Mars; EN, MN, LN – Early, Middle and Late Noachian; EH, LH – Early and Middle Hesperian; EA, MA, LA – Early, Middle and Late Amazonian (Tanaka, 1986; Head et al., 1986). Upper two bands: stratigraphy of Mars from spectroscopic and image data (Bibring et al., 2006). Crystallisation ages of meteorites from (Nyquist L.E. et al., 2001; Lapen T.J. et al., 2010). **Mars global change** marks the posited switch from a warm and wet Mars to a cold and dry planet (Bibring et al., 2006).

Given that Mars had a thicker atmosphere in the past, plus flowing water on the surface, this opens up the possibility that living organisms arose on Mars in the same way that they originated on Earth (Brack et al., 2010). This early record of the time of formation of life on Earth has either been destroyed by plate tectonics and surface activity or rendered difficult to interpret and thus controversial. Therefore in many ways, through the material revealed at its surface today, Mars offers a clearer and longer record of its geological evolution than Earth, allowing access to time periods during which life was emerging on Earth. Although Mars is believed currently to be lacking in any type of biological activity at the surface (sterile because of UV and cosmic ray irradiation), it requires only a few cm of rocky overburden to radiation to a non-destructive level, such that microorganisms might be able to survive (Cockell et al., 2002). The past two decades has seen an expansion in understanding of the limits of the biological envelope on Earth, with microbes able to survive extremes of temperature and radiation (Cavicchioli, 2002). The retrieval of a life-bearing sample is not a goal or objective of the **Master** mission. It is anticipated, though, that careful selection of the Master landing site might enable collection of material from a potentially habitable environment that could contain traces of past life.

International Context and fit to ESA strategy

ESA's space exploration strategy for the decade 2015-2025 was defined in its Cosmic Vision document published in 2005. This forward-looking plan laid out the priorities that the ESA community recognised as key to furthering our understanding of the universe, and our place within it. The research programme is arranged in four themes; the mission that we propose here, to return a sample from Mars, addresses very specifically the questions of Theme 1 (Planets and Life). Indeed, a Mars sample return (MSR) mission was highlighted in the Cosmic Vision document as one of the main mission concepts that would aid in achieving the goals of the programme (<http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=38646>). The importance of a MSR mission is well-recognised by the international community: ESA discussed a Memorandum of Understanding with NASA in 2008 to collaborate on a series of missions building towards MSR, which in itself is a stepping

stone in the path to the eventual human exploration of Mars. Thus, we believe that a MSR mission will attract wide support from other space agencies, although the concept outlined here can be achieved as an L-class mission led by ESA alone.

Background to previous MSR mission proposals

Over the past decade, there have been several separate *ad hoc* international working groups established under the auspices of the International Mars Exploration Working Group (IMEWG) to define the science goals, technology requirements and logistics of a MSR mission (see <http://mepag.jpl.nasa.gov/> for archived copies of all the relevant reports). As outlined above, one of the drivers for such a mission is to determine whether life ever arose on Mars; a second driver is assessment of potential hazards for human exploration of Mars. All the working groups recognised that:

- i. A single sample return would not answer all the questions that must be addressed to establish whether Mars was ever inhabited;
- ii. A single sample return mission would not answer all the technology requirements and reduce the risk factors associated with the human exploration of Mars;
- iii. MSR is a series of missions, and missions currently under development can be considered as part of the technology and science build-up for MSR;
- iv. The mission does not end once a sample is returned to Earth: it must be subject to stringent planetary protection controls, accompanied by scientific examination, prior to material being distributed to the scientific community.

As a result of the above factors, designs for a MSR mission have become ever more elaborate, with complex scientific payloads incorporating rovers and a requirement for the ability to select and collect suites of samples from specific localities, including the capacity for drilling and coring into bedrock. With **Master**, we are taking a different approach: although it would be advantageous to return material retrieved from Mars' subsurface, and for the precise context of the sample's location to be known, we believe that a 'grab and go' scoop from Mars' surface will yield appropriate material to answer the specific goals listed in Table 2. These goals are not only scientific in nature, but also include technological and engineering challenges that must be addressed prior to human exploration of Mars. The goals are sub-divided into more specific objectives in Table 3. Whilst the search for life on Mars is a major interest and an important component of all martian exploration missions, it is not a primary goal of the **Master** mission. However, the information that will come from **Master** will lead to an improved general understanding of the martian surface, in turn furthering our understanding of the geological context in which any potential martian life-forms might exist.

Table 2: Scientific Goals of Master:

Goal	
1.	To determine the absolute age and composition of Mars' surface at the landing site, and to relate that age and composition to the evolution of the planet and its surface and atmospheric history
2.	To measure the composition of Mars' atmosphere and to quantify how Mars' atmosphere has changed since the planet's formation
3.	To investigate the surface of Mars for signs of past biological activity, and assess the habitability-potential of Mars' surface
4.	To consider the structure and composition of Mars' surface at the landing site and assess the hazards and advantages the site presents for future human exploration of Mars

Why bring a sample back ?

From the previous sections, it is very clear that we have learnt an enormous amount about Mars and its evolutionary history from spacecraft that have visited the planet. It is also clear from the results being achieved by instruments on the Curiosity rover that reasonably sophisticated analyses can be performed on the martian surface, producing high quality and reliable data. So why do we still require material to be brought back to Earth? There are several main reasons:

1. Samples can be analysed using instrumentation that it is not possible to deploy remotely (e.g., a synchrotron facility);
2. Samples can be prepared for analysis by complex techniques not possible to achieve remotely, e.g., separation into individual mineral grains; preparation of thin sections for microscopy; preparation of solutions for elemental and isotopic analysis;
3. An individual sample can be analysed several times by an array of different instruments, e.g., scanning electron microscopy followed by ion microprobe, followed by transmission electron microscopy
4. Repeat measurements can be made (by the same technique, or by different techniques, or by different laboratories) for verification of unusual or controversial results (e.g., potential detection of life)

As well as scientific reasons for bringing material back, there is the added benefit of having a valuable resource for educational and public outreach purpose. The Apollo samples still exert an interest and fascination with students and the general public, despite having been returned more than 40 years ago.

Why Meteorites from Mars are not sufficient to achieve Master's Goals

Information about the evolutionary history of Mars from spacecraft is complemented by results from direct analysis of martian meteorites¹. These are a group of (currently ~120) meteorites² with ages from 165 Ma to 4100 Ma, spanning most of Mars' history (Nyquist et al., 2001). The meteorites are igneous, and have been sub-divided into different types based on their mineralogy and age. Some of the meteorites contain secondary minerals formed in association with water on or near the martian surface (Figure 5). The secondary minerals include phyllosilicates (clay minerals), carbonates, sulphates and hydrated iron oxides. Compositionally, the primary magmatic minerals and secondary alteration species match reasonably well with some of the rocks identified on Mars' surface (Treiman et al., 2013). Notwithstanding the similarities so far identified between martian meteorites and the surface of Mars, results from the Spirit, Opportunity and Curiosity rovers have shown that there is a much greater variation in rock types on Mars than is represented in the martian meteorite collection (McSween et al., 2004; Clark et al., 2005; Wiens et al., 2013). Specifically, sedimentary rocks are completely absent from the suites of martian meteorites, so the extent, variety and ages of these important rocks have not been studied at the level of detail required to interpret Mars' alteration history.

One of the most significant differences between martian meteorites and *in situ* martian rocks is exemplified by the increasingly-detailed maps that we have of Mars' surface. We are now able to produce maps showing the distribution of different rock types across the surface, and their location relative to landscape features such as craters, valley networks, volcanoes etc. (e.g., Lichtenberg et al., 2010; Mangold et al., 2010; Ansan et al., 2011). Construction of these maps has enabled a stratigraphy of secondary deposits to be constructed, leading to a relative chronology based on alteration mineralogy, as opposed to the more traditional stratigraphy based on crater counting (Bibring et al., 2006). The clear sequences of secondary mineralisation, visible on the surface as deposits many km² in area, demonstrates a succession, presumably resulting from changes in climate and environmental conditions over periods of many thousands to millennia of years (Ehlmann and Mustard, 2012). This situation contrasts quite dramatically with the secondary mineral assemblages identified in martian meteorites, where almost the entire sequence of alteration species can occur within a single vein, representing

¹ Evidence that the meteorites come from Mars rests on their age and the gases trapped in them during excavation from the martian surface by impact (McSween, 1985).

² There are around 45,000 known meteorites, so the martian collection accounts for less than 0.3 % of the total (Grossman, 2013).

changes in fluid composition that have occurred over 10s to 100s of years (e.g., Bridges and Grady, 2001; Changela and Bridges, 2011).

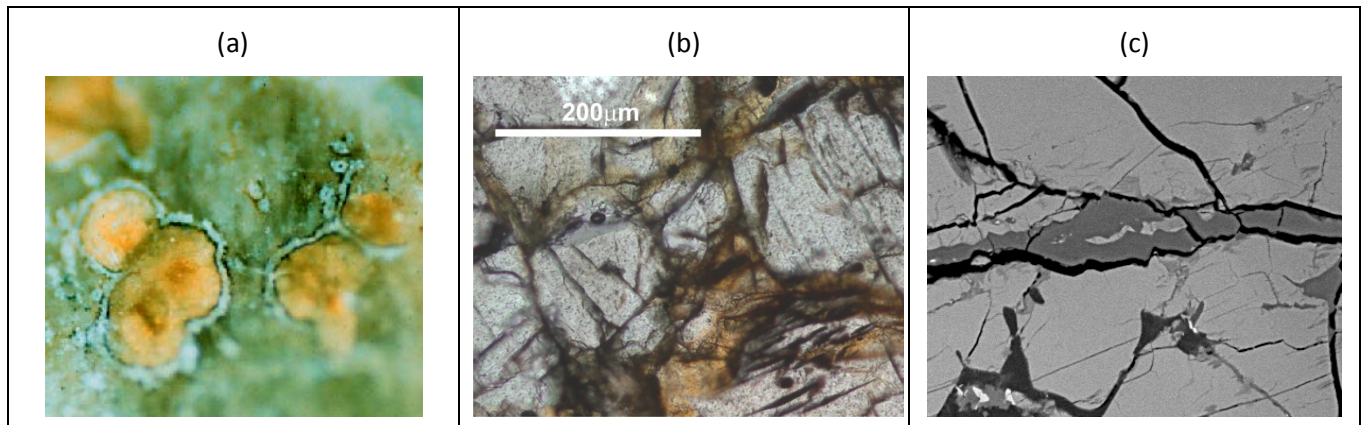


Figure 5: Secondary minerals in martian meteorites (a) Carbonates in a chip of ALH 84001. Reflected light. Field of view = 1000 μm ; (b) clay minerals in a thin section of Yamato 000593. Transmitted light. Field of view = 500 μm and (c) SEM image of an iron oxide vein in MIL 03346 image. Field of view = 100 μm . All images by MMG.

So, whilst we can gain great insight to martian processes through study of martian meteorites, they are by no means representative of the surface of Mars. We have no contextual information about their source regions on Mars or their geographical relationship to each other. Neither do we have any choice over the selection process of the rocks – they have been ejected from Mars’ surface by impact, and so have generally been altered by shock, followed by inter-planetary irradiation. Although there is a general agreement that the shergottites are apparently young (ages $< \sim 500$ Ma), there have been suggestions that the original crystallisation age of this group ranges between 4.3 and 4.1 Ga (Bouvier et al., 2008), and that the younger ages occur because of resetting by shock or fluid alteration (Bouvier et al., 2005). Attempting to solve this martian basalt age conundrum is difficult when there is no contextual information available about the source of the rocks. Hence the necessity for a sample return mission, which will enable, for the first time, age dating of martian material from a known location, leading to the definition of a chronology of the martian surface that ties absolute ages to both crater densities and mineralogy.

Science Objectives of a ‘Grab and Go’ Mission

The very first images taken at the surface of Mars were obtained by the Viking landers (Figure 2a). They showed a flat to gently undulating and dusty surface, peppered with angular rocky boulders of varying size. Over the three and a half decades since Viking, five more spacecraft have landed on Mars’ surface (Figure 6), and the quality and resolution of the data obtained have improved. We have a much greater understanding of the processes that have shaped Mars’ surface, and the variety of landscapes that result from those processes. What all the missions have shown, though, is that much of the surface is covered by a layer of fine-grained wind-blown dust. This is a seemingly homogeneous reservoir of the products of millions of years of erosional history, a mixture of all of the different rock types exposed at the surface (Yen et al., 2005). The Curiosity rover has determined a cumulative grain size distribution of ~ 1 mm to ~ 1 cm for the soil at its landing site in Gale Crater (Yingst et al., 2013)(Figure X). Chemical analyses of the soil by instruments on three different landers at three widely-separated landing sites have shown that it is homogeneous in composition over the planet’s surface (McSween et al., 2010; Yen et al., 2013). Therefore this fine-grained component of the martian regolith offers insight to the conditions and processes that have operated on a global scale. Mixed in with the dust are small pebbles. These are not the same composition at the different landing sites, but are likely to reflect the composition of the local bedrock and therefore provide a valuable record of conditions and processes in each specific region. The pebbles range from ~ 1 cm upwards to metre-sized boulders.

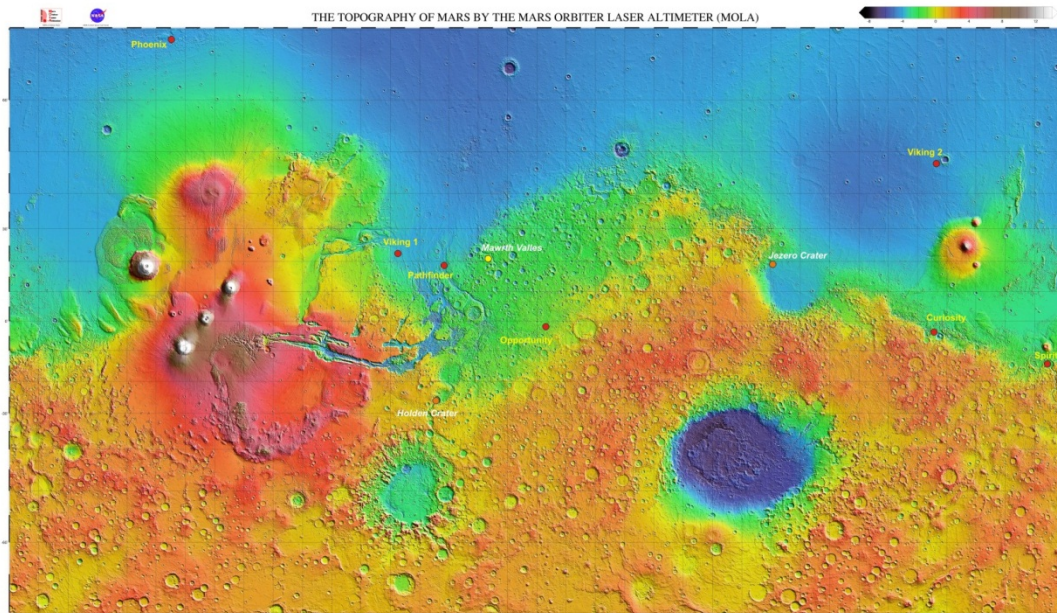


Figure 6: Topographic map of Mars, showing where previous missions have landed (yellow text, red dots). The three locations marked with white text and yellow dots are landing sites that were considered for MSL, and are potential future landing sites for **Master**.

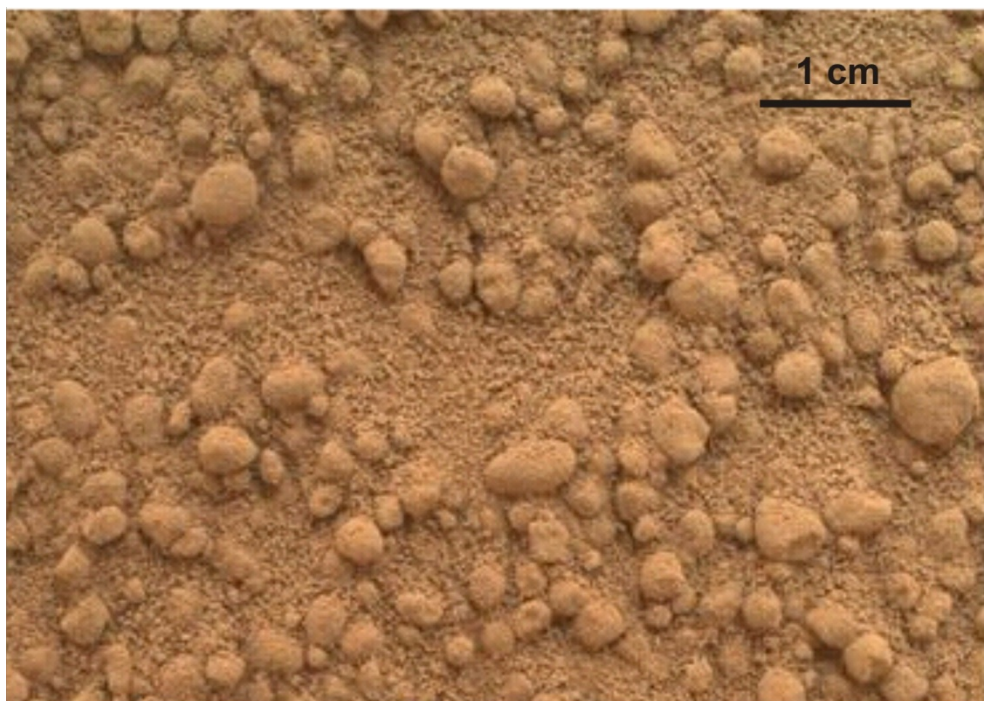


Figure 7: Image from the Mars Hand Lens Imager (MAHLI) camera of soil beside the Curiosity rover at the Rocknest site in Gale Crater. Small particles of martian regolith (dust, rock fragments) can be seen on the sub centimetre scale. Image Credit: NASA/JPL-Caltech/MSSS.

From the information we now have about the martian regolith, it is clear that a sample collected from the surface will be a mixture of both locally- and globally-derived material (pebbles and dust, respectively). Analysis of such material *in situ* is limited by the precision of the instruments deployed; it is also limited by the size of the sample that can be analysed. In contrast, analysis of samples on Earth can be undertaken on a grain-by-grain basis. Advances in laboratory instrumentation now allow measurements of grains only a few microns across. The return of a 'grab' sample of regolith from a carefully selected landing site, one that has been identified by orbital data perhaps as containing clay minerals, will answer many of the outstanding

questions concerning the evolution of the martian surface. Table 3 summarises specific objectives that a ‘grab’ sample would be designed to achieve; they are keyed to the goals given in Table 2.

Table 3: Science Objectives for a regolith (R) sample recovered by the **Master** mission

Objectives (R)	Goal
R(i). Determine the age of rock fragments and soil from a specific region of Mars	1
R(ii). Determine the mineralogy, petrology and geochemistry of rocks and dust at the landing site	1
R(iii). Determine the range of compositional and age variations within material recovered from the landing site	1
R(iv). Assess the erosional and depositional or crystallisation history of material recovered from the landing site	3, 4
R(v). Examine the surface of individual grains to search for signs of weathering or alteration	3, 4
R(vi). Assess material in terms of <i>in situ</i> resource utilisation (e.g., implanted hydrogen) for future human exploration possibilities	4

The second sample that will be returned from the mission is that of Mars’ atmosphere. There is a two-fold reason for returning an atmospheric sample: not only will its analysis assist with achievement of the science goals outlined in Table 2, its presence will stabilise the returned regolith. Material scooped from the surface will be placed into an opened canister. The solid sample will be in equilibrium with Mars’ atmosphere at the surface; by keeping the sample under the same pressure of atmosphere during the return stages of the mission, it is hoped that potential reactions between minerals will be prevented. So, for example, any hydrated sulphates present have the capacity to dehydrate, liberating water that might react with other species, e.g., perchlorate salts in the sample. Any such reactions or phase changes impact on interpretation of the mineralogy and chemistry of the regolith, as they will alter the nature of the material originally collected.

The scientific objectives to be addressed through analysis of a returned sample of martian atmosphere are given in Table 4, again keyed to the goals of the **Master** mission.

Table 4: Science Objectives for an atmosphere (A) sample recovered by the **Master** mission

Objective	Goal
A(i). Determine the abundance and elemental and isotopic compositions of the martian atmosphere, especially of the noble gases	2
A(ii). Establish limits for trace species (e.g., methane) that may be indicative of biological activity	3
A(iii). Establish limits for transient species (e.g., ozone, hydrogen peroxide) that may be indicative of oxidative activity	4

How much material should be returned?

The amount of material to be returned is related to engineering considerations, such as the mass and volume of the sample return canister etc. It is also governed by the requirements of the different techniques that will be used to analyse the samples in order to address the science goals of the mission. Fortunately, over the years, whilst laboratory instrumentation has become complex, there has been a concomitant decrease in the amount of material required for analysis. The return of material from Comet Wild II (the Stardust mission) and from Asteroid 25143 Itokawa (the Hayabusa mission) demonstrated that excellent and extensive science could be achieved on a few small grains of material. Despite continuing advances in instrumentation, there are, however, techniques that require more than a single grain for analysis. For example, one of the outstanding unknowns about Mars' surface is the abundance, distribution and speciation of carbon. Although the Curiosity rover has looked for carbon at its landing site in Gale Crater, it has not yet found anything above the detection limit of its on-board instrumentation. Thus if carbon is present (and it should be, just from asteroidal and cometary dust infall), more sophisticated detectors are required, such as are routinely used to measure carbon-bearing compounds in terrestrial laboratories.

Not all the material will be consumed by scientific analysis: planetary protection protocols call for ~ 25 % of any sample to be used for biohazard assessment. Most of the analyses under this requirement yield scientific results, but some material will have to be sacrificed for the assessment. Good curation practice also calls for ~ 30% of the material to be saved for posterity – for the development of new, as yet unforeseen measurements and instrumentation. And some material (possibly as much as 30%, but utilising as much as possible of the material that has been through the planetary protection procedures) will be used for education and public outreach activities.

Regolith

Bearing in mind that we will have one sample canister, with a volume of $\sim 1500 \text{ cm}^3$, and that sufficient space must be available for the associated atmosphere sample, a mass of 150 g would be an appropriate quantity of material to return. Assuming a mixture of unconsolidated dust plus cm-sized pebbles, giving an approximate density of 3 g cm^{-3} , then 150 g would occupy a volume of 50 cm^3 . Depending on the size of the scoop, the sampling arm might have to perform repeat manoeuvres to acquire sufficient material.

The return of such a (relatively) large sample would allow material to be distributed throughout the scientific community. It would also allow replicate, high precision analyses of the different components within the material, as well as a detailed organic analysis.

Atmosphere

Assuming a sample return vessel with a volume of 1500 cm^3 , containing 50 cm^3 of solids, then 1450 cm^3 atmosphere can be returned (taking no account pore space). In order to keep the mission profile as uncomplicated as possible, the atmosphere will be collected at ambient pressure, and not trapped or concentrated. Assuming a surface pressure of 6 mbar and temperature of 273 K, then $\sim 0.4 \times 10^{-3}$ moles of gas will be returned. This is sufficient for replicate analysis of the least abundant species, such as krypton, and to obtain precise values for $^{14}\text{N}/^{40}\text{Ar}$ and $^{40}\text{Ar}/^{36}\text{Ar}$ ratios, which are diagnostic of atmospheric evolutionary processes.

Sample Return Canister

The most important features of a sample return canister are that it has a gas-tight lid (to prevent both escape of martian atmosphere and ingress of terrestrial atmosphere) and a sample lid inside the container to fit over the solid material (to prevent material rattling round the whole canister, resulting in abrasion and erosion of the pebbles). There must be a double valve on the outer surface, to allow the gas to be removed on return to Earth. Although the solid samples are below a lid, this, quite specifically, is not gas-tight, so the atmosphere and solid can remain in contact and maintain their equilibrium. This should preclude (a) dehydration of hydrated minerals and (b) back reaction between degassed species and anhydrous minerals.

Landing Site

There are several sets of parameters that must be reconciled in the selection of a suitable landing site. These can be categorised as:

- i) Science – will the selected site fulfil the scientific goals of the mission?
- ii) Technology – what are the technology requirements that might lead to modification of the science goals of the mission?
- iii) Engineering – what are the engineering requirements of the mission, in terms of landing ellipse, and how might they modify the science goals of the mission?

The landing site that will finally be selected is a result of trade-offs amongst these three parameter sets. The Master mission does not require precision landing, although it does require a reasonably flat landing site at low latitude to enable successful return of a canister from the surface of Mars to Mars orbit. Preliminary engineering requirements identified a latitude range of $0 \pm 30^\circ$ for earlier MSR mission designs. There is another consideration that must also be taken into account: there have now been five successful landings on Mars' surface, such that several different regions of Mars have been characterised to greater detail than is possible just from orbital data. Should we consider a return trip to a region that has already been visited, or should the mission return material from a new site? The pros and cons of each option are given in Table 5.

Table 5: Arguments balancing a mission to a fresh landing site against one that has already been explored

	New site	Old site
PRO	<ul style="list-style-type: none"> – New terrain, new discoveries <i>in situ</i>, even from reduced instrument suite – Opportunity to characterise additional area, giving increasing info about terrain variation on Mars 	<ul style="list-style-type: none"> – Good geological context for samples – Complementary info will enhance data from <i>in situ</i> studies – Known technology and engineering issues already resolved
CON	<ul style="list-style-type: none"> – Limited geological context for sampled material 	<ul style="list-style-type: none"> – Not exploring a new area – More detail about a small area of Mars – Not increasing pan-regional knowledge?

On balance, it is preferred that the mission focuses on a new landing site, but one perhaps close to a previously visited region. For example, when Pathfinder landed at Ares Vallis in 1997, images of the terrain at the landing site showed angular cm-sized boulders spread across a flat plain. Dust covered the plain and the boulders, but through the dust layer, occasional lighter-coloured patches were observed, thought to be evaporates – minerals remaining after evaporation of fluid. The region was interpreted to be an outwash channel, one that flooded episodically, bringing together material from the region at the head of the channel. Evaporation of the flood waters led to precipitation of salts, and the production of 'hard pan' areas. If such a region were to be sampled by a 'grab and go' mission, it would be able to sample primary material (small boulders from local bedrock, addressing goal 1), secondary material (dust – addressing goals 1 and 3) as well as chemical precipitates (addressing goals 1 and 3).

In order to achieve the maximum number of **Master** mission objectives (Table 2), the selected landing site must cover the widest range of rock lithologies having the widest range of ages. Thus Noachian-age rocks overlain by secondary Hesperian or Amazonian clays (ie. Phyllosian overlain by siderikian or theikian rocks) would be an ideal location from which to collect material. An alternative landing site criterion could be the identification of secondary ejecta that can be directly correlated with impact craters of different ages, as successfully achieved by the sample collection undertaken by the Apollo missions to the lunar surface. Three such possible sites are described below; these areas have been recognised by landing site selection working

groups for previous MSR mission designs. Three possible landing sites are very briefly described here, to give an idea of the type of region from which samples would be sought. The sites have all been thoroughly studied as candidates for previous missions, and all are believed to be rich in clay minerals overlying a more ancient bedrock. All three sites are in cratered terrain, allowing for the opportunity to acquire fragments of basement material excavated by impact and distributed across the neighbouring plains.

Jezero Crater (18.4°N 282.4°W)

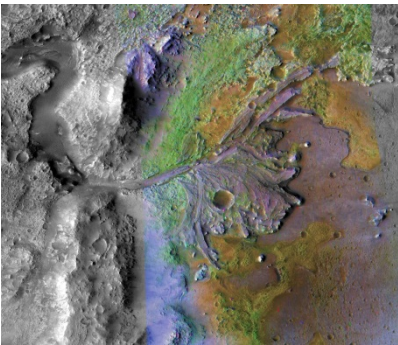
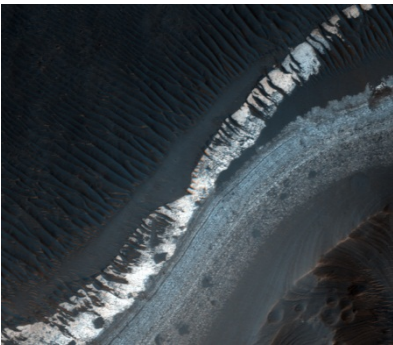
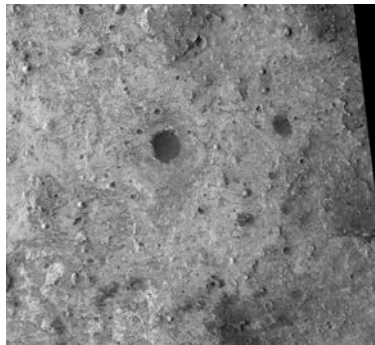
Jezero Crater (Figure 8) is on the north-western side of Isidis Planitia (see Figure 6). It is about 50 km across, and is characterised by layers of clays (including smectite) mixed with carbonates (Ehlmann et al., 2008). The crater wall is breached, and a river drains into the crater, forming a delta.

Holden Crater (26.4°S 34°W)

The stratigraphy exposed at Holden crater (Figure 9) reveals a lower layer of ancient igneous conglomerate of brecciated rock, presumably excavated during the time of crater formation. The crater was subsequently filled by a lake, from which fine-grained clay was deposited and which now overlies the megabreccia, and the clays in turn are covered by debris from a later episode of flash-flooding (Grant et al., 2008).

Mawrth Vallis (22.6°N 16.5°W)

Mawrth Vallis (Figure 10) is a valley cutting through ancient terrain. It has been infilled with a rich variety of phyllosilicates, clay mineral, hydrated silica and oxides, indicating several periods of aqueous activity with fluids of differing composition (Bishop et al., 2008).

		
<p>Figure 8: Colour-enhanced image of Jezero crater. Image PIA 15097. Copyright: NASA/JPL/JHUAPL/MSSS/Brown U</p>	<p>Figure 9: Part of the Holden crater, showing stratified clays. Image: NASA/JPL/University of Arizona</p>	<p>Figure 10: Mawrth Vallis Layered Deposits. Image No. PIA09405: taken by the HiRISE camera on MRO and is ~ 500 m across. Copyright: NASA/JPL/Univ. of Arizona</p>

Planetary Protection

The issue of planetary protection, both forwards to Mars and backwards to Earth, is an important and significant part of a sample return mission. For forward planetary protection, Earth to Mars, there are very specific protocols and procedures that have been accepted by COSPAR as mandatory for missions to Mars. They have been tested and found to be implementable through the series of successful Mars missions to date. Master would follow the directives right from the start of instrument design and build, and will take advantage of any improvements or advances in technology, in terms of sterilisation etc. that have resulted from previous missions. Specifically, the Master mission will take lessons and best practice from the planning (and possible operation) of the sample return missions to Phobos (Phootprint) and an asteroid (Marco Polo-R).

Backward planetary protection, from Mars to Earth, has also been covered by international protocols, but has not yet been tested for utility or applicability. One question that will have to be considered during mission design is the likelihood (or otherwise) of the returned sample containing living micro-organisms: will it have to be sterilised prior to landing on Earth? The ability to follow planetary protection protocols during sample

handling will be a major factor in the design specifications of the sample handling and curation facility (next section). As above, cognizance will be taken of the results for the Phootprint and Marco Polo-R missions.

Sample Curation

In order to achieve the objectives of Master, the samples will be collected with strict contamination control. The sensitivity of the instruments required to tackle these analyses is very great, and many measurements are highly sensitive to contamination. Therefore, a bespoke facility is required providing a controlled, clean environment which will not add further contamination to the samples upon their return to Earth. Without the ability to protect the samples from a wide range of contaminants found on the Earth's surface (e.g. volatile organics, microbes, moisture, aerosols and dust, etc.) much of the science will not be achievable. It is unfeasible to include the costs of the design, build and operation of a Sample Curation Facility (SCF) within the €1b cost cap of an L class mission. It is perhaps more realistic to explore an alternative source of funding for the Facility – either by regarding it as an 'instrument' to be provided by a national agency, or by seeking support from the European Union (e.g., through the Horizon 2020 programme).

No dedicated contamination-free facility with the required high level of cleanliness exists in Europe for the curation of returned extraterrestrial samples. However, prior to the return of material from Mars planned through the Master mission, there are many large institutions in Europe that host significant collections of extraterrestrial material, generally they have limited contamination-controlled environments – e.g. NHM, London, MNH Paris, Univ Siena, etc. Specialist clean room sample handling facilities for extraterrestrial materials are found in a number of European universities (e.g. the meteorite curation and characterisation laboratories at Open University), but these still fall well short of the contamination control required for material returned from Phobos and an asteroid, not to mention from Mars. Spacecraft might return material from Phobos and asteroid (341843) 2008 EV5 (the Phootprint and Marco Polo-R missions, respectively). If these missions are accepted, then a sample curation facility must be built to house the returned samples. Although the PP and curation requirements of such samples are less stringent than those for a martian sample, if the facility is designed with the possibility of a future MSR mission in mind, then costs will be significantly decreased when it comes to curating material for the Master mission.

In order to provide appropriate samples to each laboratory, with the appropriate context information and, most importantly, without adding significant contamination to the materials, a dedicated SCF is required. Such a facility will have a range of functions, including: security of the samples; Preservation of the pristine nature of samples; Characterisation and sub-sampling of materials; Documentation of samples and procedures; Administration of the samples; Focal point for public outreach activities

An important aspect of many studies will be consideration of the possibility of contamination. As well as curation and provision of samples to the community, the SCF should also contain a collection of potential contaminants. These would include witness coupons from all stages of sample collection, return and handling, as well as samples of all materials from the spacecraft sub-systems and other possible contaminants (e.g. fuels, lubricants, etc). These materials would also be available to the community for analysis.

Proposed Mission Architecture

The *Master* mission is designed as a 'grab and go' mission, with a mission architecture foreseeing a spacecraft composed of two main elements, an orbiter plus a lander. One of the reasons why previous specific proposals for a MSR mission have not been included in the forward plan of space agencies has been their cost. The missions become complex, with too many instruments in the payload, which were designed to undertake observations on the martian surface, as well as to select and collect very specific samples from closely defined rock types. The mission we propose here is much more focussed on bringing a sample back, and carrying out all the research on the material once it is back on Earth. This allows us to reduce the instrument payload which will be carried to Mars, including only instruments appropriate for characterising the landing site at a level sufficient to give context to the recovered material, and to ensure that suitable material is returned.

Mission Payload

- an orbiting element (*Orbiter*), including:

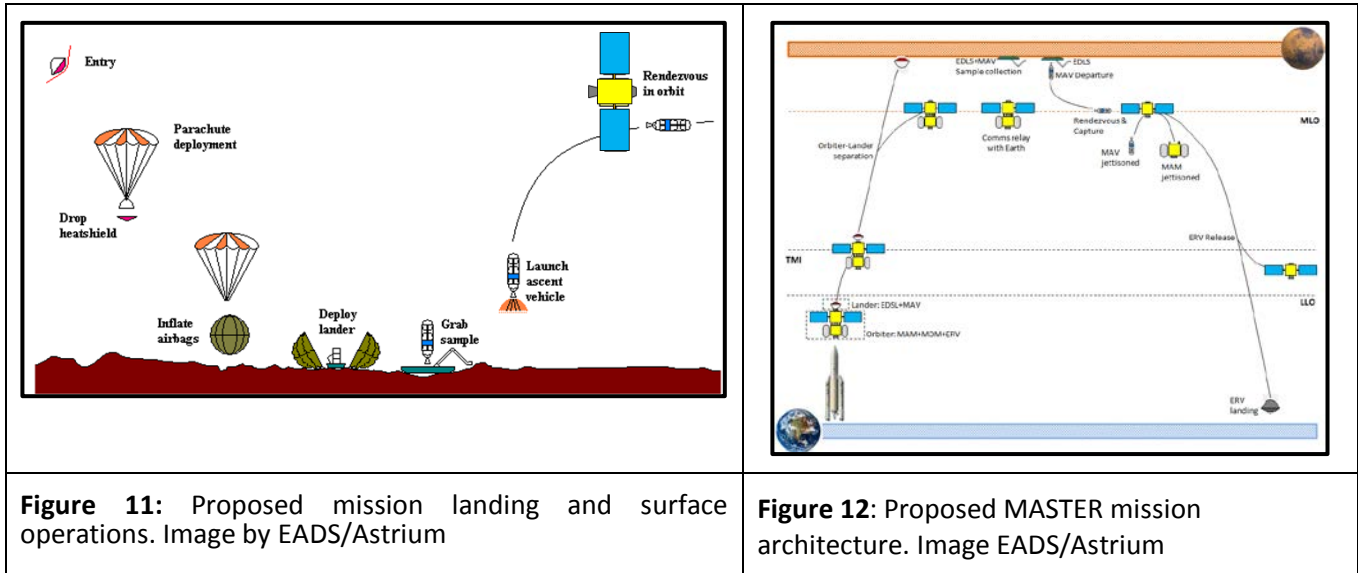
- Mars Arrival Module (MAM) carrying the propellant required to enter Mars orbit and carry out Mars orbit operations, together with the equipment needed specifically for Mars orbit operations. The MAM is discarded before the departure for return to Earth, in order to reduce overall mass and, then, the propellant needs
- Mars Departure Module (MDM) performing the Mars-to-Earth leg of the mission, and carrying the equipment needed to support the mission throughout (e.g. avionics, power and communications)
- Earth Re-entry Capsule (ERC) that performs the Earth atmospheric entry and brings the sample back to the surface of Earth
- a landing/ascending element (*Lander*), including:
 - entry, descent & surface element (EDLS) performing the landing and supporting the surface operations required to collect the sample
 - two-stage Mars ascent vehicle (MAV) carrying the sample back into orbit.

The overall strategy relies on performing a Mars Conjunction-class mission in order to reduce the propulsive requirements (minimum delta-Vs) and, therefore, the costs in terms of propellant need. The spacecraft will be launched with an ArianeV, aimed at performing the injection into a Trans-Mars Trajectory (TMI). Therefore, the two main elements, *Orbiter* and *Lander*, will separate on the approach trajectory once in proximity of Mars in order to reduce capture delta-Vs. The *Orbiter* will insert into Mars Orbit, from where it will work as a communication relay system between the *Lander* and the Earth Mission Control Centre. Both the manoeuvres of orbit injection and orbit stationkeeping are performed by the Mars Arrival Module (MAM). The *Lander* will perform a direct entry into Martian atmosphere and a soft landing through the use of parachutes and airbags. Once “softly” landed on the surface, the Entry Descent & Landing System (EDLS) will collect the required “soil + atmosphere” sample to be stored inside the Sample Container (SC) accommodated in the Mars Ascent Vehicle (MAV). The surface operations duration will be kept as short as possible, in line with a ‘grab and go’ mission. This will reduce risks associated with long stays in Martian environment, e.g. seasonal storms. Moreover, the sample shall be collected at a reasonable distance from the landing site, to avoid contamination due to landing manoeuvre, this implying that a long arm accommodated on top of a static lander is suitable for performing the mission, and there is no need of an additional element guaranteeing surface mobility. This approach would save mass and, at the same time, reduce both mission complexity and amount of time to be spent on the surface. The robotic arm will build on heritage from the Beagle 2 ‘PAW’, but will host a reduced instrument payload. The current nominal payload is as follows, although additional elements will be considered if the comply with engineering constraints:

- Stereo cameras - These will provide sets of images in colour of the surrounding landscape, in order to select the most promising collection site. Filters of different wavelengths will enable additional compositional information to be acquired.
- Close-up camera - The close-up camera will provide images of the surface at a spatial resolution of ~100 µm. This will give an impression of the texture, permeability, grain size and grain size variation of the surface.
- Near-IR Spectrometer - The spectrometer will acquire compositional data of the soil for comparison with compositional data acquired by instruments on other landers. Once the samples are returned and analysed by lab instruments, a ‘calibration’ will be established, enabling intercomparison of results from different missions.
- Sample Collection Mechanism (SCM) - There are two options for the SCM: either to be mounted directly on the lander, or to be deployed on a robotic arm. The first option is less favourable, because it restricts the area available for sample selection. The second option allows for preliminary analysis of the sample selection area prior to sample collection. Deployment on an arm would also allow several different sites to be considered before the final decision of collection site is taken. Depending on construction of the sample return canister, it might be possible to collect more than one sample, although this would add to the complexity, and thus cost of the mission. The baseline proposal is for a collection of a single regolith sample (plus ambient atmosphere).

A summary of the surface operations sequence is shown in Figure 11.

Once the sample has been collected and stored into the Sample Container (SC), the MAV will depart. Upon reaching the desired altitude, a rendezvous manoeuvre will start, with the *Orbiter* chasing the MAV until this latter is captured and the SC is transferred and sealed into a Bio Container accommodated in the Earth Re-Entry Capsule. The MAV and the Mars Arrival Module will be jettisoned, and the Mars Departure Module will provide the required delta-Vs to escape from Mars. Upon arrival of the spacecraft close to Earth, the Earth Re-Entry Capsule (ERC) will be injected into a direct entry trajectory. A summary of the main phases of the proposed mission, together with the involved elements is given in Figure 12.



The timeline for implementing the overall mission is approximately the following, where:

- i) Launch from Earth to Mars orbit;
- ii) Arrival at Mars, release of the Landing element
- iii) Soft landing with airbags
- iv) Surface activities, Communication support from Mars Orbit
- v) Launch of Sample Canister from surface of Mars to Mars Orbit
- vi) Retrieval of Canister by Orbiting element
- vii) Return to Earth Orbit
- viii) Descent of capsule to Earth's surface
- ix) Retrieval of Canister and transport to receiving facility
- x) Transfer to Curation Facility

Launch date is determined by orbital dynamics. The most favourable launch window opens in the fourth quarter of 2028, resulting in a stay-at-Mars time of ~ 300 days (Vasile et al., 2005). Most of this will be in orbit around Mars, after the sample is collected, rather than collecting the sample and keeping it isolated on the martian surface.

Table 6 – Master Mission Concept Timeline

	2028	2029				2030				2031			
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
(i)	Launch from Earth to Mars orbit												
(ii), (iii)				Arrival at Mars, Lander									

	2028	2029				2030				2031			
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
(iv)				released to surface	Surface Activities								
(v), (vi)								Launch from surface to Mars orbit					
(vii)									Canister retrieval from Mars orbit & transfer to Earth orbit				
(viii) (ix)												Canister retrieval from Earth surface & transfer to Receiving Facility	
(x)													Transfer to Curation Facility; Preliminary analysis programme

A preliminary mass assessment of this Mission Architecture has been also derived and is reported in Table 7.

Table 7: Preliminary Mass Budgets

	Total Mass (kg)	Dry Mass (kg)	Propellant Mass (kg)
MAM	2016	292	1724
MDM	1121	512	610
ERV	43	43	
Orbiter Element	3180	846	2334
EDLS	187	187	
MAV	113	34	79
Lander Element	300	221	79

The total mass allows for a 20% margin with respect to the ArianeV maximum P/L mass capability to be launched into a TMI, i.e. about 4290kg. More accurate mass break downs can be determined once the landing site and staging strategy are analysed during future mission studies. Trade-offs on the overall architecture will also need to be performed in the next study phases and might result in significant mass savings or enhanced mission capability. Among them, a trade-off shall be performed to evaluate advantages in using “Chemical vs. Solar Electric” Propulsion (e.g. using Solar Electrical Propulsion for the return journey).

Summary

We have designed a mission to bring three samples back from Mars to Earth in a single container. The samples will be of rock, dust and atmosphere, collected from a region of Mars where ancient terrain is altered by aqueous activity. The mission has a goal to determine the absolute age of a specific region of Mars, such that the relative chronology of the surface stratigraphy currently based on crater counting can be anchored and made into an absolute age scale. The mission will also measure the composition of the atmosphere, especially the least abundant noble gas species, to constrain the evolution of Mars atmosphere. Additional goals include a search for biological activity, and an assessment of the potential for human exploration of Mars. The mission will last three years, with one year in orbit around Mars. Full attention will be paid to planetary protection protocols, both backwards and forwards. Once the protocols are satisfied, material will be distributed for analysis by the international community. At least 30% of the material will be maintained in an untouched state for posterity, and a similar amount will be made available for outreach and education purposes.

References

- Ansan, V.; Loizeau, D.; Mangold, N.; Le Mouélic, S.; Carter, J.; Poulet, F.; Dromart, G.; Lucas, A.; Bibring, J.-P.; Gendrin, A.; Gondet, B.; Langevin, Y.; Masson, Ph.; Murchie, S.; Mustard, J. F.; Neukum, G. (2011). Stratigraphy, mineralogy, and origin of layered deposits inside Terby crater, Mars. *Icarus* 211, 273-304.
- Bandfield J. L., Glotch T. D. and Christensen P. R. (2003). Spectroscopic identification of carbonate minerals in the Martian dust. *Science* 301, 1084-1087.
- Bibring J.-P. et al. (2006). *Science* 312, 400 – 404
- Bishop, Janice L.; Dobrea, Eldar Z. Noe; McKeown, Nancy K.; Parente, Mario; Ehlmann, Bethany L.; Michalski, Joseph R.; Milliken, Ralph E.; Poulet, Francois; Swayze, Gregg A.; Mustard, John F.; Murchie, Scott L.; Bibring, Jean-Pierre. (2008). Phyllosilicate Diversity and Past Aqueous Activity Revealed at Mawrth Vallis, Mars. *Science* 321, 830-832.
- Bouvier A. Blichert-Toft J.; Vervoort, Jeffrey D.; Albarède, Francis. (2005). The age of SNC meteorites and the antiquity of the Martian surface. *Earth and Planetary Science Letters* 240, 221-233.
- Bouvier A., Blichert-Toft, J.; Vervoort, Jeffrey D.; Gillet, Philippe; Albarède, Francis. (2008). The case for old basaltic shergottites. *Earth and Planetary Science Letters* 266, 105-124.
- Brack, A.; Horneck, G.; Cockell, C. S.; Bérce, A.; Belisheva, N. K.; Eiroa, Carlos; Henning, Thomas; Herbst, Tom; Kaltenegger, Lisa; Léger, Alain; Liseau, Réne; Lammer, Helmut; Selsis, Franck; Beichman, Charles; Danchi, William; Fridlund, Malcolm; Lunine, Jonathan; Paresce, Francesco; Penny, Alan; Quirrenbach, Andreas; Röttgering, Huub; Schneider, Jean; Stam, Daphne; Tinetti, Giovanna; White, Glenn J. (2010). Origin and Evolution of Life on Terrestrial Planets. *Astrobiology* 10, 69-76.
- Cavicchioli, Ricardo (2002). Extremophiles and the Search for Extraterrestrial Life. *Astrobiology* 2, 281-292.
- Christensen, Philip R. and Ruff, Steven W. (2004). Formation of the hematite-bearing unit in Meridiani Planum: Evidence for deposition in standing water. *Journal of Geophysical Research* 109, E8, CiteID E08003.
- Christensen P. R., McSween H. Y., Bandfield J. L., Ruff S. W., Rogers A. D., Hamilton V. E., Gorelick N., Wyatt M. B., Jakosky B. M., Kieffer H. H., Malin M. C. and Moersch J. E. (2005). Evidence for magmatic evolution and diversity on Mars from infrared observations. *Nature* 436, 504-509.
- Clark, B. C.; Morris, R. V.; McLennan, S. M.; Gellert, R.; Jolliff, B.; Knoll, A. H.; Squyres, S. W.; Lowenstein, T. K.; Ming, D. W.; Tosca, N. J.; Yen, A.; Christensen, P. R.; Gorevan, S.; Brückner, J.; Calvin, W.; Dreibus, G.; Farrand, W.; Klingelhoefer, G.; Waenke, H.; Zipfel, J.; Bell, J. F.; Grotzinger, J.; McSween, H. Y.; Rieder, R. (2005). Chemistry and mineralogy of outcrops at Meridiani Planum. *Earth and Planetary Science Letters* 240, 73-94.
- Clifford, Stephen M. and Parker T. J. (2001). The Evolution of the Martian Hydrosphere: Implications for the Fate of a Primordial Ocean and the Current State of the Northern Plains. *Icarus* 154, 40-79.
- Cockell, C. S.; Rettberg, P.; Horneck, G.; Patel, M.; Lammer, H.; Córdoba-Jabonero, C. (2002). Ultraviolet protection in microhabitats - lessons from the terrestrial poles applied to Mars. In: *Proceedings of the First European Workshop on Exo-Astrobiology*, 16 - 19 September 2002, Graz, Austria. Ed.: Huguette Lacoste. ESA SP-518, Noordwijk, Netherlands: ESA Publications Division, ISBN 92-9092-828-X, 2002, p. 215 – 218.
- Connerney J. E. P., Acuña M. H., Wasilewski P. J., Ness N. F., Rème H., Mazelle C., Vignes D., Lin R. P., Mitchell D. L. and Cloutier P. A. (1999). Magnetic lineations in the ancient crust of Mars. *Science* 284, 794-798.
- Ehlmann, B.L.; Mustard, John F.; Fassett, Caleb I.; Schon, Samuel C.; Head, James W., III; Des Marais, David J.; Grant, John A.; Murchie, Scott L. (2008). Clay minerals in delta deposits and organic preservation potential on Mars. *Nature Geoscience* 1, 6, 355-358.
- Ehlmann B. L. and Mustard, John F. (2012). An in-situ record of major environmental transitions on early Mars at Northeast Syrtis Major. *Geophysical Research Letters* 39, L11202.
- Grant J.A.; Irwin, Rossman P., III; Grotzinger, John P.; Milliken, Ralph E.; Tornabene, Livio L.; McEwen, Alfred S.; Weitz, Catherine M.; Squyres, Steven W.; Glotch, Timothy D.; Thomson, Brad J. (2008). HiRISE imaging of impact megabreccia and sub-meter aqueous strata in Holden Crater, Mars. *Geology* 36, 195-198.
- Hartmann W. K. and Neukum G. (2001). Cratering chronology and the evolution of Mars. In: *Chronology and Evolution of Mars*. Eds. R. Kallenbach et al., Kluwer, Dordrecht. 96, 165-194.
- Head J.W. et al. (2001). *Sp. Sci. Rev.* 96, 263-292;

- iMARS (2008). Preliminary Planning for an International Mars Sample Return Mission. Report of the iMARS (International Mars Architecture for the Return of Samples) Working Group.
- Lapen T.J. et al. (2010). *Science* 328, 347-351;
- Lichtenberg, Kimberly A.; Arvidson, Raymond E.; Morris, Richard V.; Murchie, Scott L.; Bishop, Janice L.; Fernandez Remolar, David; Glotch, Timothy D.; Noe Dobrea, Eldar; Mustard, John F.; Andrews-Hanna, Jeffrey; Roach, Leah H. (2010). Stratigraphy of hydrated sulfates in the sedimentary deposits of Aram Chaos, Mars. *Journal of Geophysical Research*, Volume 115, Issue B6, CiteID E00D17
- McLennan et al. (2012). *Astrobiology*
- McSween H. Y. (1994). What we have learned about Mars from SNC meteorites. *Meteoritics* 29, 757-779.
- McSween, Harry Y., McGlynn, Ian O. and Rogers, A. Deane. (2010). Determining the modal mineralogy of Martian soils. *Journal of Geophysical Research*, Volume 115, CiteID E00F12
- Mangold, N.; Roach, L.; Milliken, R.; Le Mouélic, S.; Ansan, V.; Bibring, J. P.; Masson, Ph.; Mustard, J. F.; Murchie, S.; Neukum, G. (2010). A Late Amazonian alteration layer related to local volcanism on Mars. *Icarus* 207, 265-276.
- MEPAG ND-SAG (2008). Science Priorities for Mars Sample Return, White paper, 73 p, posted March 2008. by the Mars Exploration Program Analysis Group (MEPAG; <http://mepag.jpl.nasa.gov/reports/index.html>).
- Mustard, John F. et al. (2008). Hydrated silicate minerals on Mars observed by the Mars Reconnaissance Orbiter CRISM instrument. *Nature* 454, 305-309.
- Nimmo F. and Tanaka K (2005). Early crustal evolution of Mars. *Ann. Rev. Earth Planet. Sci.* 33, 133–61.
- Nyquist L. E., Bogard D. D., Shih C.-Y., Greshake A., Stoffler D. and Eugster O. (2001) Ages and geologic histories of martian meteorites. In: *Chronology and Evolution of Mars*. Eds. R. Kallenbach et al., Kluwer, Dordrecht. 96, 105-164.
- Orofino, V.; Goldspiel, J.; Carofalo, I.; Blanco, A.; Fonti, S.; Marzo, G. A. (2009). Evaluation of carbonate abundance in putative martian paleolake basins. *Icarus* 200, 426-435.
- Pollack J. B., Kasting J. F., Richardson S. M. and Poliakov K. (1987). The case for a wet, warm climate on early Mars. *Icarus* 71, 203-224.
- Poulet F., Bibring J.-P., Mustard J. F., Gendrin A., Mangold N., Langevin Y., Arvidson R. E., Gondet B. and Gomez C. (2005). Phyllosilicates on Mars and implications for early martian climate. *Nature* 438, 623-627.
- Tanaka K.L. (1986). The Stratigraphy of Mars. *Proc. 17th Lunar Planet. Sci. Conf.*, *J. Geophys. Res.* 91, suppl., 139–158.
- Treiman, A. H. and the MSL Science Team (2013). Basaltic Soil of Gale Crater: Crystalline Component Compared to Martian Basalts and Meteorites. *Lunar Planet. Sci. Conf.* 44, No. 1113
- Vasile M., Summerer, Leopold; de Pascale, Paolo. Design of Earth Mars transfer trajectories using evolutionary-branching technique. (2005). *Acta Astronautica* 56, 705-720.
- Wiens R.C. and the MSL Science Team (2013). Compositions determined by ChemCam along Curiosity's travers from Bradbury Station to Glenelg in Gale Crater, Mars. *Lunar Planet. Sci. Conf.* 44, No.
- Williams R.M.E. and the MSL Science Team (2013). Curiosity's mastCam images reveal conglomerate outcrops with water-transported pebbles. *Lunar Planet. Sci. Conf.* 44, No. 1617.
- Yen A.S. and the MSL Science Team (2013). Evidence for a global martian soil composition extends to Gale Crater. *Lunar Planet. Sci. Conf.* 44. No. 2495
- Yen A.S. et al. (2005). An integrated view of the chemistry and mineralogy of martian soils. *Nature* 436, 49-54.
- Yingst R. A. and the MSL Science Team (2013). Characteristics of pebble and cobble-sized clasts along the Curiosity rover traverse from Sol 0 to 90. *Lunar Planet. Sci. Conf.* 44, No. 1232.
- Zahnle K.J., Early Earth: Earth's Earliest Atmosphere. (2006). *Elements* 2, 217-222

A proposed observatory mission:

Hypertelescope Optical Observatory (HOO)

**1 to 100km flotilla for direct images at microarc-second resolution on stars,
exo-planets and deep fields**

Antoine Labeyrie (Collège de France and Observatoire de la Côte d'Azur, France) antoine.labeyrie@oca.eu
Pierre Riaud (Collège de France) riaud.pierre@gmail.com
Stefania Residori (INLN, Université de Nice-Sophia Antipolis, CNRS) Stefania.Residori@inln.cnrs.fr
Umberto Bortolozzo (INLN, Université de Nice-Sophia Antipolis, CNRS) Umberto.Bortolozzo@inln.cnrs.fr
Farrokh Vakili (Observatoire de la Côte d'Azur) farrokh.vakili@oca.eu

Abstract

Following earlier proposals of interferometric flotilla concepts for space, and hypertelescope versions for direct imaging and coronagraphy, an updated observatory-type instrument is now proposed in the form of a larger Hypertelescope Optical Observatory . Like a conventional telescope, but with the much higher resolution provided by its meta-aperture diameter expandable from 1km to 10km and perhaps 100km, it can provide rich direct images and spectro-images of varied sources. High limiting magnitudes, well beyond those of Keck, HST and JWST will be attainable if more collecting area is installed, in the form of thousands of flying mirrors which can be as small as 30mm.

The optical design can be similar to that of the ground-based hypertelescope prototypes currently tested ¹ , and the kilometric version also studied ² . Three options, previously described for the Luciola proposal ⁵ , are again considered for the flotilla's drive and control system : 1- conventional micro-thrusters attached to each mirror; as recently tested in orbit by the pair of PRISMA micro-satellites; 2- small solar sails; 3- laser trapping in interference fields, providing passive stabilization at sub-wavelength accuracy .

1. Introduction

Interferometry at radio wavelengths has greatly improved its imaging performance and discovery potential when tens and hundreds of antennas became connected, pending the 100,000 considered in some current projects . Similar gains are expected at optical wavelengths, with the novel form of interferometer called “Hypertelescope” . Ground-based prototypes are being tested toward a dilute meta-aperture spanning a kilometer and combining hundreds of small mirrors. And space versions involving a flotilla of many small mirrors have also been studied since 1996. For a given collecting area and resolution , and likely also for a given cost, many small apertures provide a better imaging performance, in terms of dynamic range, than fewer large ones.

The HOO is intended to be an updated and larger version of the Luciola hypertelescope flotilla previously proposed to ESA . It incorporates new design concepts toward more flexibility , upgradability and expandability toward a 10 km or larger meta-aperture size.

Beyond conventional interferometric and coronagraphic techniques proposed for exoplanet detection in space, hypertelescope flotillas of many small mirrors can provide direct images with a high information content and dynamic range. With a meta-aperture size reaching 100km, not to mention the 100,000km also studied for a later generation of hypertelescopes in space, morphology details are in principle resolvable on some habitable exoplanets ¹⁷ . At exoplanet sites containing

photosynthetic life, seasonal variations which may be observable spectroscopically may provide a robust signal for its detection.

The HOO, as a giant dilute telescope, is intended to be an observatory instrument with broad capabilities on many object types, including the faintest galaxies beyond the current magnitude and resolution limits of HST, JWST, and forthcoming large ground-based instruments equipped with a Laser Guide Star¹, such as ELTs and the kilometric-sized "Extremely Large Hypertelescope" (ELHyT)², currently tested with the 60-200m "Ubaye Hypertelescope" precursor³.

The HOO will thus not be restricted to exo-planet science, but usable like conventional telescopes on most celestial sources, for stellar physics, or observing neutron star and black hole environments, gravitational lensing, and deep-field imaging for cosmology, as discussed in the Luciola proposal previously submitted to ESA⁵.

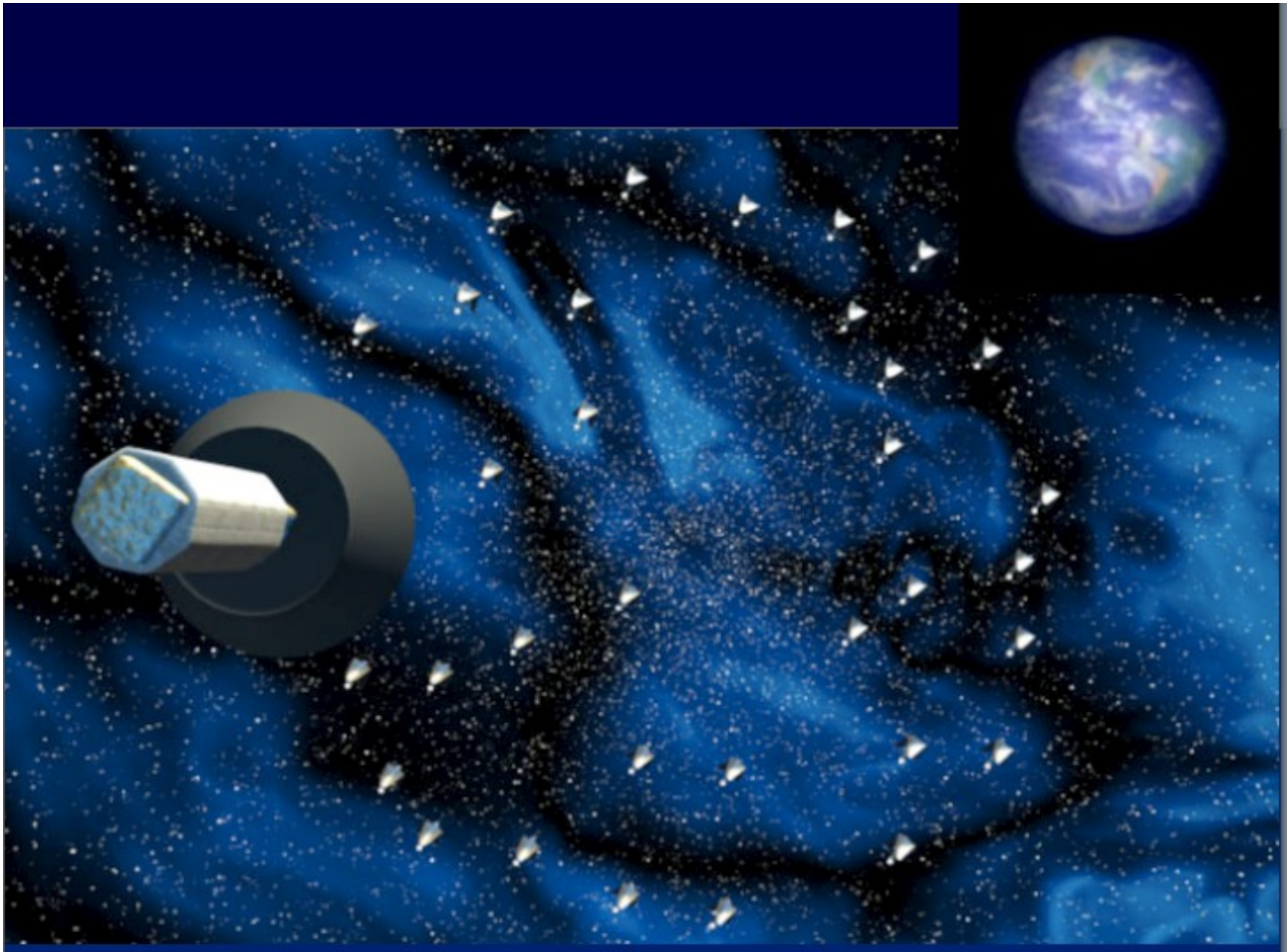


Figure 1: Artist view of the “Exo-Earth Discoverer” hypertelescope, (courtesy NASA). A flotilla of small mirrors, driven by similarly small solar sails, focuses star light toward one or more focal spaceships. Alternative driving options also considered are conventional micro-thrusters or laser trapping . Self-deploying versions, reversibly expandable between meta-aperture sizes in the range from 100m to 1, 10 or perhaps 100km , will be of interest for a broad diversity of science targets. The inset at top right is a numerical simulation of an Earth's direct image, recorded by a 100km hypertelescope at 3 parsecs. It has 100 sub-apertures of 3m.

Following the early TRIO proposal of an interferometer flotilla , the subsequent DARWIN version studied by ESA, and the initial hypertelescope proposals to NASA and ESA (LOVLI, Exo-Earth Discoverer, Epicurus and Luciola), the PRISMA test in orbit has supported the concept of a flotilla driven by micro-thrusters. As an effort to reduce the cost of Luciola, while using more mirrors of smaller size, the alternate concept of a "laser-trapped hypertelescope flotilla"

was developed ⁶ and laboratory testing undertaken in high vacuum . If it becomes validated, it may drastically simplify the flotilla hardware, and its deployment across 100 km at the Lagrangian L3 point, as well as reduce the cost. It also favors smaller mirror elements, down to perhaps 30mm, with a million of them for a collecting area comparable to an ELT. The resulting theoretical dynamic range in the direct image would then also reach 10^6 , and in fact more since a form of apodization is achievable by decreasing the mirror density toward the edge of the flotilla, as achieved in radio arrays . Coronagraphic masking is also applicable.

The HOO concept which is proposed is flexible in terms of the flotilla size, expected to be adjustable from perhaps 100m to 10km or even 100km. The optimal size of the mirror elements and their number will have to be defined , as well as the feasibility of upgrades with subsequent deliveries of additional mirrors after some years of operation, as traditionally done for radio-interferometry arrays of antennas.

Figure 2: Laboratory images of an artificial star cluster, seen through Fizeau masks having various numbers of sub-apertures , randomly patterned. From left to right: 15, 50, 235, 600 and the full meta-aperture. The decreasing contamination by the sub-aperture's broad diffractive envelope is clearly evidenced. The Fizeau mask can be rotated during the exposure , for simulating aperture supersynthesis, which improves the image obtained with few apertures, but negligibly with many . Also, cases with large or small sub-apertures providing equal collecting area have been compared (not shown here) to verify the theoretical gain of imaging performance in the latter case.

2. Science with milli- to microarcsecond resolution , direct imaging, and a high limiting magnitude

The history of discoveries in Astronomy, since telescopes began operating in the hands of Galileo, has been strongly influenced by the steadily increasing size of their aperture. As larger new telescopes became operated, they have often shown fainter sources or unexpected detail of major significance . The prospect of greatly improving both luminosity and angular resolution in future instruments such as interferometric flotillas of mirrors raises high hopes for abundant discoveries in many fields of astronomical research, including cosmology, astrobiology and perhaps SETI.

In the way of interferometry, the enhanced form called “hypertelescopes” was introduced for utilizing numerous sub-apertures and efficiently provide direct images . Its luminosity, identical to that of a monolithic telescope having equivalent collecting area, thus implies the same deep-field imaging capability as HST or the forthcoming ELTs when reaching a comparable collecting area, the cost of which should be less if the mirror elements are smaller .

2.1 Exoplanets

The spectacular recent progress of exoplanet observing is a strong stimulation toward further improving the instruments, especially for observing habitable exoplanets. Much of the rationale discussed for exoplanet science is relevant to the HOO, but its resolution, coronagraphic capabilities and the limiting magnitude reachable once upgraded bring further possibilities. Among these are:

Resolved images of exoplanet transits, and spectroscopy of the refractive arc

With a kilometric hypertelescope, transiting planets such as discovered photometrically by Kepler are opportunities for resolved images resembling those of the Sun-Venus transit events observed in the recent years. Briefly during immersion and emergence, these have shown spectacular crescent arcs likely resulting from grazing refraction of solar rays through Venus' high atmosphere.

Comparable images can be expected to become observable with the HOO on transiting exoplanets within a few parsecs. Because the crescent arc is little contaminated by direct starlight, this can in principle greatly increase the sensitivity of spectroscopy for probing the exoplanet's atmospheric absorption, and searching for bio-signature features. Instead, the coronagraphic observing of the same exoplanet at a different orbital phase, when it becomes well separated from its parent star, will likely be difficult for habitable exoplanets, even with a hypertelescope.

Bio-signature signals observable at microarc-second resolution

Beyond the 10km limit of meta-aperture size mainly proposed here, the 100km size previously discussed¹⁷ for an “Exo-Earth Imager” can in principle provide much more detail on habitable exoplanets within a few parsecs. In particular the “Indian Summer signal”, i.e. the possible seasonal spectroscopic variation of photosynthetic life in some resolved zones at mid-latitudes, on exo-planets having a spin axis suitably tilted about their orbital plane, is of interest as a detection target since it can be a robust bio-signature¹⁸.

Whether or not the necessary resolution and extreme coronagraphic camera can be planned as a possible upgrade of the HOO, or instead would require a separate mission at a later stage, will have to be studied.

Supernovae and extragalactic Cepheids

SNe are very rare events in our galaxy: typically a few per century. Conversely 2 or more SNe are discovered per year within the VIRGO cluster alone and in a 3 times larger volume, a hundred SNe could be detected at least, by an interferometer for an equivalent collecting area of a 10m telescope. Direct spatial information on the structure of the expanding ejecta, across different spectral lines would be obtained by a kilo to deca-kilometric imaging array. Thus enabling us to follow the details of the SN explosion and the complex mechanisms that govern the structural evolution within the local univers.

With 10 km baselines we will measure the individual members of star-burst HII regions in the Magellanic Clouds, the kinematics of compact clusters like in 30 Doradus by measuring the proper motion of member stars bringing news constrains on their IMF, multiplicity of stellar population and finally test the evolutionary scenarii of such clusters like star evaporation for instance.

2.2 Deep fields and cosmology

The modest limiting magnitude of existing interferometers, yet observing through the atmosphere in the absence of a laser guide star, has precluded work on deep sky sources. Hypertelescopes on Earth however may become capable of blind phasing with a Laser Guide Star, and thus reach high limiting magnitudes. In space, the wide isoplanatic patch, allows phasing hypertelescopes with a natural guide star, and reaching a high limiting magnitude, same on unresolved sources as with a monolithic telescope of equivalent collecting area.

This opens the way to a vast new realm of high-resolution imaging on:

- galaxy structures, including AGNs and the environment of central black holes,
- gravitational lensing bodies and background sources, including the predicted brief diffractive transients from “free floating planets” at sub-parsec distances¹⁹.
- cosmologic sources, including little resolved galaxies fainter than those seen in the HST Ultra Deep Field, if the collecting area exceeds its 6m^2 .
- optical counterparts of gamma ray bursts
- extragalactic supernovae and Cepheids

3. The HOO concept

Like *Luciola* ⁵, the HOO has a dilute flotilla of many small mirrors, belonging to a common spherical or paraboloidal locus, and feeding light to one or more focal spaceships. For accurate positioning, it uses laser metrology systems such as discussed for *Luciola*, also studied by ESA, and possibly including the coarse positioning technique tested with the pair of PRISMA satellites. The new option, absent in the original *Luciola* proposal, of laser-trapped mirrors does not require a metrology system.

3.1 Basic specifications

The terrestrial hypertelescopes currently studied, such as the ELHyT, are unlikely to have a meta-aperture much larger than a kilometer, in the absence of large and deep enough craters or other concave sites on Earth. Larger flat sites are available and can conceivably carry hundreds of telescopes, but long delay lines would be needed and the cost would be prohibitive for hundreds of them.

A reasonable range of size for a first-generation HOO, the space version in flotilla form, may therefore be from 100m to 1km and 10km, with potential expandability toward 100km in the longer term. An adjustable size would be most useful if its feasibility is confirmed since different science questions need different amounts of resolution. As demonstrated by the PRISMA testing, even a rather modest investment during the coming decade can validate the basic technology and be strongly conclusive toward justifying ambitious upgrades. Much preliminary testing can be done in the laboratory, including for the Laser Trapped version. The ground-based prototype “Ubaye Hypertelescope”, currently tested in a southern Alpine valley ^{1,3}, can also serve as a test bench and provide useful experience for the alignment and cophasing techniques.

The number N of sub-apertures can be as small as nine initially if provisions are made for upgrading to many more, especially if the flotilla size is also expanded, once the basic operation and science potential are verified. Up to thousands of *Luciola*-type mirrors, or perhaps a million “laser-trapped” ones as small as 30mm, can in principle be incorporated in a large flotilla for a huge increase in imaging performance.

Among the possible sites in space the L2 Lagrange point of Sun-Earth is particularly attractive for the solar driven version, given its low and uniform level of microgravity. L3, being partially shadowed by the Earth, is of interest for the laser-trapped version, in which case the laser source should reside with its photovoltaic generator some distance away in full sunlight ⁶.

Closer to Earth, high orbits may offer sufficiently weak gravity gradients for stabilizing the flotilla with conventional micro-thrusters at the scale of several kilometers, as demonstrated at the smaller scale of their 100m spacing by the pair of PRISMA satellites.

3.2 Wavelength range and spectral imaging

The spectral range exploited covers much of the ultra-violet, visible and near to mid infra-red. The prospect of laser cooling for the laser-trapped version, if it becomes verified, may further extend the infra-red range.

Dedicated cameras and auxiliary instruments can be installed within the single or multiple focal spaceships. Arecibo-like optical designs, with a spherical but dilute primary mirror, indeed allow the simultaneous use of several such focal spaceships, independantly movable along the focal surface, and which can be specialized for various spectral ranges and auxiliary instruments. This has been successfully achieved at Arecibo and found most useful.

The narrow field of view which can be exploited within one focal spaceship is compatible with

hyperspectral (also called spatio-spectral) imaging, a highly desirable feature providing a spectrum of each resel .

3.3 Optical concept

For designing the optics of a hypertelescope, one typically begins by designing the “meta-telescope”, i.e. a giant monolithic telescope having the same overall aperture size (the meta-aperture size) . Best efforts are made to achieve the widest diffraction-limited field of view, using the usual design recipes.

One then virtually adds a multi-hole Fizeau mask in front of the aperture, with hole sizes d and distribution similar to the desired hypertelescope. The hole spacing is typically much wider than their size, and this greatly attenuates the image intensity. With hundreds of them, it however typically leaves the Airy peak in the point spread function (PSF) little affected in terms of its width, thus preserving the angular resolution, but attenuated . The Airy rings become distorted into “speckle” sidelobes, and intensified relative to the peak . This degrades the dynamic range, thus affecting the detection of faint stellar companions such as exoplanets, and the contrast in images of extended or clustered sources.

Indeed, the Fizeau interferometer thus obtained, while retaining the angular resolution of the meta-telescope and its diffraction-limited field of view, is affected since the image-forming convolution degrades the image contrast when too many point sources are present within the diffractive envelope of the PSF. Also called “sub-aperture lobe”, this envelope is generated by diffraction through individual sub-apertures of size d , and its angular size is λ/d , where λ is the wavelength.

This “image crowding” effect limits the number of point sources allowed within the lobe for images retaining some contrast. The direct image of a star cluster smaller than the λ/d sub-aperture lobe cannot be contrasted unless it contains fewer than N^2 stars , considered as point sources. The phenomenon affects all types of interferometers, as a consequence of their incomplete sampling of the incoming wavefronts.

Transforming the Fizeau interferometer into a hypertelescope with a multi-field camera then consists in adding small optical elements near the focal camera:

1 - a micro-lens array in the focal plane, with pitch matching the sub-aperture lobe. With an appropriate image scale, this can separate independant imaging channels for each sub-field thus selected.

2- a pupil densifier, typically an array of micro-scale Galilean refractors, is also inserted in each imaging channel . These provide in each channel a direct intensified image, but which covers only a fraction λ/s of the channel's sky coverage λ/d , if the pupil is fully densified, where s is the sub-aperture spacing in the entrance aperture. The fractional sky coverage, in solid angle, is then $\{(\lambda/s) / (\lambda/d)\}^2 = (d/s)^2$. This loss of sky coverage is the cost to be payed for the image intensification, which occurs in the same ratio. Fortunately, the central “Direct Imaging Field” , also called “Clean Field”, thus created within each imaging channel can be offset toward any compact celestial source of interest located within the corresponding lobe, such as a resolvable star, exoplanet, or cosmological source.

The trade-off of intensification vs. sky coverage is adjustable, for fitting various types of sources, by varying the pupil densification. In addition, Aime (2012) and Mary (2012) have developed special deconvolution algorithms which can greatly extend the DIF coverage.

With respect to the full meta-telescope, the unavoidable loss of performance with the hypertelescope resides in the reduced sky coverage, the limiting magnitude, the dynamic range and the related image crowding. If upgrades prove feasible, following “early science” programs, by adding mirror elements through additional delivery missions, then all four performance aspects become improved.

Some design concepts allow “meta-aperture zooming”, i.e. varying its size as may be needed to adapt the resolution for various object types.

Apart from gas or dust nebulae, most astronomical sources of interest for high-resolution imaging have compact components, often little resolved. The simulated direct-images of an Exo-Earth at 3pc (figure 1, inset) , using a 100km hypertelescope with 100 apertures, illustrates the power achievable: continents are seen as well as oceans, cloud formations, and large forested areas such as the Amazon and Congo basins.

3.4 Sky coverage with the peculiar dilute and segmented field-of-view

When a Fizeau array is converted to a hypertelescope, by adding a multi-field separator with pupil densifiers, the dilute segmented aperture causes a comparable dilution and segmentation in the field of view captured by the direct-imaging camera in a focal spaceship. The image appearance effect resembles that of a printed image carrying a grid mask overlay with broad opaque lines, thus showing an array of narrow sub-fields. It does not prevent obtaining a full image, which requires stitching exposures made with slightly offset pointing.

3.5 The three driving options

Relevant hardware architectures and optical design concepts have been studied in much detail by different groups, mostly in Europe where a strong expertise has been acquired in the wake of its large terrestrial interferometers, the early TRIO proposal and the 1996 description of hypertelescopes .

For the HOO, three main options are considered for driving the hypertelescope flotilla:

a- Propellant thrusters: chemical micro-rockets, cold gas jets and ion thrusters are well known techniques . Their main limitation is the propellant volume which can be used, limiting the mission lifetime.

b- Small solar sails

Also proposed as an option for Luciola , according to the initial study for the TRIO proposal, small solar sails are of interest for driving the slow and accurate motions of the “flying mirrors”. Sails not much larger than each mirror element can suffice at sites such as the L1 Lagrange point of Sun-Earth or an Earth-trailing solar orbit, where microgravity fluctuations are well below those in Earth orbit. In comparison with conventional thrusters, solar sails can potentially extend the operational lifetime and reduce the pollution of optical surfaces..

c- Laser trapping ⁶

It uses a multitude of very small flying mirrors, typically 30 mm in diameter. Each is trapped by standing waves formed by a pair of laser beams propagating in nearly opposite directions . The scaling laws indeed indicate that the acceleration achievable for the small mirrors varies as their inverse size , at given total collecting area and laser power. This favors small mirrors for fast global repointing of the flotilla, but also for fighting gravity gradients, disturbances from micro-meteorite impacts, the competing solar radiation pressure, etc...

The more numerous smaller mirrors providing a given collecting area , at given flotilla size, also greatly benefit to the direct-imaging performance . Both the DIF extent and the dynamic range are improved. A practical minimal size, of the order of 30mm, is however imposed to keep the auxiliary optics, i.e. the focal optics , the laser beam launcher and associated diverging mirrors, within reasonable dimensions such as one meter.

A third benefit of smaller mirrors is to decrease the bulk and mass of the orbital delivery package,

expected to be self-deploying with the laser beams. Typically, a delivery package smaller than a cubic meter can contain a million 30mm mirrors, 1mm thick, providing together a 1000m² aperture area comparable to that of a 30m ELT, and providing a comparable limiting magnitude for the deployed hypertelescope.

3.6 Feasibility of adjustable flotilla size

Somewhat like the zooming of ordinary camera lenses, it would be desirable to have an adjustable flotilla size. Both the resolution and sub-field width would be influenced by such variation. Preliminary studies have indicated that this is likely feasible by expanding or contracting the flotilla. For the laser-trapped version, this appears feasible with active optical elements in the laser launcher package which modify the angular fan of the many laser beams.

4. Implementation steps

Like radio interferometers, optical hypertelescopes are highly modular and flexible in design, and they can grow or be up-graded, both in terms of meta-aperture diameter and of mirror density, by delivering additional elements. If the flotilla has a spherical geometry like the Arecibo radiotelescope, an additional upgrading possibility consists in adding focal spaceships for simultaneously and independantly observing different sky areas, as already demonstrated at Arecibo with its several detecors. This requires, on each focal spaceship, a corrector of spherical aberration in addition to the camera, spectrograph, coronagraph, etc... , which imposes a rather slow primary focal ratio if the corrector size is to remain manageable .

Since initially proposed, the concept of interferometric flotillas in space has been much studied, and tested in space with PRISMA for some of the basic control techniques needed. Only in 1996 did the “Hypertelescope” concept become proposed for efficient direct imaging. It became analyzed in theoretical detail, with different groups contributing theoretical analysis and numerical or laboratory simulations, confirmed by sky verifications at reduced scale ^{4,9,10} . Two ground-based prototypes are currently tested , with meta-aperture sizes expected to reach 200m, and a larger 1km version is under preliminary design . Like radio arrays of antennas, hypertelescopes have a flexible geometry which is upgradable. This is of interest on Earth for risk reduction and cost management, which may also be the case for a space flotilla if upgrades prove feasible at scheduled intervals , as achieved for HST.

Technology readiness level

The TRD has been discussed in the original Luciola proposal to ESA, but this did not include the Laser Trapping option which appeared later. For all three driving options, testing steps in the laboratory and possibly in space are needed for a robust evaluation.

5. Hypertelescope prospects for the longer term

The following extrapolated concepts, mentioned for perspective, were explored in preliminary detail but raise significant technical issues, unlikely to become solved before several decades. They are not part of the present proposal.

The “Dilute Bubble Hypertelescope”

Similar to the Arecibo radiotelescope, but in dilute form and with a complete dilute sphere for full-sky coverage, it features a static spherical array of hypertelescope “tiles” and a number of focal satellites, independantly moving along the half-sized focal sphere . The static mirrors save

pointing time since the pointing of a source is achieved by moving one of the focal spaceships, which can be available near the new position .

The 100,000 km “Neutron Star Imager” hypertelescope

The 100,000km “Neutron Star Imager” appears feasible in principle, but at a later stage . It is sized for resolving highly compact sources such as the Crab Pulsar, the size of which is believed to be about 20km. Its very high intrinsic luminance, much beyond that of ordinary stars, provides enough photons per resel, unlike ordinary stars which would not be easily observed with the nanoarcsecond angular resolution. The science program would thus concentrate on sources such as supernovae, QSO's and other violent objects.

For an adequate focal ratio of the sub-apertures, limiting to a few meters the size of their focal Airy peak, the “Neutron Star Imager” needs large component mirrors of 8m, not necessarily many, but costly and requiring significant technical developments.

Conclusion

The high science potential of a HOO and the concept flexibility appear to justify further testing and development in the laboratory and with ground-based hypertelescopes, toward defining a robust concept for a space instrument . Upgradability, if it proves feasible, can be most valuable.

References

1. A. Labeyrie. [Hypertelescopes : The challenge of direct imaging at high resolution](#). EAS Publication series, vol. 59 : 5-23, March 2013.
2. A. Labeyrie, D. Mourard, F. Allouche, R. Chakraborty, J. Dejonghe, A. Surya, Y. Bresson, C. Aime, D. Mary, and A. Carlotti. [Concept study of an Extremely Large Hyper Telescope \(ELHyT\) with 1200m sparse aperture for direct imaging at 100 micro-arcsecond resolution](#). In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8445, July 2012.
3. A. Labeyrie, F. Allouche, D. Mourard, F. Bolgar, R. Chakraborty, J. Maillot, N. Palitzyne, J. R. Poletti, J.-P. Rochaix, R. Prud'homme, A. Rondi, M. Roussel, and A. Surya. [Construction of a 57m hypertelescope in the Southern Alps](#). In society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8445, July 2012.
4. H. Le Coroller, J. Dejonghe, et al. [Tests with a Carlina-type diluted telescope. Primary coherencing](#). Astronomy & Astrophysics, vol. 539, March 2012.
5. A. Labeyrie, H. Le Coroller, J. Dejonghe, O. Lardière, C. Aime, K. Dohlen, D. Mourard, R. Lyon, and K. G. Carpenter. [Luciola hypertelescope space observatory : versatile, upgradable high-resolution imaging, from stars to deep-field cosmology](#). Experimental Astronomy, vol. 23:463–490, March 2009.
<http://link.springer.com/content/pdf/10.1007%2Fs10686-008-9123-8.pdf>
(the proposal previously submitted to ESA is posted at:
<http://www.oamp.fr/infoglueDeliverLive/www/OHP/Actualit%E9s?contentId=1148>)
6. A. Labeyrie, H. L. Coroller, S. Residori, U. Bortolozzo, J. Huignard, and P. Riaud. Resolved Imaging of Extra-Solar Photosynthesis Patches with a “Laser Driven Hypertelescope Flotilla”. In Pathways Towards Habitable Planets, Astronomical Society of the Pacific Conference Series [V. Coude Du Foresto, D. M. Gelino, and I. Ribas, editors], vol. 430 : 239, October 2010.
http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?

7. A.Labeyrie et al., “EPICURUS- a hypertelescope concept” , proposal submitted to ESA (2000)
<http://www.oamp.fr/lise/publis/epicurusProposal.pdf>
8. A. Labeyrie. [Resolved imaging of extra-solar planets with future 10-100km optical interferometric arrays](#). A&A, vol. 118:517–524, September 1996.
9. E. Pedretti, A. Labeyrie, L. Arnold, N. Thureau, O. Lardiere, A. Boccaletti, and P. Riaud. [First images on the sky from a hyper telescope](#). A&A, vol. 147:285–290, December 2000.
10. S. Gillet, P. Riaud, O. Lardiere, J. Dejonghe, J. Schmitt, L. Arnold, A. Boccaletti, D. Horville, and A. Labeyrie. [Imaging capabilities of hypertelescopes with a pair of micro-lens arrays](#). A&A, vol. 400:393–396, March 2003.
11. O. Lardiere, F. Martinache, and F. Patru. [Direct imaging with highly diluted apertures - I. Field-of-view limitations](#). MNRAS, vol. 375:977–988, March 2007.
12. F. Patru, N. Tarmoul, D. Mourard, and O. Lardiere. [Direct imaging with highly diluted apertures - II. Properties of the point spread function of a hypertelescope](#). MNRAS, vol. 395:2363–2372, June 2009.
13. Labeyrie, A., Authier, B., de Graauw, T., Kibblewhite, E. Weigelt, G., "TRIO, a kilometric optical array stabilized by solar sails" 1985, Proc. ESA coll. Kilometric Optical Arrays n Space, SP 226, Cargèse, 27-33
14. D. Ricci, H. Le Coroller, A. Labeyrie, and P. Piron. [Simulations of coronagraphy with a dynamic hologram for the direct detection of exo-planets](#). In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 7731, July 2010.
15. Carpenter et al., “Technology Development for Future Sparse Aperture Telescopes and Interferometers in Space”, 2009, White paper for the US Decadal Survey
16. Boccaletti, A., Riaud, P., Moutou, C., & Labeyrie, A., 2000, Icarus, 145, 2, 628
17. A.Labeyrie “Snapshots of Alien Worlds--The Future of Interferometry”, Science 17 September 1999, Vol. 285 no. 5435 pp. 1864-1865
18. A. Léger, M. Fontecave, A. Labeyrie, B. Samuel, O. Demangeon, and D. Valencia. “Is the Presence of Oxygen on an Exoplanet a Reliable Biosignature?” Astrobiology. May 2011, 11(4)
19. Labeyrie, A., “Gravitational lenses as giant diffractive telescopes”, Astronomy and Astrophysics (ISSN 0004-6361), vol. 284, no. 2, p. 689-692

The ODINUS Mission Concept

The Scientific Case for a Mission to the Ice Giant Planets with Twin spacecraft to Unveil the History of our Solar System



Spokeperson:

Diego Turrini
Istituto di Astrofisica e Planetologia Spaziali INAF-IAPS
Via del Fosso del Cavaliere 100
00133 - Rome, Italy
Email: diego.turrini@iaps.inaf.it

Authors list

Diego Turrini¹, Romolo Politi¹, Roberto Peron¹, Davide Grassi¹, Christina Plainaki¹, Mauro Barbieri², David M. Lucchesi¹, Gianfranco Magni¹, Francesca Altieri¹, Valeria Cottini³, Nicolas Gorius⁴, Patrick Gaulme⁵, François-Xavier Schmider⁶, Alberto Adriani¹, Giuseppe Piccioni¹

- (1) Institute for Space Astrophysics and Planetology INAF-IAPS, Italy.
- (2) Center of Studies and Activities for Space CISAS, University of Padova, Italy.
- (3) University of Maryland, USA.
- (4) Catholic University of America, USA
- (5) New Mexico State University, USA
- (6) Laboratoire Lagrange, Observatoire de la Côte d'Azur, France

Supporters list

Roberto Orosei (INAF/IAPS, Italy), Giovanna Rinaldi (IAPS-INAF, Italy), Ernesto Palomba (IAPS-INAF, Italy), Alessandra Migliorini (IAPS-INAF, Italy), Maria Cristina De Sanctis (IAPS-INAF, Italy), Emiliano D'Aversa (INAF/IAPS, ITALY), Fabrizio Oliva (IAPS-INAF, Italy), Anna Milillo (INAF-IAPS, Italy), Agustín Sánchez-Lavega (Universidad del País Vasco UPV/EHU, Spain), Yves Langevin (Institut d'Astrophysique Spatiale, France), Patrick Irwin (University of Oxford, United Kingdom), Joern Helbert (DLR, Germany), J.-C. Gerard (LPAP- Univ. de Liege, Belgium), Rumi Nakamura (IWF/OEAW, Austria), Bortolino Saggini (Politecnico di Milano, Italy), Bianca Maria Dinelli (ISAC-CNR, Italy), Nikolay Ignatiev (Space Research Institute of Russian Academy of Sciences, Russia), David Luz (CAAUL/University of Lisbon, Portugal), Ioannis A. Daglis (Department of Physics - University of Athens, Greece), Glenn Orton (Jet Propulsion Laboratory- California Institute of Technology, USA), Alexander Rodin (Moscow Institute of Physics and Technology, Russia), Themelis Konstantinos (National Observatory of Athens, Greece), Britney E Schmidt (Georgia Institute of Technology, United States), Thomas B. McCord (Bear Fight Institute, USA), Randy Gladstone (SwRI, USA), H. MAVROMICHALAKI (University of Athens, Greece), anna maria fioretti (CNR - Institute of Geosciences and Earth Resources, Italy), Kurt Retherford (Southwest Research Institute, USA), Michael Davis (Southwest Research Institute, United States of America), Bertrand Bonfond (Université de Liège, Belgium), Iannis Dandouras (IRAP, France), Dario De Fazio (IMIP-CNR, Italia), Leo Girardi (INAF - Osservatorio Astronomico di Padova, Italy), Tiziano Maestri (University of Bologna, Italy), Mauro Focardi (INAF-OAA, Italy), Jean-Baptiste Vincent (MPS, Germany), Neil Murphy (Jet Propulsion Lab, US), Mark Hofstadter (JPL/Caltech, United States), Anna Cinzia Marra (CNR-ISAC, ITALY), Vincenzo della Corte (Università degli Studi di Napoli Parthenope, Italy), Arnold (DLR, Germany), Gian Paolo Marra (CNR-ISAC , ITALY), Maurizio Pajola (CISAS - University of Padova, Italy), Maarten Roos (Lightcurve Films, Portugal), Cesare Grava (Southwest Research Institute, USA), and many others (see ODINUS website for the full list).

ODINUS website and full list of supporters

<http://odinus.iaps.inaf.it/>

Overview

The planets of our Solar System are divided in two main classes: the terrestrial planets, populating the inner Solar System, and the giant planets, which dominate the outer Solar System. The giant planets, in turn, can be divided between the gas giants Jupiter and Saturn, whose mass is mostly constituted by H and He, and the ice giants Uranus and Neptune, whose bulk composition is instead dominated by the combination of the astrophysical ices H_2O , NH_3 and CH_4 with metals and silicates. While in the case of the gas giants H and He constitutes more than 90% of their masses, in the case of the ice giants these gaseous envelopes are more limited, amounting to only 1-4 Earth masses (De Pater and Lissauer 2010). The terrestrial planets and the gas giants have been extensively studied with ground-based observations and with a large numbers of dedicated space missions. The bulk of the data on the ice giants, on the contrary, has been supplied by the Voyager 2 mission, which performed a fly-by of Uranus in 1986 followed by one of Neptune in 1989.

The giant planets appeared extremely early in the history of the Solar System, forming across the short time-span when the Sun was still surrounded by a circumstellar disk of gas and dust and therefore predating the terrestrial planets. The role of the giant planets in shaping the formation and evolution of the young Solar System was already recognized in the pioneering works by Oort and Safronov in 1950-1960. In particular, Safronov (1969) suggested that the formation of Jupiter would inject new material, in the form of planetesimals scattered by the gas giant, in the formation regions of Uranus and Neptune. More recently, the renewed understanding of planetary formation we obtained by the study of extrasolar planetary systems gave rise to the idea that the Solar System could have undergone a much more violent evolution than previously imagined (e.g. the Nice Model for the Late Heavy Bombardment, Tsiganis et al. 2005), in which the giant planets played the role of the main actors in shaping the current structure of the the Solar System.

The purpose of this document is to discuss the scientific case of a space mission to the ice giants Uranus and Neptune and their satellite systems and its relevance to advance our understanding of the ancient past of the Solar System and, more generally, of how planetary systems form and evolve. As a consequence, the leading theme of this proposal will be the first scientific theme of the Cosmic Vision 2015-2025 program:

- What are the conditions for planetary formation and the emergence of life?

In pursuing its goals, the present proposal will also address the second and third scientific theme of the Cosmic Vision 2015-2025 program, i.e.:

- How does the Solar System work?
- What are the fundamental physical laws of the Universe?

The mission concept we will illustrate in the following will be referred to through the acronym **ODINUS**, this acronym being derived from its main fields of scientific investigation: **Origins, Dynamics and Interiors of Neptunian and Uranian Systems**. As the name suggests, the ODINUS mission is based on the use of two twin spacecraft to perform the exploration of the ice giants and their regular and irregular satellites with the same set of instruments. This will allow to perform a comparative study of these two systems so similar and yet so different and to unveil their histories and that of the Solar System.

Theme 1: What are the conditions for planetary formation and the emergence of life?

In this section we will briefly summarize how our understanding of the processes of planetary formation has evolved across the years, discuss their chronological sequence for what concerns the Solar System and highlight how the exploration of Uranus, Neptune and their satellite systems can provide deeper insight and better understanding of the history of the Solar System.

The Evolving View of Planetary Formation: Solar System and Exoplanets

The original view of the set of events and mechanisms that characterize the process of planetary formation (Safronov 1969) was derived from the observation of the Solar System as it is today. This brought to the assumption that planetary formation was a local, orderly process that produced regular, well-spaced and, above all, stable planetary systems and orbital configurations. However, with the discovery of more and more planetary systems through ground-based and space-based observations, it is becoming apparent that planetary formation can result in a wide range of outcomes, most of them not necessarily consistent with the picture derived from the observations of the Solar System.

The orbital structure of the majority of the discovered planetary systems seems to be strongly affected by planetary migration due to the exchange of angular momentum with the circumstellar disks (see e.g. Papaloizou et al. 2007 and references therein), in which the forming planets are embedded, and by the so-called “Jumping Jupiters” mechanism (Weidenschilling & Marzari 1996; Marzari & Weidenschilling 2002), which invoke multiple planetary encounters, generally after the dispersal of the circumstellar disk, with a chaotic exchange of angular momentum between the different bodies involved.

The growing body of evidence that dynamical and collisional processes, often chaotic and violent, can dramatically influence the evolution of young planetary systems gave rise to the idea that also our Solar System could have undergone the same kind of evolution and represent a “lucky” case in which the end result was a stable and regular planetary system. The most successful attempt to describe the evolution of the Solar System to the present epoch has been the so-called Nice Model (Gomes et al. 2005; Tsiganis et al. 2005; Morbidelli et al. 2005; Morbidelli et al. 2007; Levison et al. 2011). The Nice Model is a Jumping Jupiter scenario formulated to link the event known as the Late Heavy Bombardment (LHB, see e.g. Hartmann et al. 2000 for a review) to a migration event involving all the giant planets.

In the Nice Model, the giant planets of the Solar System are postulated to be initially located on a more compact orbital configuration than their present one and to interact with a massive primordial trans-Neptunian region. The gravitational perturbations among the giant planets are initially mitigated by the trans-Neptunian disk, whose population in turn is eroded. Once the trans-Neptunian disk becomes unable to mitigate the effects of the interactions among the giant planets, the orbits of the latter become excited and a series of close encounters takes place. The net result of the Jumping Jupiters mechanism in the Nice Model is a small inward migration of Jupiter and marked outward migration of Saturn, Uranus and Neptune (Tsiganis et al. 2005).

The importance of the Nice Model lies in the fact that it strongly supports the idea that the giant planets did not form where we see them today or, in other words, that what we observe today is not necessarily a reflection of the Solar System as it was immediately after the end of its formation process. Particularly interesting in the context of this proposal is that, in about half the cases considered in the Nice Model scenario, Uranus and Neptune swapped their

orbits (Tsiganis et al. 2005). The success of the Nice Model in explaining several features of the Solar System opened the road to more extreme scenarios, also based on the Jumping Jupiters mechanism, either postulating the existence of a now lost fifth giant planet (Nesvorný et al. 2011) or postulating an earlier phase of migration and chaotic evolution more violent and extreme than the one described in the Nice Model (Walsh et al. 2011).

In strict relation with the idea of the giant planets migration, one of the most fascinating aspects of these scenarios is that they all invoke a certain degree of mixing of the solid materials that compose the Solar System. The mixing is generally the larger the more the causing event is located toward the beginning of the Solar System lifetime. As an example, the “Grand Tack” scenario by Walsh et al. (2011, 2012) implies a much stronger remixing than the one that the LHB would cause in the framework of the Nice Model (see e.g. Levison 2009). However, a more or less extensive migration of the giant planets is not required to have a remixing of the solid material in the Solar Nebula. Safronov (1969) pointed out that the formation of Jupiter would scatter the planetesimals in its vicinity both inward and outward respect to its orbital region. The outward flux of ejected material was postulated to rise the density of solid material in the formation regions of Uranus and Neptune and increase their accretion rate.

The inward flux instead crosses the regions of the terrestrial planets and the asteroid belt, with potentially important implications for the collisional evolution of the primordial planetesimals (Weidenschilling 1975, Weidenschilling et al. 2001; Turrini et al. 2011, 2012). The influence of Jupiter's formation, however, is not limited to the scattering of neighboring planetesimals: the orbital resonances with the planet would extract planetesimals from farther away regions and put them on orbits crossing those of the other forming giant planets. One of the regions affected by the orbital resonances is the asteroid belt (Turrini et al. 2011, 2012): rocky material is therefore extracted from the inner Solar System and, as in the original idea from Safronov (1969), possibly accreted by the forming cores of Uranus and Neptune or captured in their circumplanetary disk and incorporated in their satellites.

The Role of Ice Giants in Unveiling the Past of Solar System

As discussed in the previous section, during its history the Solar System went through a series of violent processes that shaped its present structure. The main actors of these processes were the giant planets. Due to their smaller masses and their likely later formation, Uranus and Neptune were also strongly affected by these very same processes. In this section, we will reorganize the events discussed in the previous section in a chronological order and discuss their implications for Uranus and Neptune and their satellite systems. If we follow the description of the history of the Solar System by Coradini et al. (2011), we can divide it into three main phases: the Solar Nebula, the Primordial and the Modern Solar System. This schematic view of the evolution of the Solar System is summarized in Fig. 1, where we reports the main events that took place across the different phases.

The Solar Nebula

From the point of view of the giant planets, the Solar Nebula (see Fig. 1) is the period across which they were forming in the circumsolar disk and migrating due to disk-planet interactions. While the giant planets are forming, their gravitational perturbations on the protoplanetary disk cause a sequence of bombardment events that Coradini et al. (2011) called the Primordial Heavy Bombardment. One of the consequences of this Primordial Heavy Bombardment is that, after the formation of the first giant planet, each successive giant planet forms from a more and more evolved and remixed disk, in which the abundances of different elements and materials are different from the original ones, with implications for

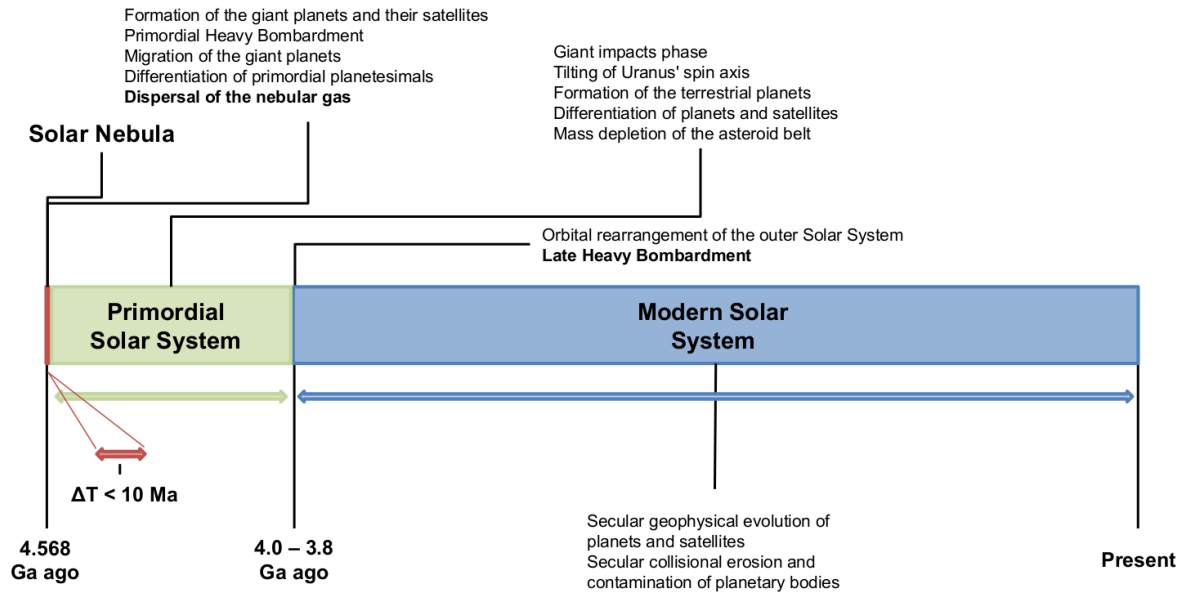


Figure 1: time-line of the history of the Solar System following the division in three phases (Solar Nebula, Primordial and Modern Solar System) proposed by Coradini et al. (2011). The events marking the transition between the different phases are in bold characters.

the rock/ice ratio and the ratio between different ices in the cores of the giant planets and in the material available for the forming satellites. In the standard view of the Solar System formation (Safronov 1969), the migration of the giant planets due to their exchange of angular momentum with the circumsolar disk was limited and the main role in reshuffling the protoplanetary disk was played by the Primordial Heavy Bombardment. However, in the alternative views we presented in the previous section, the migration of the giant planets could have played a significant role in the reshuffling of the different materials in the Solar System. In the “Grand Tack” scenario (Walsh et al. 2011, 2012) the giant planets are hypothesized to migrate extensively across the Solar System. Their formation regions, in this case, would be markedly different from those assumed by the standard scenario and the composition of their planetary cores would be affected by it. Moreover, part of the planetesimals that the giant planets scatter while migrating would collide with the giant planets themselves, contributing to the late accretion of high-Z elements first hypothesized by Owen et al. (1999) to explain the super-solar abundances of C, N, S, Ar, Kr and Xe in the atmosphere of Jupiter. All these remixing events, moreover, affect the source materials, captured in the form of planetesimals by the circumplanetary disks, from which the regular satellites of the giant planets can form (see Coradini et al. 2010 for a review). Depending on the formation time of the relevant giant planet and on the amount of radiogenic sources (incorporated in the rocky fraction of the source material), the regular satellites could already differentiate across this phase of the life of the Solar System. Finally, across the Solar Nebula phase a first generation of irregular satellites of the giant planets could have been captured from the protoplanetary disk due to collisions, the effects of gas drag or a combination of the two (see e.g. Mosqueira et al. 2010 for a discussion). This first generation of irregular satellites, however, would not survive the LHB if the latter is associated to a violent rearrangement of the Solar System like the one hypothesized by the Nice Model.

The Primordial Solar System

Somewhere between the Solar Nebula and the Primordial Solar System phases, two events contributed to shape the Uranian and Neptunian satellite systems. One was the giant impact of a planetary embryo with Uranus, suggested to be responsible for its 98° obliquity. As discussed by Coradini et al. (2010), it is possible that the original satellite system of the ice giant was destroyed during this event and new satellites formed from the debris of the original ones. The second event was the capture of Triton by Neptune and the following shrinking and circularizing of its orbit, which caused the removal of most of the original regular satellites of the ice giant. Across these events and throughout the Primordial Solar System, the Nice Model predicts that the giant planets would still be on different, closer orbits with respect to their present ones. Once the dynamical instability responsible for the LHB takes place, icy planetesimals from what will become the trans-neptunian region are excited into high-eccentricity, giant planet-crossing orbits analogous to those of the present-day Centaurs. A fraction of these planetesimals will impact against the giant planets, possibly contributing to the late enrichment of their atmospheres (Matter et al. 2009). A fraction of these planetesimals will also impact on the satellites of the giant planets, contributing to their contamination by exogenous material and possibly supplying energy for their late differentiation (Barr & Canup 2010). Barr & Canup (2010) argue that the LHB can bring to the differentiation of Ganymede but not to that of Callisto, in agreement with the available data on their internal structure. Matter et al. (2009) assessed instead the amount of high-Z elements that would be accreted by the four giant planets during the LHB, finding that it is insufficient to explain or significantly contribute to the observed values. Another implication of the Nice Model is that any pre-existing population of irregular satellites would be destroyed as a consequence of the close encounters between the giant planets (Tsiganis et al. 2005). Nesvorný et al. (2007) however showed that three-body effects between the giant planets and the planetesimals during the planetary encounters invoked by the Nice Model would naturally supply a way to re-populate the satellite systems of the giant planets by irregular satellites. It must be noted that these studies are based on the earlier formulation of the Nice Model (Tsiganis et al. 2005; Gomes et al. 2005; Morbidelli et al. 2005) and that the implications of its more recent formulation (Morbidelli et al. 2007; Levison et al. 2011) are still to be addressed. Nevertheless, they show that the evolution of the Solar System across the Primordial Solar System phase could have a non-negligible role in shaping the present-day Uranus and Neptune and their satellite systems.

The Modern Solar System

The Modern Solar System phase starts after the end of the LHB and, differently from the previous two phases, instead of violent processes it is dominated by more regular, secular ones. Moreover, the population of small bodies in the outer Solar System is significantly smaller than that at earlier times, so that collisional processes are less intense than before. Most of the information that we can gather through crater counting on the surface of the satellites of the giant planets refers to this long, more quiescent phase, especially if the satellites are still geophysically active and undergo resurfacing, as it appears to be the case of Triton (see Schubert et al. 2010 for a discussion). In the case of geophysically active satellites, moreover, the surface features and composition supply us information on their more recent internal state, i.e. they again give us insight on the processes that acted across the Modern Solar System phase. Depending on the degree of geophysical activity and the flux of impactors (being them planetocentric, i.e. other satellites, or heliocentric, e.g. comets and Centaurs), the surfaces of the satellites can be contaminated to various degrees by exogenous

material (see e.g. Mosqueira et al. 2010, Schubert et al. 2010 for a discussion), an effect that has to be taken into account while interpreting e.g. spectral data, as spectrometers allow to probe the composition of a very thin layer (~cm-sized) of the satellites surfaces. Across the Modern Solar System, moreover, the secular effects of space-weathering due to various exogenic sources (e.g. solar wind, magnetospheric plasma, cosmic rays) contributed to the surface evolution of the satellites in ways that are still poorly quantified or even understood.

The ODINUS mission and the history of the Solar System

As the previous sections highlight, our view of the processes of planetary formation and of the evolution of the Solar System has greatly changed across the last twenty years but most of the new ideas are in the process of growing to full maturity or need new observational data to test them against. The ODINUS mission aims to address these open problems by exploring the systems of Uranus and Neptune, as they are the most affected from the violent processes that sculpted the early Solar System and yet they are the least explored and more mysterious ones.

The primary information that the ODINUS mission wants to gather by exploring the Uranian and Neptunian systems are:

- What is the atmospheric composition and enrichment with respect to solar abundances of the two planets?
- What are the bulk densities and the masses of the ice giants and their satellites?
- What are the interior structures and density profiles of the ice giants and their satellites?
- What is the surface composition of the regular and irregular satellites?
- Which satellites are fully or partially differentiated and which ones are undifferentiated?

Using these data, the open questions that ODINUS aims to answer are:

- When and where did the planets form? Did they migrate? If so, how much? Did Uranus and Neptune swap their positions as hypothesized by the Nice Model?
- What is the ice-to-rock ratio of the cores of ice giants and of their satellites? How much “non-local” material was available to them when they formed? Where did this “non-local” material originated from?
- Are the satellites of Uranus primordial or they reformed after the planet tilted its spin axis? What were the effects of the capture of Triton for the Neptunian satellites?
- Where did the irregular satellites originate? Can they be used to constrain the dynamical evolution of the ice giants?

Theme 2: How does the Solar System work?

In gathering the data that will allow to address Theme 1 of the Cosmic Vision 2015-2025 program, the ODINUS mission will gather a wealth of data on the present status of the Uranian and Neptunian systems. While the twin spacecraft setup constrains the number of instruments on-board each spacecraft, the goal of the ODINUS mission is to perform a global survey as complete as possible of the two giant planets and their satellites. The data, which ODINUS will collect, will allow to gain a more complete understanding of how icy satellites so far away from the Sun evolve both for what it concerns their surfaces and their interiors. Moreover, the coupled investigation of these two planets, so similar and yet so different, will allow to better understand the sources of their different atmospheric and thermal behavior.

Atmospheres of Uranus and Neptune

The Herschel observations of Uranus and Neptune (Feuchtgruber et al., 2013) confirmed that the ice giants have a remarkably similar D/H content ($4.4 \pm 0.4 \times 10^{-5}$ and $4.1 \pm 0.4 \times 10^{-5}$ respectively), suggesting a common source of icy planetesimals in the protoplanetary disk. Further insight on the conditions of the disk in its outer regions can be derived from the relative enrichment (with respect to the Solar values) of C, N, S and O, by determination of the abundances of the corresponding reduced forms. At the current date, methane is still the only minor atmospheric constituent that has been directly detected in both ice giants (e.g.: Baines et al., 1994); an extensive investigation in this field is therefore extremely urgent to ultimately characterize the emergence of our solar system.

The post-Voyager 2 observations of Uranus by ground-based and space telescopes revealed a progressive increase of meteorological activity (cloud and dark spots occurrence) in the proximity of Northern Spring equinox (see, e.g. Sromovsky et al., 2012). While this evolution is undoubtedly related to the extreme obliquity of the planet, the relative roles of solar illumination and internal heating (and its possible variations) remain to be assessed by detailed studies at high spatial resolution.

The possibility to compare the atmospheric behaviour of Uranus with the extremely dynamic meteorology of Neptune – apparently characterized by a slower long-term evolution – provides a unique opportunity to gain insights on the response of thick atmospheres to time-variable forcing, representing therefore a new area of tests for future atmospheric global circulation models, in conditions not found in terrestrial planets or gas giants.

Uranus zonal winds are currently characterized by moderately retrograde values (-50 m/s) at the equator that progressively become prograde, to reach a maximum value of 200 m/s at 50N (Sromovsky et al., 2012). On Neptune, a similar pattern is observed, but the absolute speed values are strongly amplified, to reach – despite the limited solar energy input – the extreme values (400m m/s or more) observed in the Solar System (Shuleen Chau et al., 2012). Wind speed fields are the most immediate proxy for atmospheric circulation and their modeling can provide constraints on very general properties of the atmosphere, such as the extent of deep convection (Suomi et al., 1991).

While the efforts of ground based observers has allowed to considerably expand the results of Voyager 2, an extensive, long-term, and high spatial resolution cloud tracking remains essential to study the ultimate causes of these extreme phenomena.

Neptune shows an unexpected temperature of 750 K in its stratosphere (Broadfoot et al., 1989) that cannot be justified by the small solar UV flux available at that heliocentric distance. More complex mechanisms – such as energy exchange with magnetospheric ions – shall become predominant in these regions. Uranus, on the other hand, offers unique magnetospheric geometries because of its high obliquity and strong inclination of magnetic axis.

The satellites of Uranus and Neptune

The satellites of Uranus and Neptune are poorly known, mostly due to the limited coverage and resolution of the Voyager 2 observations. The Uranian satellites Ariel and Miranda showed a complex surface geology, dominated by extensional tectonic structures linked to their thermal and internal evolution (Prockter et al. 2010 and references therein). Umbriel appeared featureless and dark, but the analysis of the images suggests an ancient tectonic system (Prockter et al. 2010 and references therein). Little is known about Titania and Oberon, as the resolution of the images taken by Voyager 2 was not enough to distinguish tectonic features. The partial coverage of the surface of Triton revealed one of the youngest

surfaces of the Solar System, suggesting the satellite is possibly more active than Europa (Schubert et al. 2010 and references therein). Notwithstanding this, the surface of Triton showed a variety of cryovolcanic, tectonic and atmospheric features and processes (Prockter et al. 2010 and references therein).

From the point of view of their surface composition, the Uranian satellites are characterized by the presence of crystalline H₂O ice (Dalton et al. 2010). The spectral features of Ariel, Umbriel and Titania showed also the presence of CO₂ ice, while CO₂ ice was not observed on Oberon (Dalton et al. 2010 and references therein). In the case of Miranda, the possible presence of ammonia hydrate was observed but both the presence of the spectral band and its interpretation are to be confirmed (Dalton et al. 2010 and references therein). The confirmation of the presence of ammonia would be of great importance due to its anti-freezing role in the interior of the satellites. The spectra of Triton possess the absorption bands of five ices: N₂, CH₄, CO, CO₂, and H₂O (Dalton et al. 2010). The detection of the HCN ice band has been reported, which could imply the presence of more complex materials of astrobiological interest (see Dalton et al. 2010 and references therein). Triton also possesses a tenuous atmosphere mainly composed by N₂ and CO, which undergoes seasonal cycles of sublimation and recondensation (see Dalton et al. 2010 and references therein). Images taken by Voyager 2 revealed active geyser-like vents on the surface of Triton, indicating that the satellite is still geologically active even if at present it is not tidally heated (Schubert et al. 2010).

Both Uranus and Neptune possess a family of irregular satellites. Neptune, in particular, possesses the largest irregular satellite in the outer Solar System (not counting Triton), i.e. Nereid. Aside their estimated sizes and the fact that observational data suggest they might be more abundant than those of Jupiter and Saturn (Haghighipour and Jewitt 2007), almost nothing is known of these bodies.

Magnetosphere-Exosphere-Ionosphere Coupling in the Uranian and Neptunian systems

The highly non-symmetric internal magnetic fields of Uranus and Neptune, coupled with the relatively fast rotation and the unusual inclination of the rotation axes to the orbital planes imply that their magnetospheres are subject to drastic geometrical variations on both diurnal and seasonal timescales. The relative orientations of the planets' spin axis, their magnetic dipole axis and the direction of the solar wind flow determine the configuration of each magnetosphere and, consequently, the plasma dynamics in these regions.

Due to the planet's large obliquity, Uranus' asymmetric magnetosphere varies from a pole-on to orthogonal configuration during an Uranian year (84 Earth years) and changes from an "open" to a "closed" configuration during an Uranian day. At solstice (when Uranus' magnetic dipole simply rotates around the vector of the direction of the solar wind flow) plasma motions produced by the rotation of the planet and by the solar wind are effectively decoupled (Selesnick and Richardson, 1986; Vasyliunas, 1986). Moreover, the Voyager 2 plasma observations showed that when the Uranus dipole field is oppositely directed to the interplanetary field, injection events to the inner magnetosphere (likely driven by reconnection every planetary rotation period) are present (Sittler et al., 1987). The time-dependent modulation of the magnetic reconnection sites, the details of the solar wind plasma entry in the inner magnetosphere of Uranus and the properties of the plasma precipitation to the planet's exosphere and ionosphere are unknown. Models indicate that Uranus' ionosphere is dominated by H⁺ at higher altitudes and H₃⁺ lower down (Capone et al., 1977; Chandler and Waite, 1986; Majeed et al., 2004), produced by either energetic particle precipitation or solar ultraviolet (UV) radiation. Our current knowledge on the aurora of Uranus is limited since it

is based only on: one spatially resolved observation of the UV aurora (by the Ultraviolet Spectrograph data on board Voyager 2, Herbert 2009); observations of the FUV and IR aurora with the Hubble Space Telescope (Ballester, 1998); and on observations from ground-based telescopes (e.g., Trafton et al., 1999). The details of the solar wind plasma interaction with the planet's exosphere, ionosphere and upper atmosphere (through charge-exchange, atmospheric sputtering, pick-up by the local field), the seasonal and diurnal variation of the efficiency of each mechanism as well as the total energy balance (deposition/loss) due to magnetosphere-exosphere-ionosphere coupling are unknown. Since the exact mechanism providing the required additional heating of the upper atmosphere of Uranus is also unknown, new in situ plasma and energetic neutral particles observations could become of particular importance in order to determine whether plasma precipitation play a key role in this context. The magnetospheric interaction with the Uranian moons can be studied through in situ measurements of magnetic field, particles, and energetic neutrals emitted from the surfaces. Finally, remote imaging of charge exchange energetic neutral atoms would offer a unique opportunity to monitor the plasma circulation where moons and/or Uranus exosphere are present.

Neptune's magnetic field has a complex geometry that includes relatively large contributions from non-dipolar components, including a strong quadrupole moment that may exceed the dipole moment in strength. Unlike Uranus, however, Neptune has shown no evidence of UV emission that could be associated with auroral activity. Although this non-observation did not rule out an active magnetosphere *per se*, it ruled out processes similar to those associated with the aurora observed at Uranus. Whereas the plasma in the magnetosphere of Uranus has a relatively low density and is thought to be primarily of solar-wind origin, at Neptune, the distribution of plasma is generally interpreted as indicating that Triton is a major source (Krimigis et al., 1989; Mauk et al., 1991, 1994; Belcher et al., 1989; Richardson et al., 1991). Escape of neutral hydrogen and nitrogen from Triton maintains a large neutral cloud (Triton torus) that is believed to be source of neutral hydrogen and nitrogen (Decker and Cheng, 1994). The escape of neutrals from Triton could be an additional plasma source for the Neptune's magnetosphere (through ionization). Our knowledge on the plasma dynamics in the magnetosphere of Neptune as well as on the neutral particles production in Triton's atmosphere is limited. New in situ plasma and energetic neutral particles observations focused in the Triton region can provide important information on the role of the combined effects of photoionization, electron impact ionization, and charge exchange in the context of the coupling of a complex asymmetric planetary magnetosphere with a moon exosphere at large distances from the Sun.

Planetary and satellite interiors

The available constraints on interior models of Uranus and Neptune are limited. The gravitational harmonics of these planets have been measured only up to fourth degree (J₂, J₄), and the planetary shapes and rotation periods are not well known (see e.g. Helled et al. 2011 and references therein). The response coefficients of Uranus and Neptune suggest that the latter is less centrally condensed than the former (De Pater and Lissauer 2010).

The thermal structures of these planets are also intriguing (see e.g. Helled et al. 2011 and references therein). Uranus stands among the planets for the extremely low value of $0.042 \pm 0.047 \text{ W/m}^2$ of its internal energy flux (Pearl et al., 1990). This figure sharply contrasts with Neptune, where Voyager 2 determined a value of $0.433 \pm 0.046 \text{ W/m}^2$ (Pearl et al., 1991). The two ice giants must therefore differ in their internal structure, heat transport mechanisms, and/or in their formation history. Substantial differences in internal structures are suggested by the analysis of available gravitational data for the two planets (Podolak et al., 1995).

Namely, the Uranus gravity data are compatible with layered convection in the shell, which inhibits the transport of heat. Alternative views call – among the others – for a later formation age of Neptune (Gudkova et al., 1988). Consequently, heat fluxes represent, along with gravity and magnetic data, the key experimental constraints to characterize the interior of Uranus and Neptune and their evolution.

The information on the interior structure of the satellites of Uranus and Neptune is even more limited and is mostly derived from their average densities, which are used to infer the rock-to-ice ratios, and their surface geology, which suggests that across their lives they possessed partially or completely molten interiors (De Pater and Lissauer 2010). As a consequence, the data that can be collected by the ODINUS mission on their interiors will play an important role in filling up this gap in our understanding of the icy satellites in the outer Solar System.

Gravity data can indeed be used to constrain the internal structure and composition of the planets. Deviations of the primary body gravitational field from the spherical symmetry (due to its rotational state and internal structure and composition) perturb the orbit of the spacecraft and can be extracted via a precise orbit determination and parameter estimation procedure from the tracking data, usually the range and the range rate in a typical Radio Science Experiment. Fundamental to this objective is a proper modeling of the spacecraft dynamics, both gravitational (e.g., gravitational multipoles) and non gravitational (e.g., radiation pressure). This could be non trivial in case of a complex spacecraft (the ideal would be a test mass) and – in selected cases – could require also the use of an on-board accelerometer (Iafolla et al., 2010). In the case of Uranus measurements of the precession of its elliptical rings should add to the list of observables. What said for the primaries extends to their satellites as well. Selected fly-bys to the satellites will allow for the determination of at least their lowest-degree multipoles.

An alternate and complementary method to probe the internal structures of Uranus and Neptune consists of using seismic techniques that were developed for the Sun (helioseismology, see e.g. Goldreich & Keeley 1977), then successfully applied to stars with the CoRoT and Kepler space missions (Michel et al. 2008, Borucki 2009), and tested on Jupiter (Gaulme et al. 2011). Seismology consists of identifying the acoustic eigen-modes, whose frequency distribution reflects the inner sound speed profile. The main advantage of seismic methods with respect to gravity moments is that waves propagate down to the central region of the planet, while gravitational moments are mainly sensitive to the external 20% of the planetary radius. The second advantage is that the inversion problem is not model dependent, neither on the equation of state or on the abundances that we want to measure. As regards Uranus and Neptune, the difference in internal energy flux should appear as a difference in the amplitude of acoustic modes. As for helioseismology, two approaches may be used to perform such seismic measurements, either with Doppler spectro-imaging (e.g. Schmider et al. 2007), or visible photometry (Gaulme & Mosser 2005). A dedicated study must be led to determine which method is the most appropriate for these two planets.

Heliosphere science

During the ODINUS mission cruise phase, it will be possible to obtain important information on the interplanetary medium properties at different distances from the Sun as well as on the heliosphere structure and its interactions with the interstellar medium. Although there is plenty of information on how solar wind and coronal mass ejections interact with the interplanetary medium at 1 AU from the Sun, little is known on how this interaction works at larger distances. The ODINUS measurements of the interplanetary magnetic field fluctuations and plasma densities variations, at different distances from the

Sun, can provide information for understanding the origin of turbulence in the solar wind and its evolution from its source to the heliopause. ODINUS, therefore, will give an opportunity to study space weather in the outer heliosphere and to understand how the interplanetary medium properties are modified in space and time.

The prevailing models of the shape of the heliosphere suggest a cometary-type interaction with a possible bow shock and/or heliopause, heliosheath, and termination shock (Axford, 1973; Fichtner et al., 2000). However, recent energetic neutral atom images obtained by the Ion and Neutral Camera (INCA) onboard Cassini did not conform to these models (Krimigis et al., 2009). Specifically, the map obtained by Cassini/INCA revealed a broad belt of energetic protons with non-thermal pressure comparable to that of the local interstellar magnetic field (Krimigis et al., 2009). In October 2008, Interstellar Boundary Explorer (IBEX) was launched with energetic neutral atom cameras specifically designed to map the heliospheric boundary at lower (<6 keV) energies (McComas et al., 2009; Funsten et al., 2009). Both IBEX and INCA identified in the energetic neutral atom images dominant topological features (ribbon or belt) that can be explained on the basis of a model that considers an energetic neutral atom-inferred non-thermal proton pressure filling the heliosheath from the termination shock to the heliopause (Krimigis et al., 2009).

During the cruise phase, the two spacecraft can be used measure the energetic neutral atoms produced by energetic singly charged particles in the heliosheath that charge-exchange with interstellar neutral hydrogen and enter the heliosphere unimpeded by the interplanetary magnetic field (Hsieh et al., 1992; Gruntman et al., 2001). Using also magnetic field measurements, the ODINUS can address the question whether the interaction of the heliosphere with the interstellar magnetic field takes place at the termination shock or at the heliopause.

How well do we know the distribution of mass in the Kuiper Belt?

The cruise phase of the two spacecraft to Uranus and Neptune offers the possibility to improve our current knowledge of the total mass and the mass distribution of the Kuiper Belt. Among the various methods used for constraining this distribution, the study of heliocentric orbits of objects in the Solar System (Anderson et al., 1995) applies well to ODINUS. The spacecraft approaching Uranus and Neptune in their cruise could be considered (as in the fundamental physics experiments) as test masses subject to the gravitational attraction of the Kuiper belt objects: the accurate tracking of the spacecraft will therefore help to further constrain the total mass and the mass distribution of these objects.

Theme 3: What are the fundamental physical laws of the Universe?

Since the early interplanetary exploration missions, spacecraft are used as (nearly) test masses to probe the gravitational machinery of Solar System and more in general to test for fundamental physics. Though general relativity is currently regarded as a very effective description of gravitational phenomena and it has passed all the experimental tests (both in the weak- and strong-field regimes) so far, it is challenged by theoretical (e.g. Grand Unification, Strings) scenarios and by cosmological findings (Turyshev, 2008). Stringent tests of general relativity have been obtained in the past by studying the motion of spacecraft in cruise, as well as the propagation of electromagnetic waves between spacecraft and Earth (see e.g. Bertotti et al. 2003). In this respect, the spacecraft are considered as test mass subject (mainly) to the gravitational attraction of Solar System bodies. Well-established equations of motions can then be tested against the experimental data, in order to place strong constraints

to possible deviations from what is predicted by general relativity. Also for what it concerns electromagnetic waves propagation experiments, the spacecraft act as a virtual bouncing point for microwave pulses, enabling a measure of the Shapiro time delay. Being very effective in the past in ruling out possibilities of "exotic physics" (i.e., the so-called "Pioneer Anomaly"), such tests could be used in the future to further pursue experiments in this way. The very-weak-field environment of the more external regions of the Solar System is particularly interesting, in that "exotic" phenomenology such as MOND could be probed. These tests would help extend the scale at which precision information on gravitational dynamics is available; this will contribute to bridge the "local" scale (in which precise measurements on gravitational dynamics are available) to more "global" scales (subject to puzzling phenomenology as dark matter and dark energy).

Scientific rationale of the twin spacecraft approach

The approach proposed for the ODINUS mission is to use a set of twin spacecraft (see Fig. 2), each to be placed in orbit around one of the two ice giant planets. The traditional approach for the exploration of the giant planets in the Solar System is to focus either on the study of a planetary body and its satellites (e.g. the Galileo and Cassini missions to the Jovian and Saturnian systems) or on the investigation of more specific problems (e.g. the Juno mission to study the interior of Jupiter and the JUICE mission to explore the Jovian moons Ganymede, Callisto and Europa). This is a well tested approach that allows for a thorough investigation of the subject under study and to collect large quantities of highly detailed data. The only drawback of this approach is that comparative studies of the different giant planets are possible only after decades, especially since the datasets provided by the different missions are not necessarily homogeneous or characterized by the same level of completeness, as the different missions generally focus on different investigations. In the case of the well-studied Jovian and Saturnian systems, about 10 years passed before it became possible to compare the dataset supplied by the Galileo mission with the first data supplied by the Cassini mission.

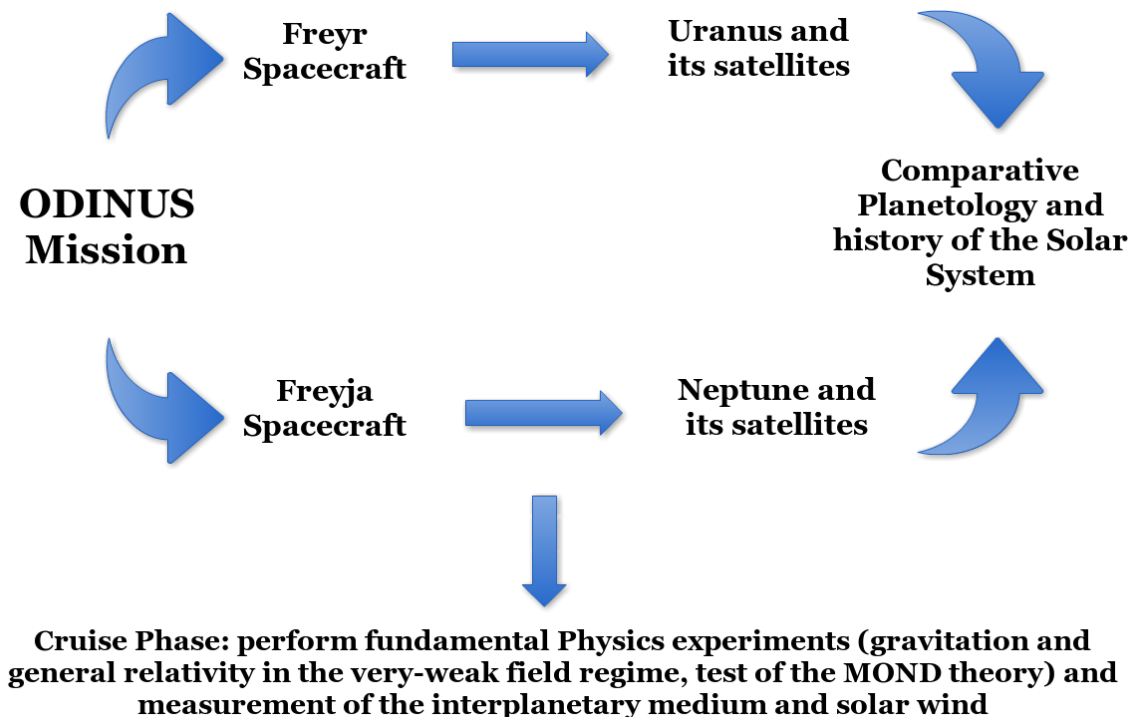


Figure 2: schematics of the twin spacecraft approach of the ODINUS mission.

However, in order to be able to perform a detailed comparative study of the satellites of these two giant planets it will be necessary to wait until the completion of the JUICE mission, due to the limited coverage of the data from Galileo. As a consequence, about half a century will be required before we can fully address the differences and similarities between the Jovian and Saturnian systems.

Exploring the Uranian and Neptunian systems with the traditional approach would require either half a century of efforts or the focus on this exclusive goal over the L2 and L3 missions of Cosmic Vision. In a scenario in which, to balance between the different needs of the astrophysics community, ESA would devote the L3 and L5 missions to the exploration of these two giant planets, the launch of the L5 missions would occur in 2046 or later (assuming a temporal distance between L5 and L4 and between L4 and L3 analogous to that between L3 and L2): assuming a travel time to Uranus and Neptune of about 13-15 years, as in the scenarios assumed for the Uranus Pathfinder (Arridge et al. 2012) and OSS (Christophe et al. 2012) mission proposals, the completion of the two missions would occur in about 2060, i.e. about half a century from now. In the unrealistic scenario of devoting both L2 and L3 missions to the exploration of the ice giants, it would be possible to complete this task by about 2050 but at the cost of not having L-class missions devoted to astrophysics before L4.

The approach proposed for the ODINUS mission is different from the traditional one in that it focuses on the use of two M-class spacecraft to be launched toward two different targets in the framework of the same mission. The use of two twin spacecraft, aside limiting the development cost of the mission, will allow to perform measurements with the same set of instruments in the Uranian and Neptunian systems, supplying data of similar quality and potentially completeness. Obviously, the need to produce and manage two spacecraft in place of one will limit the amount of instruments to be included in the scientific payload: this will translate in a less in-depth exploration of the two systems with respect to what would be possible with two dedicated missions. As we will discuss in the template mission profile, a careful selection of the instruments and design of the spacecraft can limit the importance of this drawback. Finally, we want to emphasize that, due to the different travel time to reach the two planets, the two spacecraft will not be operating at the same time except for short periods during the cruise phase, thus limiting the complexity of the mission management.

A model mission profile for ODINUS

To illustrate the feasibility and the critical aspects of the ODINUS mission concept, in the following we will discuss a model mission profile. We will illustrate the possible configuration of the two spacecraft and their scientific payload, the orbital paths that we believe could maximise the scientific return and the launch slot that could best suit the ODINUS mission.

The twin spacecraft

As we mentioned previously, the founding idea of the ODINUS mission concept is to have a set of twin spacecraft (which we dubbed Freyr and Freyja from the twin gods of the Norse pantheon) to be placed in orbit of Uranus and Neptune respectively. In order to fit the budget of an L-class mission, a conservative, straw-man configuration for the ODINUS mission could be based on two New Horizons-like spacecraft, i.e.:

- about 6 instruments in the scientific payload + radio science;
- about 500-600 kg of dry mass for each spacecraft;
- hybrid (ionic and chemical) propulsion;
- radioisotope-powered energy source.

The limitations on the scientific payload and the dry mass of the spacecraft come from a worst-case scenario evaluation of the fuel budget needed to reach the ice giants and to insert them on planetocentric orbits. If we consider the Hohmann transfer orbit between Earth and Uranus (or Neptune) with an orbital insertion at about 2×10^7 km from the relevant planet on a highly eccentric orbit, the required Δv of about 5 km/s translates into a wet-to-dry mass ratio of about 5 for each spacecraft. This implies that 600 kg of dry mass requires a wet mass at launch of about 3000 kg. Such a wet mass at launch would make the mission feasible either considering a single launch of the Freyr and Freyja spacecraft with an Ariane V rocket or two separate launches with Soyuz rockets. The scenario contemplating two separate launches allows the two trajectories to be optimized independently, thus allowing for the largest savings of either fuel or travel time, but a preliminary check of the orbital positions of Uranus and Neptune showed that the two ice giants will be in a favorable position to launch the two spacecraft together and then separate their paths at Uranus.

The post-insertion orbital paths of the spacecraft and the exploration strategy of the Uranian and Neptunian systems

The choices of the insertion orbit and of the hybrid propulsion system are motivated by the exploration strategy of the Uranian and Neptunian systems. The basic idea is to have the spacecraft enter their planetocentric orbits thanks to the chemical propulsion and then to take advantage of the ionic propulsion to slowly spiral inward toward the respective planets. The insertion orbits are chosen to insert the spacecraft in the orbital regions populated by the irregular satellites and have one or more fly-bys with members of this family of small bodies. The spacecraft will then spiral toward the regions populated by the regular satellites, possibly maintaining highly eccentric orbits to allow for the contemporary observation of the regular satellites and the planets or their ring systems.

The high obliquity values of Uranus and Neptune imply that the regular satellites orbit on planes significantly inclined with respect to the ecliptic plane. As a consequence, unless the fuel budget and the orbital studies indicate the possibility of inserting the spacecraft on high-inclination orbits, the orbital path of the spacecraft will need to be optimized to allow for as many close encounters as possible with the regular satellites in the lifetime of the mission. This is particularly important in the case of Uranus, where the satellites orbit almost perpendicularly to the ecliptic plane: a spacecraft orbiting near the latter would therefore allow only for short close encounters with the regular satellites when they are approaching and crossing the ecliptic plane itself.

A possible solution could be to take advantage of the ionic propulsion to make the orbits of the spacecraft precess: the resulting rosetta orbit should be optimized to allow the most close-encounters with the regular satellites. After the completion of the exploration of the regular satellites, the spacecraft would shrink their orbits again in order to approach the planets and focus the next phase of the mission to their study. A possible end-mission scenario would then be to take advantage of the ionic propulsion to slowly spiral the spacecraft inside the atmospheres of the planets and use the two spacecraft as atmospheric probes. If feasible, the use of the ionic propulsion to slow down the atmospheric descent would allow to circumvent the needs of heat shields on the spacecraft, thus reducing their weight.

The straw-man payload

A possible straw-man payload for the two spacecraft, which could allow for the achievement of the goals of the ODINUS mission, is composed by:

- Camera (Wide and Narrow Angle);

- VIS-NIR Image Spectrometer;
- Magnetometer;
- Mass Spectrometer (Ions and Neutrals, INMS);
- Doppler Spectro-Imager (for seismic measurements) or Microwave Radiometer;
- Radio-science package.

The choice to limit the number of instruments on-board the spacecraft is due to the budget constraint, i.e. to the need of keeping the ODINUS mission inside the total cost for an L-class mission. As we will discuss also in the next section, given the long times required to explore the ice giant planets, the development of a highly integrated payload, in order to maximize the number of instruments that can be fit in the spacecraft and thus the scientific return of the mission, is critical for the success of ODINUS. Two instruments that would significantly improve the completeness of the exploration of Uranus and Neptune and their satellites and the scientific return of the mission would be:

- Energetic Neutral Atoms Detector (to complement the measurements of the INMS);
- High-sensitivity Accelerometer (for the atmospheric descent phase).

As discussed in the section devoted to the study of the planetary interiors, an alternative approach based on seismologic measurements can be coupled to the more traditional study of the gravitational momenta to study the interiors of Uranus and Neptune. The ODINUS mission would be the ideal test-bed for this new kind of measurements, as the launch slot we suggest (2034, as discussed in the next section) would allow to assess which of the possible approaches (doppler-spectro imaging or visible photometry) is the most appropriate for ODINUS. Should visible photometry prove to be the technique of choice, the Doppler-Spectro Imager we indicated in the straw-man payload could be replaced by one (or more) of the alternative instruments we discussed (microwave radiometer, ENA detector, accelerometer).

Launch slot and timeline of the ODINUS mission

Given the technological challenges that the two-spacecraft approach of the ODINUS mission rises and the need to assess how to include seismological measurements among those performed by the spacecraft, we think that the optimal slot for ODINUS would be as the L3 mission of ESA Cosmic Vision 2015-2025 program, with the indicative launch foreseen for 2034. This would allow for enough time to develop the required enabling technologies (e.g. the radioisotope-powered energy source or a flight-qualified doppler spectro imager) and nevertheless, assuming an indicative time of flight of about 9 years to reach Uranus and 12 years to reach Neptune as achieved by the Voyager 2 mission, to complete the exploration of the outer Solar System by the first half of the century.

Critical aspects and enabling technologies of the ODINUS mission

As we highlighted in the previous sections, the ODINUS mission is in principle feasible with the present-day technology. The two spacecraft are modeled after the one of the ongoing New Horizons mission and their wet masses, according to our first order estimates, would fit either the Soyuz (two launches scenario) or the Ariane V (single launch scenario) payload capabilities. With an estimated final cost of about 550 MEuro (source: [NASA](#)) for the New Horizons mission and taking into account that the development costs would be shared between the two spacecraft, the ODINUS mission would be feasible also from the point of view of the expected cost.

The two most critical aspects for the success of the ODINUS mission are:

- the availability of radioisotope-powered energy sources;

- the possibility to achieve times of flight comparable with those of the Voyager 2 mission.

The first critical aspect is due to the large distances of Uranus and Neptune from the Sun, which make the use of solar panels for energy generation unpractical: the development of the required technology and the identification of an affordable and reliable energy source compliant with ESA's policies is therefore mandatory for the feasibility of the ODINUS mission. The second aspect is not critical for the feasibility of the mission: the Uranus Pathfinder (Arridge et al. 2012) and OSS (Christophe et al. 2012) mission studies already showed that the mission could be feasible even if on longer timescales (13-15 years of time of flight). Nevertheless, the duration of the mission is of major importance since it determines the possibility to perform a comparative study of the two systems in a reasonable timespan as well as it influences the management cost of the mission.

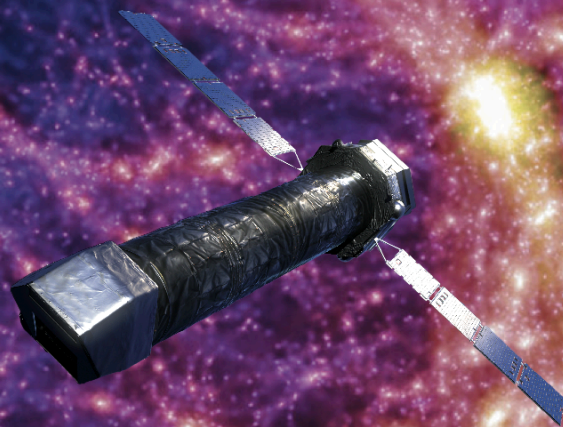
References

- Anderson, J.D. et al. 1995, *Astrophys. J.* 448, 885-892
- Arridge C. et al. 2012, *Experimental Astronomy* 33, 753-791
- Axford, W.I., 1973, *Space Sci. Rev.* 14, 582
- Baines, K.H. et al. 1995, *Icarus*, vol. 114, no. 2, p. 328-340
- Ballester, G.E., in: Wamsteker, W., Gonzalez Riestra, R. (eds.) *Ultraviolet Astrophysics Beyond the IUE Final Archive*, Proceedings of the Conference held in Sevilla, Spain, from 11–14 November 1997, ESA SP, vol. 413, p. 21. ESA Publications Division (1998)
- Barr, A. C., Canup, R. M. 2010, *Nature Geoscience* 3, 164-167.
- Belcher, J. W., et al. 1989, *Science*, 246, 1478,
- Bertotti, B et al. 2003, *Nature* 425, 374-376
- Borucki W.J., Koch D., Jenkins J., et al., 2009, *Science*, 325, 709
- Broadfoot, A.L. et al. (1989), *Science*, vol. 246, p. 1459-1466
- Capone, L.A., Whitten, R.C., Prasad, S.S., Dubach, J. 1977, *Astrophys. J.* 215, 977–983
- Chandler, M.O., Waite, J.H. 1986, *Geophys. Res. Lett.* 13, 6–9
- Christophe B. et al. 2012, *Experimental Astronomy* 34, 203-242
- Coradini, A., Magni, G., Turrini, D. 2010, *Space Science Reviews* 153, 411-429.
- Coradini, A., Turrini, D., Federico, C., Magni, G. 2011, *Space Science Reviews* 163, 25-40.
- Dalton, J. B. 2010, *Space Science Reviews* 153, 219-247.
- De Pater, I., Lissauer J.J. 2010. *Planetary Science - Second Edition*. Cambridge University Press, Cambridge, UK
- Decker, R.B., Cheng, A.F. 1994, *JGR*, 99, E9, 19027–19045
- Feuchtgruber, H. et al. 2013, *Astronomy & Astrophysics*, Volume 551, id.A126, 9 pp.,
- Fichtner, H. et al., 2000. *AIP Conf. Proc.* 528, 345
- Funsten, H.O. et al., 2009. *Science* 326, 964
- Gaulme P., Mosser B., 2005, *Icarus*, 178, 84
- Gaulme P., Schmider F.X., Gay J., Guillot T., Jacob C., 2011, *A&A*, 531, A104
- Goldreich, P., & Keeley, D. A. 1977, *ApJ*, 212, 243
- Gomes R., Levison H.F., Tsiganis K., Morbidelli A., 2005, *Nature*, 435, 466-469.
- Gruntman, M.A. et al., 2001. *J. Geophys. Res.* 106, 15767
- Gudkova, T.V. et al. 1988, *Astronomicheskii Vestnik*, vol. 22, p. 23-40
- Guillot, T., 2005, *Ann. Rev. Earth Planet. Sci.* 33, 493–530
- Hartmann, W. K., Ryder, G., Dones, L., Grinspoon, D. 2000, in *Origin of the Earth and Moon* 493-512.
- Helled, Ravit; Anderson, John D.; Podolak, Morris; Schubert, Gerald, (2011). *The Astrophysical Journal*, Volume 726, Issue 1, article id. 15, 7 pp.
- Herbert, F., *J. Geophys. Res.* 114, A11206 (2009)
- Hsieh, K.C., Shih, K.L., Jokipii, J.R., Grzedzielski, S., 1992. *Astrophys. J.* 393, 756
- Iafolla, V. et al., *Plan. Space Sci.* 58, 300-308 (2010)
- Jewitt, D., Haghighipour, N. 2007, *Ann. Rev. of Astron. and Astrophys.*, 45, 261-295.
- Krimigis, S. M., et al., *Science*, 246, 1483, 1989
- Krimigis et al., 2009. *Science*, 326, 971-973
- Levison, H. F., Bottke, W. F., Gounelle, M., Morbidelli, A., Nesvorniy, D., Tsiganis, K. 2009, *Nature* 460, 364-366.
- Levison, H. F., Morbidelli, A., Tsiganis, K., Nesvorniy, D., Gomes, R. 2011, *AJ* 142, 152.
- Lissauer, J.J., and Stevenson, D.J., 2007, in *Protostars and Planets V*, Reipurth, Jewitt, and Keil, (Eds.). Univ. of Arizona Press.
- Mauk, B. H., E. P. Keath, M. Kane, S. M. Krimigis, A. F. Cheng, M. H. Acufia, T. P. Armstrong, and N. F. Ness, *J. Geophys. Res.*, 96, 19,061, 1991

- Mauk, B. H., S. M. Krimigis, A. F. Cheng, and R. S. Selesnick, in *Neptune and Triton*, edited by D. Cruikshank, University of Arizona Press, Tucson, in press, 1994
- Majeed, T., Waite, J.H., Bougher, S.W., Yelle, R.V., Gladstone, G.R., McConnell, J.C., Bhardwaj, A., *Adv. Space Res.* 33(2), 197–211 (2004)
- Martin, S. C. et al., *Astrophysics and Space Science*, 337 (1), 2011
- Marzari F., Weidenschilling S. J. 2002, *Icarus* 156, 570-579
- Matter, A., Guillot, T., Morbidelli, A. 2009. *Planetary and Space Science* 57, 816-821
- McComas, D.J. et al., 2009. *Space Sci. Rev.* 146, 11
- Michel E., Baglin A., Auvergne M., et al., Oct. 2008, *Science*, 322, 558
- Morbidelli A., Levison H. F., Tsiganis K., Gomes R., 2005, *Nature* 435, 462-465.
- Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F., Gomes, R. 2007, *AJ* 134, 1790-1798.
- Mosqueira, I., Estrada, P., Turrini, D. 2010. *Space Science Reviews* 153, 431-446.
- Nesvorn{\y}, D. 2011. Young Solar System's Fifth Giant Planet?. *ApJ* 742, L22.
- Nesvorny, D., Vokrouhlicky, D., Morbidelli, A. 2007. *AJ* 133, 1962-1976.
- Owen, T., Mahaffy, P., Niemann, H. B., Atreya, S., Donahue, T., Bar-Nun, A., de Pater, I. 1999, *Nature* 402, 269-270.
- Papaloizou, J. C. B., Nelson, R. P., Kley, W., Masset, F. S., Artymowicz, P. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil, Univ. Arizona Press, Tucson, AZ, 655
- Pearl, J.C. et al., *Icarus*, vol. 84, 1990, p. 12-28
- Pearl, J.C. et al., *Journal of Geophysical Research Supplement*, vol. 96, Oct. 30, 1991, p. 18,921-18,930
- Podolak, M. et al., *Planetary and Space Science* v. 43, p. 1517-1522, 1995
- Prockter, L. M., et al. 2010. *Space Science Reviews* 153, 63-111.
- Richardson, J. D., J. W. Belcher, M. Zhang, and R. L. McNutt, *J. Geophys. Res.*, 96, 18,993, 1991
- Safronov, V. S. (1969). *Evolution of the protoplanetary cloud and formation of the earth and planets*. Translated from Russian in 1972. Keter Publishing House, 212 pp.
- Schmider F.X., Gay J., Gaulme P., et al., Nov. 2007, *A&A*, 474, 1073
- Schubert, G., Hussmann, H., Lainey, V., Matson, D. L., McKinnon, W. B., Sohl, F., Sotin, C., Tobie, G., Turrini, D., van Hoolst, T. 2010. *Space Science Reviews* 153, 447-484.
- Selesnick, R.S., Richardson, J.D., *Geophys. Res. Lett.* 13, 624–627 (1986)
- Shuleen Chau et al., 2012, *Astrophys. Space Sci.* 337, 65-78
- Sittler Jr., E.C., Ogilvie, K.W., Selesnick, R., *J. Geophys. Res.* 92, 15263 (1987)
- Sromovsky, L.A. et al., *Icarus*, Volume 220, Issue 2, p. 694-712, 2012
- Suomi, V.E. et al., *Science*, vol. 251, Feb. 22, 1991, p. 929-932
- Trafton, L.M., Miller, S., Geballe, T.R., Tennyson, J., Ballester, G.E. 1999, *ApJ* 524, 1059–1083
- Tsiganis K., Gomes R., Morbidelli A., Levison H. F., 2005, *Nature*, 435,459-461.
- Turrini, D., Magni, G., Coradini, A. 2011, *MNRAS* 413, 2439-2466.
- Turrini, D., Coradini, A., Magni, G. 2012, *ApJ* 750, 8.
- Turyshev, S.G., *Annu. Rev. Nucl. Part. Sci.* 58, 207-248 (2008)
- Vasyliunas, V.M., *Geophys. Res. Lett.* 13, 621–623 (1986)
- Walsh K. J., Morbidelli A., Raymond S. N., O'Brien D. P., Mandell A. M., 2012, *M&PS*, 47, 1941
- Walsh K. J., Morbidelli A., Raymond S. N., O'Brien D. P., Mandell A. M., 2011, *Nature*, 475, 206
- Weidenschilling, S. J. 1975, *Icarus* 26, 361-366
- Weidenschilling, S. J., Davis, D. R., Marzari, F. 2001. *Earth, Planets, and Space* 53, 1093-1097
- Weidenschilling S. J., Marzari F., 1996, *Nature*, 384, 619-621

THE HOT AND ENERGETIC UNIVERSE

A White Paper presenting the
science theme motivating
the Athena+ mission



Professor Kirpal Nandra

Max Planck Institute for Extraterrestrial Physics
Giessenbachstrasse, 847541 Garching, Germany
E-mail: knandra@mpe.mpg.de
Tel: +49 89 30000-3401

© Volker Springel/MPA-Chandra/CXC-NASA

AUTHORS AND CONTRIBUTORS

The Athena+ Co-ordination Group: Xavier Barcons (ES), Didier Barret (FR), Andy Fabian (UK), Jan-Willem den Herder (NL), Kirpal Nandra (DE), Luigi Piro (IT), Mike Watson (UK)

The Athena+ Working Groups¹: Christophe Adami (FR), **James Aird (UK)**, Jose Manuel Afonso (PT), Dave Alexander (UK), Costanza Argiroffi (IT), Lorenzo Amati (IT), Monique Arnaud (FR), Jean-Luc Atteia (FR), Marc Audard (CH), Carles Badenes (US), Jean Ballet (FR), Lucia Ballo (IT), Aya Bamba (JP), Anil Bhardwaj (IN), Elia Stefano Battistelli (IT), Werner Becker (DE), Michaël De Becker (BE), Ehud Behar (IL), Stefano Bianchi (IT), Veronica Biffi (IT), Laura Bîrzan (NL), Fabrizio Bocchino (IT), Slavko Bogdanov (US), Laurence Boirin (FR), Thomas Boller (DE), Stefano Borgani (IT), Katharina Borm (DE), Nicolas Bouché (FR), Hervé Bourdin (IT), Richard Bower (UK), Valentina Braitto (IT), Enzo Branchini (IT), **Graziella Branduardi-Raymont (UK)**, Joel Bregman (US), Laura Brenneman (US), Murray Brightman (DE), Marcus Brüggen (DE), Johannes Buchner (DE), Esra Bulbul (US), Marcella Brusa (IT), Michal Bursa (CZ), Alessandro Caccianiga (IT), Ed Cackett (US), Sergio Campana (IT), Nico Cappelluti (IT), **Massimo Cappi (IT)**, **Francisco Carrera (ES)**, Maite Ceballos (ES), Finn Christensen (DK), You-Hua Chu (US), Eugene Churazov (DE), Nicolas Clerc (DE), Stephane Corbel (FR), Amalia Corral (GR), **Andrea Comastri (IT)**, **Elisa Costantini (NL)**, **Judith Croston (UK)**, Mauro Dadina (IT), Antonino D’Ai (IT), **Anne Decourchelle (FR)**, Roberto Della Ceca (IT), Konrad Dennerl (DE), Klaus Dolag (DE), **Chris Done (UK)**, **Michal Dovciak (CZ)**, Jeremy Drake (US), Dominique Eckert (CH), Alastair Edge (UK), **Stefano Etori (IT)**, Yuichiro Ezoe (JP), Eric Feigelson (US), Rob Fender (UK), Chiara Feruglio (FR), **Alexis Finoguenov (FI)**, Fabrizio Fiore (IT), Massimiliano Galeazzi (IT), Sarah Gallagher (CA), Poshak Gandhi (UK), Massimo Gaspari (IT), Fabio Gastaldello (IT), **Antonis Georgakakis (DE)**, Ioannis Georgantopoulos (GR), Marat Gilfanov (DE), Myriam Gitti (IT), Randy Gladstone (US), Rene Goosmann (FR), Eric Gosset (BE), Nicolas Grosso (FR), Manuel Guedel (AT), Martin Guerrero (ES), Frank Haberl (DE), Martin Hardcastle (UK), Sebastian Heinz (US), Almudena Alonso Herrero (ES), Anthony Hervé (FR), Mats Holmstrom (SE), Kazushi Iwasawa (ES), **Peter Jonker (NL)**, **Jelle Kaastra (NL)**, Erin Kara (UK), Vladimir Karas (CZ), Joel Kastner (US), Andrew King (UK), Daria Kosenko (FR), Dimita Koutroumpa (FR), Ralph Kraft (US), Ingo Kreykenbohm (D), Rosine Lallement (FR), Giorgio Lanzuisi (GR), J. Lee (US), Marianne Lemoine-Goumard (FR), Andrew Lobban (UK), Giuseppe Lodato (IT), Lorenzo Lovisari (DE), Simone Lotti (IT), Ian McCherthy (UK), Brian McNamara (CA), Antonio Maggio (IT), Roberto Maiolino (UK), Barbara De Marco (DE), Domitilla de Martino (IT), Silvia Mateos (ES), **Giorgio Matt (IT)**, Ben Maughan (UK), Pasquale Mazzotta (IT), Mariano Mendez (NL), Andrea Merloni (DE), Giuseppina Micela (IT), Marco Miceli (IT), Robert Mignani (IT), Jon Miller (US), Giovanni Miniutti (ES), Silvano Molendi (IT), Rodolfo Montez (ES), Alberto Moretti (IT), **Christian Motch (FR)**, Yaël Nazé (BE), Jukka Nevalainen (FI), Fabrizio Nicastro (IT), Paul Nulsen (US), Takaya Ohashi (JP), **Paul O’Brien (UK)**, Julian Osborne (UK), Lida Oskina (DE), Florian Pacaud (DE), Frederik Paerels (US), Mat Page (UK), Iossif Papadakis (GR), Giovanni Pareschi (IT), Robert Petre (US), Pierre-Olivier Petrucci (FR), Enrico Piconcelli (IT), Ignazio Pillitteri (IT), C. Pinto (UK), Jelle de Plaa (NL), **Etienne Pointecouteau (FR)**, Trevor Ponman (UK), Gabriele Ponti (DE), Delphine Porquet (FR), Ken Pounds (UK), **Gabriel Pratt (FR)**, Peter Predehl (DE), Daniel Proga (US), Dimitrios Psaltis (US), David Rafferty (NL), Miriam Ramos-Ceja (DE), Piero Ranalli (IT), Elena Rasia (US), Arne Rau (DE), **Gregor Rauw (BE)**, Nanda Rea (IT), Andy Read (UK), James Reeves (UK), **Thomas Reiprich (DE)**, Matthieu Renaud (FR), Chris Reynolds (US), Guido Risaliti (IT), Jerome Rodriguez (FR), Paola Rodriguez Hidalgo (CA), Mauro Roncarelli (IT), David Rosario (DE), Mariachiara Rossetti (IT), Agata Roszanska (PL), Emmanouil Rovilos (UK), Ruben Salvaterra (IT), Mara Salvato (DE), Tiziana Di Salvo (IT), **Jeremy Sanders (DE)**, Jorge Sanz-Forcada (ES), Kevin Schawinski (CH), Joop Schaye (NL), Axel Schwöpe (D), **Salvatore Sciortino (IT)**, Paola Severgnini (IT), Francesco Shankar (FR), Debora Sijacki (UK), Stuart Sim (IE), Christian Schmid (DE), Randall Smith (US), Andrew Steiner (US), Beate Stelzer (IT), Gordon Stewart (UK), Tod Strohmayer (US), Lothar Strüder (DE), Ming Sun (US), Yoh Takei (JP), Andreas Tiengo (IT), Francesco Tombesi (US), Ginevra Trinchieri (IT), Asif ud-Doula (US), Eugenio Ursino (NL), Lynne Valencic (US), Eros Vanzella (IT), Simon Vaughan (UK), Cristian Vignali (IT), Jacco Vink (NL), Fabio Vito (IT), Marta Volonteri (FR), Daniel Wang (US), Natalie Webb (FR), Richard Willingale (UK), **Joern Wilms (DE)**, Michael Wise (NL), Diana Worrall (UK), Andrew Young (UK), Luca Zampieri (IT), Jean In’t Zand (NL), Silvia Zane (UK), Andreas Zezas (GR), Yuying Zhang (DE), Irina Zhuravleva (US).

The Japanese X-ray astronomy community, as represented by the High Energy Astrophysics Association in Japan (HEAPA) is pleased to indicate its endorsement and support for the selection of the "Hot and Energetic Universe" as the science theme for an ESA large mission in 2028, to be implemented via an X-ray observatory.

¹ Bold face denotes Working Group chairs. For the full list of the 1000+ Athena+ supporters see: <http://www.the-athena-x-ray-observatory.eu>

1. EXECUTIVE SUMMARY

This White Paper advocates the need for a transformational leap in our understanding of two key questions in astrophysics: 1) How does ordinary matter assemble into the large scale structures that we see today? and 2) How do black holes grow and shape the Universe?

To understand the first of these questions, we must determine the physical evolution of clusters and groups of galaxies from their formation epoch at $z \sim 2-3$ to the present day. These structures grow over cosmic time by accretion of gas from the intergalactic medium, with the endpoint of their evolution being today's massive clusters of galaxies, the largest bound structures in the Universe. Hot gas in clusters, groups and the intergalactic medium dominates the baryonic content of the local Universe, so understanding how this component forms and evolves is a crucial goal.

While the framework for the growth of structure is set by the large scale dark matter distribution, processes of an astrophysical origin also have a major effect. To understand them, it is necessary to measure the velocities, thermodynamics and chemical composition of the gas to quantify the importance of non-gravitational heating and turbulence in the structure assembly process. The temperature of the hot gas is such that it emits copiously in the X-ray band, but current and planned facilities do not provide sufficient collecting area and spectral resolution to settle the issue of how ordinary matter forms the large scale structures that we see today. The key breakthrough is to enable spectroscopic observations of clusters beyond the local Universe, out to $z=1$ and beyond, and spatially resolved spectroscopy to map the physical parameters of bound baryonic structures. Technological advances in X-ray optics and instrumentation can deliver simultaneously a factor 10 increase in both telescope throughput and spatial resolving power for high resolution spectroscopy, allowing the necessary physical diagnostics to be determined at cosmologically relevant distances for the first time.

One of the critical processes shaping hot baryon evolution is energy input – commonly known as feedback – from supermassive black holes. Remarkably, processes originating at the scale of the black hole event horizon seem able to influence structures on scales 10 orders of magnitude larger. This feedback is an essential ingredient of galaxy evolution models, but it is not well understood. X-ray observations are again the key to further progress, revealing the mechanisms which launch winds close to black holes and determining the coupling of the energy and matter flows on larger galactic and galaxy cluster scales.

The widespread importance of black hole feedback means that we cannot have a complete understanding of galaxies without tracking the growth of their central supermassive black holes through cosmic time. A key goal is to push the frontiers of black hole evolution to the redshifts where the first galaxies are forming, at $z=6-10$. X-ray emission is the most reliable and complete way of revealing accreting black holes in galaxies, but survey capabilities need to be improved by a factor ~ 100 over current facilities to reach these early epochs and perform a census of black hole growth. This requires a combination of high sensitivity, which in turn depends on large throughput and good angular resolution, and wide field of view. Again, the required technologies to provide this leap in wide field X-ray spectral imaging are now within our grasp. The same high throughput needed to detect these early black holes will also yield the first X-ray spectra of accreting black holes at the peak of galaxy growth at $z=1-4$, measurements which are impossible with current instrumentation. These spectra will show, for example, if the heavily obscured phase of black hole evolution is associated with the termination of star formation in galaxies via feedback.

These topics comprise the major elements of the science theme **The Hot and Energetic Universe** (Figure 1). The Advanced Telescope for High-energy Astrophysics (hereafter *Athena+*)² mission provides the necessary angular resolution, spectral resolution, throughput, detection sensitivity, and survey grasp needed to revolutionize our understanding of these issues. These capabilities will also provide a powerful observatory to be used in all areas of astrophysics. The technologies for the mission are mature, being based on much previous heritage and major technology developments. A lightweight X-ray telescope based on ESA's Silicon Pore Optics (SPO) technology provides large effective area with excellent angular resolution, combining with state-of-the-art instrumentation for spatially resolved high resolution X-ray spectroscopy (provided by the X-ray Integral Field Unit, X-IFU) and wide field X-ray imaging (provided by the Wide Field Imager, WFI). *Athena+* will open up a vast discovery space leading to completely new areas of scientific investigation, continuing the legacy of discovery that has characterized X-ray astronomy since its inception. The implementation of *Athena+* for launch in 2028 will guarantee a transformation in our understanding of **The Hot and Energetic Universe**, and establish European leadership in high energy astrophysics for the foreseeable future.

² *Athena+* as the successor of Athena, considered for L1, but with enhanced capabilities in terms of angular resolution, effective area, and instrument fields of view.

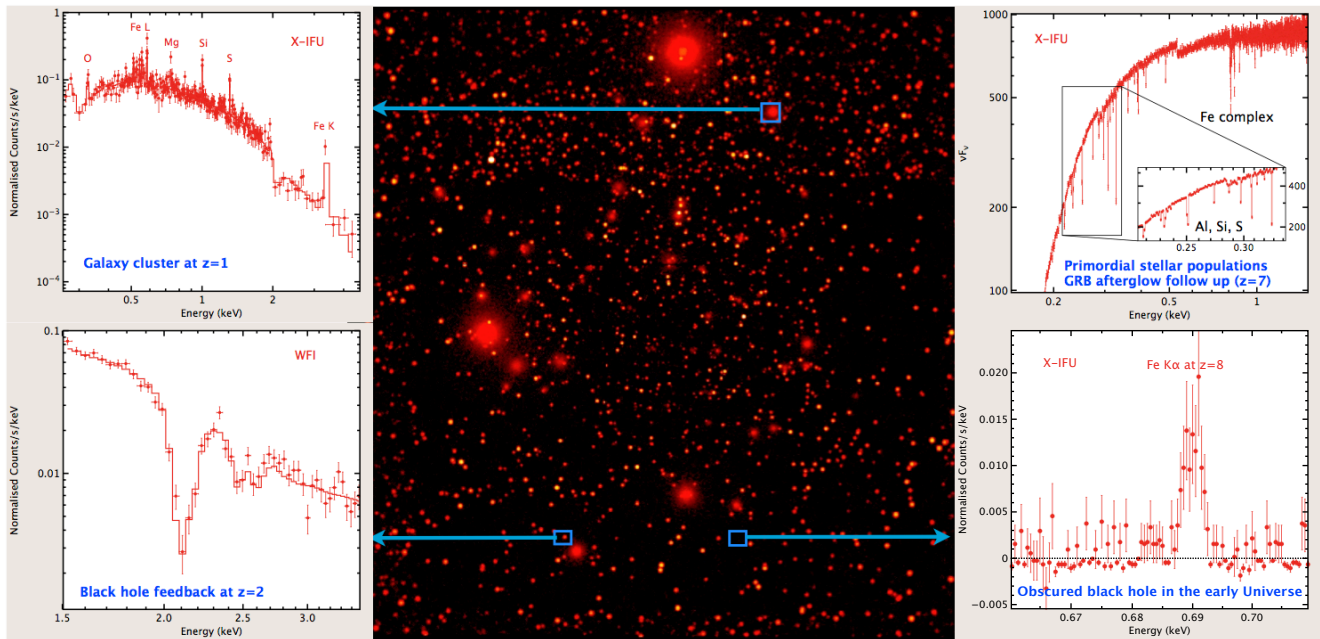


Figure 1: *Athena+* will provide revolutionary advances in our knowledge of the Hot and Energetic Universe. The central panel is a simulated deep WFI observation, while the four surrounding spectra illustrate advances in different science areas, none of which are possible with current facilities.

2. SCIENCE THEME: THE HOT AND ENERGETIC UNIVERSE

2.1. The Hot Universe: How does ordinary matter assemble into the large scale structures that we see today?

In the past decade or more, major observational and theoretical resources have been focused on the understanding of the formation and evolution of galaxies, and this will continue in the future as major facilities are implemented. A significant fraction of galaxies are trapped in larger scale structures – groups, clusters– whose baryonic content is nonetheless completely dominated by hot gas, with stars accounting for less than 15%. In fact, hot gas may dominate the total baryonic content of the local Universe. Groups and clusters of galaxies, as the next step up in the hierarchy of the Universe from galaxies themselves, are fundamental components of the Universe. While the backbone of the large scale structure of the Universe is determined by its cosmological parameters and by the gravitational interaction of the dominant dark matter, the assembly and evolution of baryonic structures are strongly affected by processes of astrophysical origin, which are often poorly known. A complete understanding of **The Hot Universe** - the baryonic gas that traces the most massive structures and drives the formation of galaxies within them - is a fundamental requirement of theories of structure formation and can only be achieved via X-ray observations. **Major astrophysical questions include:**

- How do baryons in groups and clusters accrete and dynamically evolve in the dark matter haloes?
- What drives the chemical and thermodynamic evolution of the Universe's largest structures?
- What is the interplay of galaxy and supermassive black hole evolution in groups and clusters?
- Where are the missing baryons at low redshift and what is their physical state?

2.1.1. The formation and evolution of groups and clusters of galaxies

In the Λ CDM cosmology, the first dark matter haloes are seeded by density fluctuations in the early Universe. These haloes accrete primordial gas and grow over cosmic time via hierarchical gravitational collapse. This process heats the gas to X-ray temperatures. Lying at the nodes of the cosmic web in today's highly structured Universe, galaxy clusters are the final product of this process. Over 80% of their total mass is in the form of dark matter. The remainder is composed of baryons trapped in the dark matter potential well, around 85% of which is diffuse, hot, metal-enriched, X-ray emitting plasma of the intra-cluster medium (ICM). The radiation from this gas and from the member galaxies

reveals the interplay between the dark matter, the hot ICM and the cold baryons locked in stars and interstellar medium (e.g., Kravtsov & Borgani 2012).

By 2028, *Euclid* and LSST, and advances in numerical modelling, should have shown how dark matter structures assembled. However, X-ray observations are needed to understand the evolution of the baryons in the dark matter potential. Key observables include the gas temperature, abundance, velocity and ionisation stage, all of which are provided uniquely via observations of the X-ray continuum and emission lines. These observables reveal how gravitational energy shapes cluster assembly, showing how it is converted into thermal and non-thermal components in the ICM, and generates turbulence and kpc-scale bulk motions (e.g., Vazza et al. 2011). These processes have yet to be observed conclusively at the relevant spatial scales. Moreover, the contribution of the non-thermal component of the energy budget over time, and its influence on the formation and intrinsic properties of galaxy groups and clusters, cannot be determined with current and planned instruments. The spectral capabilities of the *Athena+* X-IFU will allow measurement of the projected ICM velocity field to the required precision through determination of spectral line broadening due to turbulence (see Figure 2) and measurements of line shift due to bulk motion. Fully constraining these basic astrophysical processes will determine the large scale properties of the ICM for nearby clusters, the first step to reveal how the baryons evolve in dark matter structures.

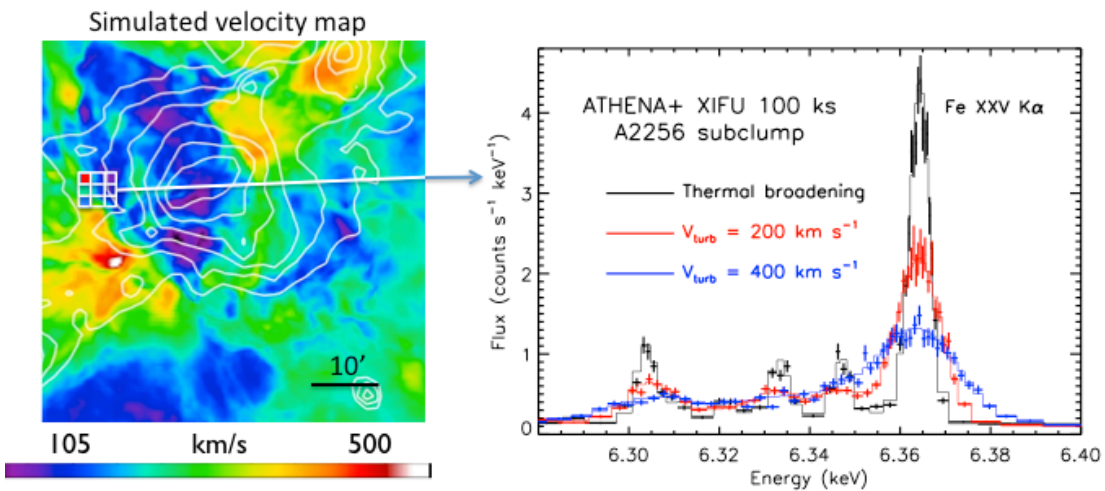


Figure 2: *Athena+* X-IFU spectrum of a subclump in the galaxy cluster A2256, demonstrating the high precision measurements possible for the ICM velocity field. **Left:** Velocity map of a cosmological hydrodynamical simulation of a perturbed galaxy cluster of about $M_{200} \sim 10^{15} M_{\odot}$ with X-ray surface brightness contours overlaid. **Right:** Simulated spectrum for a 100 ks observation with *Athena+* X-IFU for a 1.5 arcmin. region (one of the 9 small regions shown on the image), showing the turbulent broadening of the Fe XXV K α line. Simulated data with $v=200 \text{ km s}^{-1}$ are shown in red. Black and blue represent the model with $v=0$ and $v=400 \text{ km s}^{-1}$, respectively. For an input velocity of 0, 200, 400 km s^{-1} , the 1σ statistical uncertainty is $+20, \pm 5, \pm 10 \text{ km/s}$, respectively.

The interplay of the cluster member galaxies with the ICM can further modify its properties beyond the simple expectations from pure gravitational collapse. For instance, supernova winds can add energy and eject enriched gas into the cluster atmosphere. It is also becoming clear that supermassive black holes inject enough energy to affect the ICM on Mpc scales. Cooling of the denser parts of the ICM feeds the central black hole, resulting in a feedback cycle. The history of this feedback over cosmic time is unknown, but can be quantified via the measurement of the entropy distribution and its evolution (e.g., Ettori et al. 2004). Entropy can be directly derived only from measurements of X-ray surface brightness and gas temperature. The *Athena+* X-IFU and WFI will measure the gas entropy profiles for groups and clusters on all mass scales out to $z \sim 2$. This will determine the non-gravitational energy input over the whole volume from the centre to the outskirts back to the epoch when star formation and accretion activity - and hence feedback processes - were most active.

Clusters are still forming today, and the cluster outskirts, lying across the virial radius and occupying about 85% of the cluster volume, are thus expected to be undergoing strong energetic activity as material is accreted into the dark matter potential (e.g., Reiprich et al. 2013). These are the regions where energy is first transferred into the ICM through merging events. This process will be resolved by *Athena+*. WFI and X-IFU observations will allow routine mapping of the outskirts of nearby clusters, providing their emission measure, temperature and metallicity. This will reveal the physical state of this accreting material, leading to a complete picture of the structure assembly process.

Athena+ will probe the evolution of the scaling relations (e.g., Giodini et al., 2013) that link the total mass of a cluster and its observable properties across a wide range of masses up to $z \sim 2$. Competing feedback and energy input models, which are degenerate at low redshift, will be constrained for the first time at high redshift (e.g., Short et al. 2010). Finally, while *eROSITA* will detect the X-ray emission of all massive clusters to $z \sim 1.5$, much deeper X-ray observations are required to reveal low mass galaxy groups out to the formation epoch. Detection of an extended hot ICM is the unmistakable proof that a galaxy group is fully collapsed and possesses a deep gravitational potential well. The sensitivity and large field of view of the *Athena+* WFI will open up a discovery space for baryonic structures in their infancy at $z > 2$ (e.g., Gobat et al., 2011). *Athena+* observations provide not only the means to discover these objects, but also measurements of the physical properties needed to understand how they formed.

2.1.2. The chemical history of hot baryons

Elements like carbon and beyond are generated via stellar processes, and form the basis of everything around us. Much effort has been devoted to understanding how these "metals" are generated and their distribution within galaxies. On

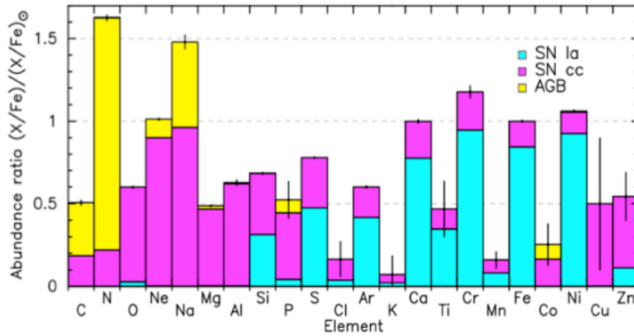


Figure 3: **Abundance measurements for a typical cluster of galaxies (AS 1101, 100 ksec), illustrating the power of high precision *Athena+* X-IFU observations.** The expected abundance ratios relative to solar are shown for SNIa, SNcc, and AGB stars. Abundance measurements will strongly constrain the origin of the metals, the IMF and thus the star formation history.

the other hand, metals can easily be expelled from galaxies, for example by the action of supernovae and AGN winds. Clusters are ideal laboratories to study the production and distribution of heavy metals, having been enriched by member galaxies throughout their lifetime (e.g., de Plaa et al. 2007, Werner et al. 2008). In addition the ICM contains as many baryons as all the stars in the Universe. The metals enter the ICM via the gravitational action of ram-pressure stripping of infalling galaxies, merger-induced gas sloshing and galaxy-galaxy interactions, and feedback from super-winds in starburst galaxies and from AGN (e.g., Schindler & Diaferio, 2008, Böhringer & Werner 2010). How the gas and metals mix depends on transport processes in the magnetized ICM, which are currently

poorly understood. X-ray observations of the lines emitted by the hot ICM are the only way to access information on their abundance and to probe its evolution to high redshift.

With *Athena+* we will observe the epoch when the metals produced in the galaxies are ejected and redistributed in the ICM. The X-IFU will map the metal distribution and explore its relation to the metal ejection and transport processes, catching enrichment processes in the act. Radial profiles of the most abundant elements (O, Ne, Mg, Si, S) out to unprecedented distances will further show how elements are being distributed in the entire cluster volume. Looking at the metal distribution in the cluster outskirts, *Athena+* will determine the role of AGN feedback in shaping the metallicity profiles through expulsion of pre-enriched gas over cosmic time. *Athena+* will also determine the historical origins of ICM enrichment. Each source of metals (e.g., SNIa, SNcc, AGB stars) synthesizes heavy elements in different proportions, so their relative role can be assessed using abundance ratio measurements (e.g., of O/Fe and Si/Fe, Figure 3). *Athena+* will determine the main source of C and N, which can originate from a wide variety of sources including stellar mass loss from intermediate mass stars, and whose cosmic history is poorly known. Abundances of trace elements such as Cr and Mn, widely accessible for the first time, depend on the initial metallicity of the SNIa progenitor system, while N and Na determine the AGB star contribution.

Finally, current instruments have provided hints about cluster Fe abundance evolution from $z=1$ to the present (e.g., Baldi et al. 2012). Observations suggest that about half of the metals found in the ICM were released into the IGM and ICM prior to $z \sim 1$. *Athena+* will measure the evolution of the most abundant elements with redshift, and their ratios, with unprecedented precision, tracing chemical evolution over cosmic time, telling the full story of how, where, and when the ICM was enriched. This is directly related to the metals that have been lost from the cluster member galaxies, so these observations reveal an important aspect of the evolution of galaxies in high-density environments.

2.1.3. Feedback in clusters of galaxies

The mechanical energy carried by jets from central AGN is now believed to control hot-gas cooling in massive ellipticals and groups and clusters of galaxies via a feedback loop in which jets heat the hot gas, suppressing star formation and regulating their own fuel supply. Current X-ray observations have revealed compelling evidence for such AGN feedback (e.g. McNamara & Nulsen 2007). In detail, however, there is a complex interplay between cooling and heating. The physics of how the balance between these processes is established and maintained, and how it evolves with time, is poorly understood. *Athena+* will make the first kinematic measurements on small spatial scales of the hot gas in galaxy, group and cluster haloes as it absorbs the impact of AGN jets. Combined with vastly improved ability to map thermodynamic conditions on angular scales well-matched to the jets, lobes and gas disturbances produced by them, this will provide answers to two key outstanding questions: how energy input from jets is dissipated and distributed throughout the ICM, and how the energy balance between cooling and heating is maintained and established in regions where the most massive galaxies are being formed.

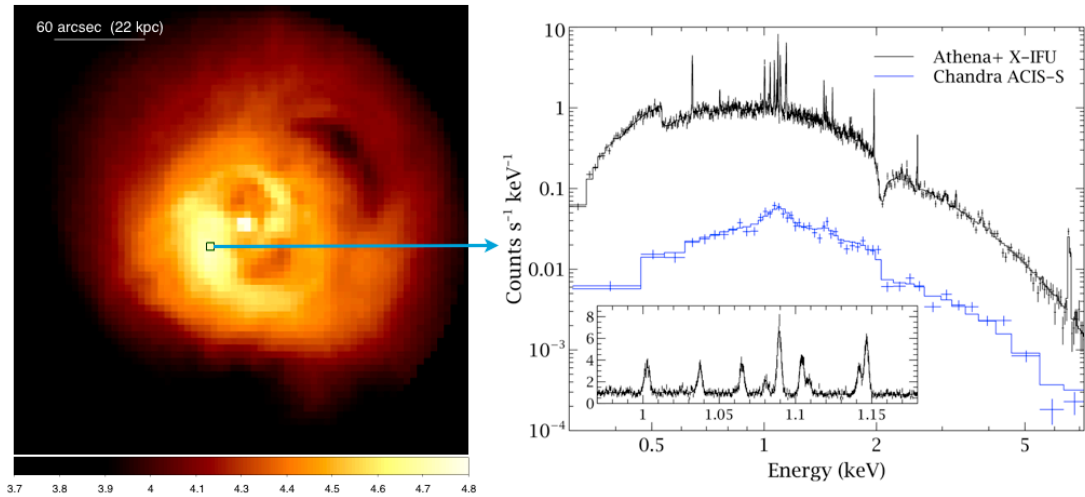


Figure 4: **Simulated *Athena+* observations of the Perseus cluster, highlighting the advanced capabilities for revealing the intricacies of the physical mechanisms at play.** The left panel shows a simulated 50ks X-IFU observation (0.5-7 keV), displayed on a log scale. The spectrum on the right is from the single 5'' \times 5'' region marked by the box, with the existing *Chandra* ACIS spectrum for comparison. The inset shows the region around the iron L complex. With such observations velocity broadening is measured to 10-20 km s⁻¹, the temperature to 1.5% and the metallicity to 3% on scales <10kpc in 20-30 nearby systems, and on <50kpc scales in hundreds of clusters and groups. Such measurements will allow us to pinpoint the locations of jet energy dissipation, determine the total energy stored in bulk motions and weak shocks, and test models of AGN fuelling so as to determine how feedback regulates hot gas cooling.

The X-IFU will map velocity structures and gas conditions on kpc scales in the cores of galaxy groups and clusters where feedback is regulating cooling. The locations of heating and cooling will be pinpointed for the first time and the energy dissipation determined (Figure 4). With the WFI it will be possible to carry out the first population studies of the AGN-induced ripples, disturbances and weak shocks that are assumed to distribute the jet energy isotropically (e.g. Fabian et al. 2003), relating the mechanical energy stored in these disturbances, and its subsequent dissipation, to the environmental and AGN properties across a wide mass range. In addition to establishing the microphysics of AGN heating for the first time, *Athena+* also has the potential to determine how the AGN fuelling process is linked to the thermodynamical properties of the hot gas that absorbs the jet energy input, as is required if a self-regulated feedback process operates to suppress gas cooling and star formation. With an efficiency more than two orders of magnitude higher than the *XMM-Newton* RGS, the X-IFU will determine the gas cooling rates across a wide temperature range on spatial scales matched to the filamentary nebulae of cooler material observed to coincide with the regions of strongest X-ray cooling (e.g. Crawford et al. 1999). X-IFU measurements of the dynamics of the hot gas in the vicinity of the filaments will establish whether their motions are correlated, and distinguish locations where filaments are evaporated by the hot gas, where gas is thermally unstable to cooling and where mixing and possibly charge exchange are occurring, thus determining the relative importance of these processes. The role of AGN-induced turbulence in seeding thermal instabilities will be investigated via population studies, and robust jet power estimates from total mechanical energy input can be compared with accretion rates from hot and cold accretion models for the first time.

Beyond the core region, the energetic impact of radio jets and their role in building up entropy in group and cluster gas is poorly understood. The energy input from strong shocks expected to occur in typical environments is not taken into account in the scaling relations between radio luminosity and jet power (e.g. Birzan et al. 2008), and cannot be reliably determined from radio data. *Athena+* will enable the dynamics and source age and thus jet power to be assessed robustly via direct bulk velocity measurements of expanding hot gas shells around radio lobes extending up to Mpc scales. At higher redshifts, the identification of characteristic features associated with strong shocks in high-resolution WFI temperature maps will measure age, power and energetic impact for large representative samples.

2.1.4. The missing baryons and the Warm-Hot Intergalactic Medium

The intergalactic medium contains 90% of the baryons at the current epoch, and is the visible tracer of the large scale dark matter structure of the local Universe. Theory predicts that the state of most of these baryons evolves from low

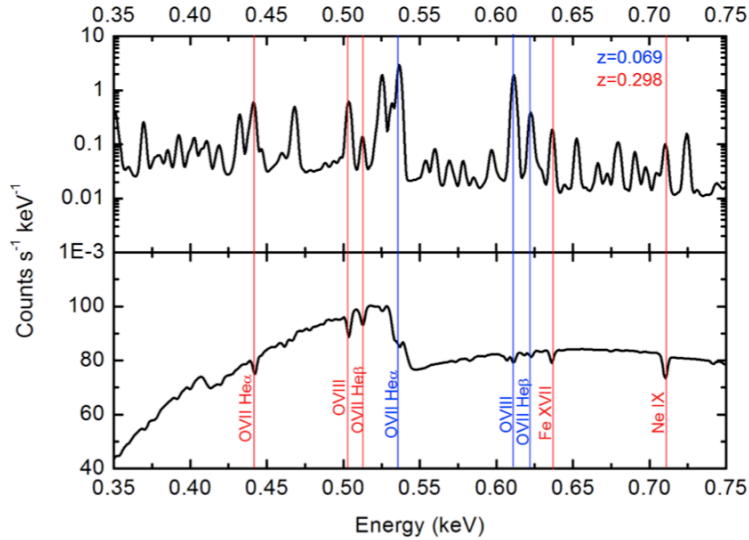


Figure 5: **Simulated emission and absorption line spectra captured in a single *Athena+* observation for two filaments at different redshifts.** Lower panel: absorption spectrum from a sight line where two different filamentary systems are illuminated by a bright background source. Upper Panel: corresponding emission from a $2' \times 2'$ region from the same filaments for a 1 Ms exposure time. The high spectral resolution allows us to distinguish both components. *Athena+* will be able to study ~ 100 of these sight lines in detail.

temperatures, as manifested in the Ly α forest at $z > 2$, to a warm-hot phase (10^5 - 10^7 K) at later times shaped by the filamentary structure of dark matter (Cen & Ostriker 2006). Most of the metals are predicted to reside in the warm-hot phase already at $z \sim 4$. Thermal continuum emission from this gas is extremely hard to detect. The only characteristic radiation from this medium will be in the discrete transitions of highly ionized metals. Evidence for the warm tail of the WHIM, where 10-15% of the missing baryons reside, has been obtained via UV-absorption line studies with FUSE and HST-COS (Shull et al 2012). However, around 50% of the baryons at redshift $z < 2$ and 90% of the metals at redshifts $z < 3$, locked in the hot phase, remain unobserved. In order to reveal the underlying mechanisms driving the distribution of this gas on various scales, as well as different metal circulation and feedback processes the chemical and physical states of about a hundred filaments must be characterized. This can only be done in

X-rays. Present facilities can marginally detect a few filaments (Nicastro et al 2013), but not characterize their physical properties. *Athena+* will probe these baryons in three dimensions, through a combination of absorption and emission studies using the X-IFU. Deep observations of bright AGN combined with Gamma-Ray Burst (GRB) afterglows caught with a 2-4 hour reaction time will be used as backlights for absorption studies through the warm and hot gas. Lines from the high ionization states of O, Ne, Si and Fe, seen simultaneously, enable unique identification of the filamentary structures of the cosmic web (Figure 5), with the detection and characterization of about hundred filaments. At the same time the emission from these structures is mapped by X-ray lines. Combining the two measurements allows the projected size of the structures to be derived while the shapes of the lines and their position reveal the kinematics of the baryons, which, together with the clustering information from the emission lines, pinpoints their origin for the first time.

In Table 1, we summarize the key issues addressed in this section.

Table 1: The Hot Universe: Key issues and key observations.

How does ordinary matter assemble into the large scale structures that we see today?	
Key issue	<i>Athena+</i> key observation
The formation and evolution of groups and clusters of galaxies	
Understand how baryons accrete and evolve in the largest dark matter potential wells of groups and clusters. Determine how and when the energy contained in the hot intra-cluster medium was generated.	Map the structure of the hot gas trapped in galaxy clusters at various redshifts out to the virial radius, resolving gas density and temperature with the WFI. Measure the gas motions and turbulence through X-IFU spatially resolved spectroscopy.
The chemical history of the hot baryons	
Determine when the largest baryon reservoirs in galaxy clusters were chemically enriched. Infer the relative contributions of supernova types, and the initial stellar mass function in protoclusters. Identify the locations in clusters where most of the metals are generated, and determine how they are dispersed.	Measure elemental abundances of heavy elements like O, Ne, Mg, Si, S and Fe, through X-IFU X-ray spectroscopy of groups and clusters at different redshifts. Synthesize the abundances using yields of various SN types and AGB stars. Determine where metals are produced in clusters via spatially resolved spectroscopy of nearby objects.
Cluster feedback	
Understand how jets from active galactic nuclei dissipate their mechanical energy in the intracluster medium, and how this affects the hot gas distribution.	Measure hot gas bulk motions and energy stored in turbulence directly associated with the expanding radio lobes in the innermost parts of nearby clusters with X-IFU. Use sensitive WFI imaging to detect and characterize large scale ripples and weak shocks in nearby groups and clusters.
Determine whether jets from powerful radio-loud AGN are the dominant non-gravitational process affecting the evolution of hot gas in galaxy groups and clusters.	Use WFI to obtain temperature maps of clusters around radio-loud AGN out to intermediate redshifts and map shock structures. Test jet evolution models and infer their impact at the epoch of group and cluster formation.
Establish how AGN feedback regulates gas cooling in groups and clusters and AGN fuelling	Compare jet power estimates by determining total energy budget and dynamical timescales from X-IFU velocity measurements, with accretion rates for competing fuelling models tuned to precisely measured thermodynamical conditions. Determine importance of AGN-induced turbulence in driving thermal instabilities, via mapping of turbulent velocities in a range of systems.
The Warm-Hot Intergalactic Medium	
Find the missing 50% of baryons at $z < 2$ and reveal the underlying mechanisms driving the distribution of this gas on various scales, from galaxies to galaxy clusters, as well as metal circulation and feedback processes.	Determine the distribution of filaments via X-ray absorption spectroscopy against bright distant objects with X-IFU. For the fraction that can be seen in emission, measure their chemical composition, density, size, temperature, ionization and turbulence.

2.2. The Energetic Universe: how do black holes grow and influence the Universe?

All massive galaxies, not just those in clusters and groups, host a supermassive black hole (SMBH) at their centre, the mass of which is tightly correlated with the galaxy bulge properties (e.g. via the $M_{\text{BH}}-\sigma$ relation). This observation has revolutionized our view of the formation and evolution of galaxies, implying a profound influence of black hole accretion throughout the Universe (Kormendy & Ho, 2013). The energy released during the build up of the SMBH exceeds the binding of the entire galaxy by a factor of 10-100, but the relationship requires a self-regulating mechanism connecting the accretion-powered growth of the SMBH at the event horizon level to the star-formation powered growth of the galaxy at much larger scales. Determining the nature and prevalence of this feedback is key to understand the growth and co-evolution of black holes and their host galaxies. This **Energetic Universe** is revealed in a unique manner by observations in the X-ray band, as this is where contamination from the host galaxy is smaller. X-rays provide the clearest and most robust way of performing a census of black hole growth in the Universe, accounting for obscured objects. Despite the progress made currently by *Chandra* and *XMM-Newton*, order of magnitude increases in both survey power and photon collecting area are required to address the most pressing issues in the global black hole growth history, namely its evolution at high redshift ($z=4-10$) and the importance of ultra-obscured, Compton-thick objects. On smaller scales, X-rays produced by gravitational release near the event horizon of black holes are able to diagnose the accretion flow in the strong gravity regime. These observations are needed to reveal not only the secrets of

how matter flows on to the black hole, but how and why the various types of outflow (jets and winds) are launched. The capabilities to probe both the event-horizon scale and the most distant AGN are provided by *Athena+*, offering a unique and comprehensive view of **The Energetic Universe**. With these capabilities, we will address the following questions:

- How do early supermassive black holes form, evolve and affect the distant Universe?
- What is the role of (obscured) black hole growth in the evolution of galaxies?
- How do accretion powered outflows affect larger scales via feedback?
- How do accretion and ejection processes operate in the near environment of black holes?

Athena+ will explore many aspects of black hole accretion and provide definitive evidence of how black holes grow, where in the Universe that growth occurs, and how it affects the wider cosmos.

2.2.1. Formation and Early Growth of Supermassive Black Holes

The processes responsible for the early growth of SMBH are currently unknown. The remnants of the first generation of stars (PopIII stars) may be the seeds of SMBHs and must grow rapidly through frequent periods of intense accretion in early galaxies (Li et al. 2007). Alternatively, massive seeds may form from the monolithic collapse of primordial gas clouds (Begelman et al. 2006), and grow through extended periods of more moderate accretion (Figure 6). How these early AGN shape the evolution of their host galaxy via feedback and the role they may play in the reionisation of the Universe is also unknown. Rapid growth requires considerable fuel for accretion so is likely to imply obscuration. Current and future facilities at (e.g. ALMA, JWST, E-ELT) have a strong focus on finding and studying galaxies, the high redshift Universe, and may sometimes reveal spectroscopically whether these galaxies have a growing black hole. To perform a census, however, requires a selection technique which traces the bolometric luminosity of the system, while minimizing the effects of obscuration and contamination from star formation processes. X-rays provide this, and hence the most powerful AGN selection method at high redshift. We must also cover very large sky areas at very high sensitivity, something that current X-ray instruments (and the above-mentioned longer wavelength facilities), cannot do.

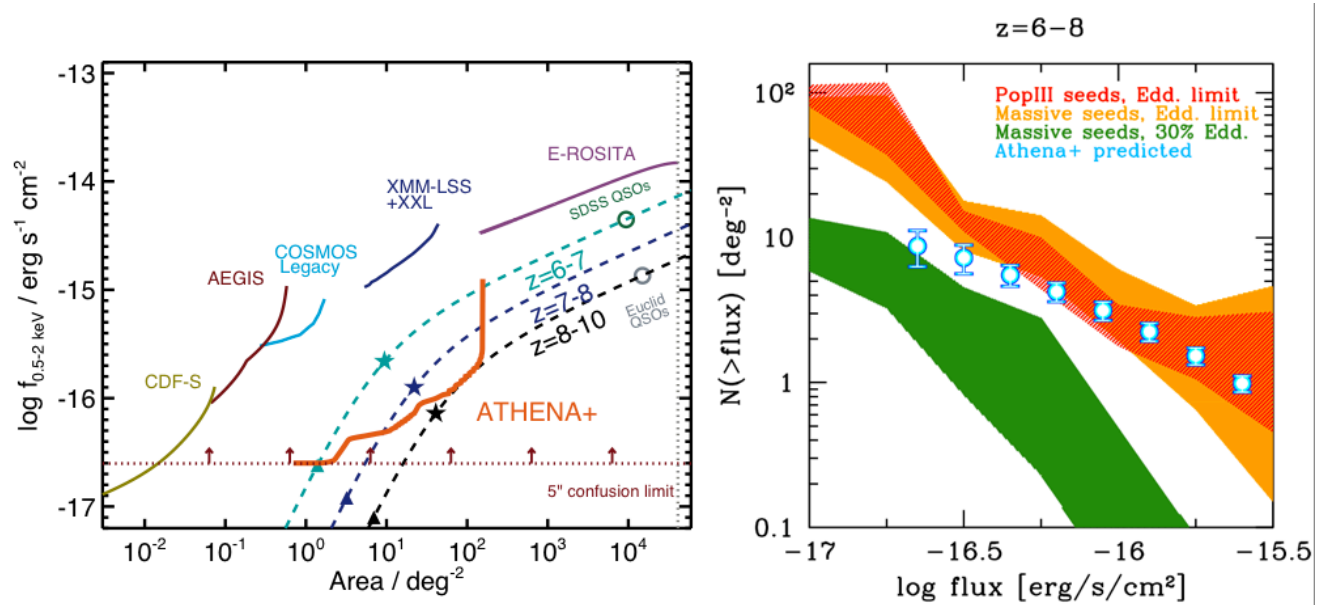


Figure 6: The new discovery space that will be accessed by *Athena+* surveys, which will place essential constraints on physical models for early SMBH growth and formation. *Left*: Area-flux coverage for a multi-tiered survey with the *Athena+* WFI (red line; 4x1Ms, 20x300ks, 75x100ks, 250x30ks), compared with existing and planned X-ray surveys. *Athena+* will break through to the high redshift Universe for the first time, with a survey power a factor ~ 100 better than existing surveys (other solid lines). The dashed lines show the coverage required to detect at least ten sources in the given redshift ranges. *Athena+* will probe population $\sim 2-3$ orders of magnitude fainter than the SDSS or Euclid QSOs (open circles), reaching the population of objects which dominate the accretion power. *Right*: Expected number counts of $z=6-8$ AGN from the survey (circles). Note that at present no purely X-ray selected objects have been found in this redshift range. The shaded regions show predictions based on theoretical models that differ by black hole formation mechanism and growth rate (Volonteri & Begelman, 2010).

Despite large time investments, current X-ray facilities have been able to determine AGN demographics only to around $z \sim 4$ (Aird et al. 2010). With its combination of collecting area at ~ 1 keV, field-of-view and angular resolution, *Athena+* will perform X-ray surveys more than two orders of magnitude faster than *Chandra* or *XMM-Newton*. *Athena+* will access new discovery space, so that models of seed formation and growth mechanisms for early SMBHs in the $z=6-10$ redshift range can be discriminated. Currently there is no X-ray discovered object confirmed above $z=6$, i.e. into the epoch of reionization where the first galaxies and black holes formed. A suitably designed multi-tiered survey with the *Athena+* WFI will yield over 400 $z = 6-8$ X-ray selected AGN and around 30 at $z > 8$ (Figure 6). Crucially, these will include obscured objects, and those whose host galaxy light prevents identification of the AGN in other wavebands. Note that performing an equivalent survey with *Chandra*, the only current X-ray facility sufficiently sensitive to access the necessary flux range, would take approximately 100 years of dedicated survey observations.

Athena+ surveys will also pinpoint active SMBHs (to $\sim 1''$ positional accuracy) within samples of galaxies from the large, state-of-the-art optical and near-infrared imaging surveys (e.g. LSST, Hyper Suprime Cam, Euclid). Further follow-up of these X-ray AGN with ALMA, E-ELT and JWST will yield redshifts, stellar masses, star formation rates, cold gas properties, dust masses, and other important properties of the host galaxies. *Athena+* X-IFU follow-up of samples of $z \sim 6-9$ galaxies may even reveal the intense iron line emission using the superb spectral resolution to improve the sensitivity and push below the confusion limit. This emission characterizes the most obscured AGN, which may signpost a critical phase in the formation of the earliest galaxies.

An entirely complementary way of exploring the seed population of SMBH at the highest redshifts can be achieved via fast follow up observations of X-ray afterglows of Gamma-Ray Bursts. The chemical fingerprint of Pop III star explosions is distinct from that of later generations (Heger & Woosley 2010). Tracing the first generation of stars is also crucial for understanding cosmic re-ionization and the dissemination of the first metals in the Universe. High resolution absorption line X-ray spectroscopy with the *Athena+* X-IFU will enable us to determine the redshift and the typical masses of early stars, with fundamental impact on the models of early star populations and the onset of the accretion power from SMBH in the Universe.

2.2.2. Obscured Accretion and Galaxy Evolution

Following their formation at high redshifts, galaxies and black holes continue to grow, with the peak of both star formation and accretion activity occurring at $z \sim 1-4$. The general picture proposed involves an early phase of intense black hole growth that takes place behind large obscuring columns of inflowing dust and gas clouds. This is then followed by a stage in which some form of AGN feedback controls the fate of the interstellar medium and hence, the

evolution of the galaxy. Open questions that relate to our current understanding of black hole growth and its relation to the build-up of galaxies include: what are the physical conditions (e.g. fuelling mode, triggering mechanism) that initiate major black hole accretion events and how they are connected to star-formation on larger scales; what is the nature of AGN feedback and whether it plays a significant role in the evolution of galaxies in the range $z \sim 1-4$, where most of the stellar and black hole mass was put in place.

Observations at high energies provide unique information about both the early heavily obscured stage and the later blow-out phase. Detection of the intense iron $K\alpha$ emission (Figures 1 and 7), characteristic of the most obscured “Compton thick” phase, will yield clean samples of deeply shrouded AGN with well-defined selection functions and determinations of their accretion

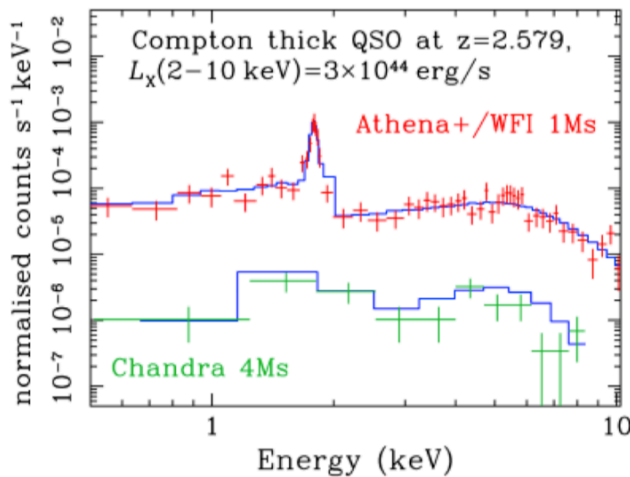


Figure 7: *Athena+* WFI spectrum of a Compton thick AGN at $z=2.6$. The spectrum accumulated by *Chandra* in 4 Ms is shown for comparison.

luminosity and obscuring column. Existing X-ray surveys have yielded only a few tens of the most obscured sources. *Athena+* WFI surveys will yield around 15,000 of the most heavily obscured AGN to $z \sim 4$, and perhaps beyond (Figure 1), determining their demographics for the first time.

X-rays also probe the most highly ionized component of feedback flows from AGN, which can dominate in terms of energy and mass flux, by detecting highly ionised material, via absorption line spectroscopy (Figure 1). This material is invisible at other wavelengths. Current studies of these phenomena are largely limited to the local Universe. The high throughput of *Athena*+ will enable detailed X-ray spectral studies of AGN up to $z \sim 4$, including the prevalence and energetics of these feedback flows, for the first time.

Having identified the key evolutionary stage via X-ray spectroscopy, future observatories that focus primarily on galaxy properties will provide the necessary complementary information (e.g. morphology, star formation) to understand the link between black hole growth and galaxy formation when the Universe was experiencing its most active phase.

The same surveys, designed to yield significant samples of relatively rare populations (e.g. high z or Compton thick AGN), will also yield of order 600,000 mildly and unobscured AGN at all redshifts. This will enable detailed statistical investigations, e.g. of AGN host galaxies, clustering, and the link between black hole accretion and large scale structures as a function of redshift, luminosity or obscuration. The huge improvement in sample sizes will open a new discovery space in supermassive black hole studies, similar to the progress achieved in galaxy evolution work when huge datasets (e.g. SDSS) became available (e.g. the discovery of star-formation main sequence, galaxy colour bimodality). As an example, we note that the baseline *Athena*+ WFI survey strategy will yield approximately 10,000 X-ray AGN at $z=4-6$, compared to about 100 current examples from combined *Chandra* and *XMM-Newton* surveys.

2.2.3. Galaxy-scale Feedback

While AGN feedback is invoked in almost all models of galaxy evolution, the physical mechanisms by which the energy output from the AGN emerges and couples with the surrounding medium at larger scales is not yet established. Jet feedback is well established in clusters of galaxies, where it heats gas in the halo preventing gas cooling onto the central galaxy (Croton et al. 2006, see Fabian 2012 for a recent review). The AGN feedback mechanism that is thought to quench star formation in more typical massive galaxies is likely to be different. In this context, the high velocity winds recently discovered in the X-ray spectra of AGN are potentially a very effective way of transporting energy from the nuclear scale to galaxy (King & Pounds, 2003). These are important, because they probe the phase of the wind/outflow which carries most of the kinetic energy, and is otherwise too highly ionized to be seen at longer wavelengths. According to existing models, the energy of such powerful AGN-driven winds is deposited into the host galaxy ISM, and may contribute to the powering of the recently discovered galactic-scale molecular outflows, which are able to sweep away the galaxy's reservoir of gas and quench the star formation activity (Wagner et al. 2013).

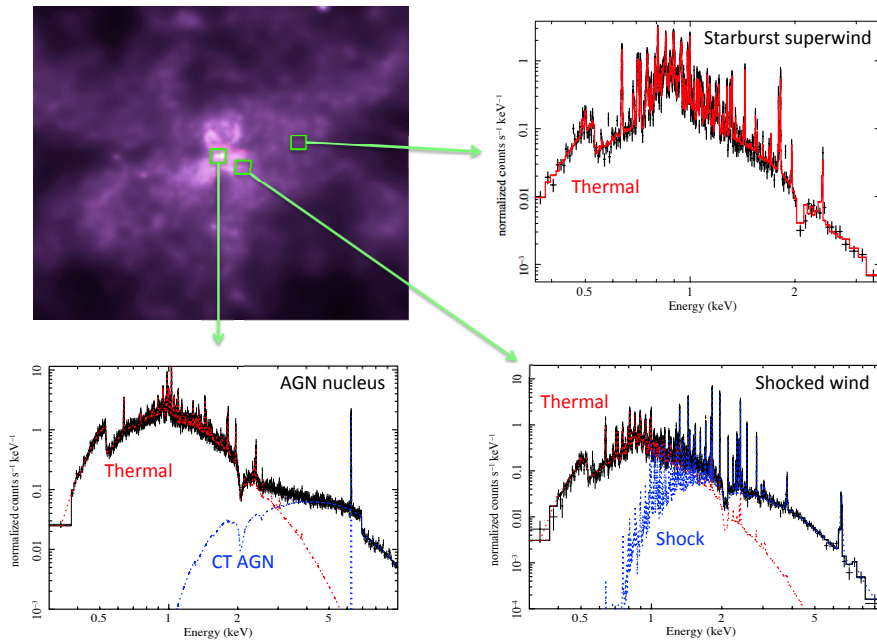


Figure 8: Simulated *Athena*+ X-IFU spectra of different regions in the ULIRG NGC 6240, illustrating the ability to disentangle the complex mix of excitation mechanisms. The observations can clearly distinguish starburst-driven excitation and AGN-heated shocks in the large scale galactic outflows of molecular and cold gas, as well as the buried component originating from the Compton thick AGN double nuclei. The top left panel shows the *Chandra* X-ray image of NGC6240 (Credit: NASA/CXC/SAO/E. Nardini et al.), with the green boxes indicating the representative 5" x 5" regions for which the spectra are simulated.

Athena+ is sufficiently sensitive to detect these winds at cosmological distances. In the local Universe, exquisite detail of the feedback process will be revealed. The absorption lines from AGN winds and outflows and, crucially, their variability will determine the ionization state, density, temperature, abundances, velocities and geometry of the winds down to the inner regions where they are launched (Proga et al. 2000). We expect different correlations between the physical parameters (e.g. density, ionization parameter) as a function of distance, maximum outflow velocity, and X-ray or UV luminosity. These can then be compared to theoretical models of accretion disk winds.

With the spatially-resolved spectroscopic power of the X-IFU, the interactions between hot and ionized feedback winds – whether originating from AGN or star formation – can be resolved and distinguished (Figure 8). The ability to separate spatially and spectrally the two types of activity will distinguish the dominant feedback mode and reveal if and how these phenomena are linked. The *Athena+* X-IFU will clearly distinguish spectrally between starburst superwinds from AGN-driven shocks on scales down to few kpc in 40-50 nearby AGN, ULIRG and starburst galaxies. At the same time it will map the velocity field of the outflowing gas to within 20-30 km s⁻¹. These results will provide a template for AGN at higher redshift where wind shocks can be spectrally, but not spatially, resolved.

Supernovae-driven hot winds are in themselves a crucial element in our understanding of galaxy evolution and baryonic structures. Hydrodynamical simulations show that they are able to expel 90% (or more) of the baryons in star-forming galaxies, transporting metals into the intragroup, intracluster or inter-galactic medium. *Athena+* will observe this process directly via spatially-resolved spectroscopy of nearby galaxies, and in particular star-forming galaxies. Currently, we have crude constraints on the X-ray halo of either very nearby and bright examples (i.e. M82) or very massive galaxies through stacked analysis. *Athena+* will perform high resolution spectroscopy on typical galaxies with high spatial resolution, revealing how much mass and metals are carried in the hot superwind component.

2.2.4. The Physics of Accretion

The black hole feedback processes that shape the evolution of galaxies ultimately originate on scales 10 orders of magnitude smaller, as a result of accretion processes close to the event horizon. To discover this mechanism we must access a very small region, within a few tens of gravitational radii of the compact object. It is here where matter in the accretion disk may release almost half of its rest-mass energy, where winds and jets are launched, and where General and Special Relativity leave their mark on the emitted radiation. To understand accretion we must therefore understand the close environment of the black hole, how strong gravity affects the behaviour of matter and radiation, whether or not the black hole is spinning, and the relationship between accretion and ejection. Important questions still remain, such as whether the disk always extends down to the innermost stable orbit and, if not, whether any disk truncation is related to the launching of winds and jets. Similarly, how the jets themselves are accelerated remains largely a mystery, with wound-up magnetic fields usually invoked. It is also possible that black hole spin is tapped by a variant of the Penrose process (Blandford & Znajek, 1977).

X-ray emission is produced copiously in the black hole environment, via Comptonisation of thermal disk photons by electrons in a hot corona. The resulting continuum illuminates the disk, where it is reprocessed and scattered, producing signatures in the X-ray spectrum such as the iron K α line which encode the signatures of the strong gravity environment (Nandra et al. 2007). Spectroscopy and timing of these reflection features is the key to further understanding of the accretion process. The most promising technique involves mapping of the inner accretion flow via the reverberation expected when intrinsic changes in the luminosity of the corona are seen in the reflection spectrum (Stella 1990). Differing path lengths mean that the reflection is seen to lag behind the coronal continuum. Since different parts of the reflection spectrum come from different radii of the disc, the lag energy spectrum depends on both time and energy and can be used to determine both the mass and spin of the black hole, and to map the central regions. The predicted lags have now been measured by *XMM-Newton* (Fabian et al. 2009) and confirmed by microlensing observations, placing limits on the size of the corona to a few tens of gravitational radii.

Athena+ will take the next step, improving the data quality sufficiently to determine the so-called transfer function of the process (Blandford & McKee 1982), which encodes the underlying geometry of the corona-disk system. For example, it will be possible to distinguish between a compact corona on the black hole spin axis (suggestive of e.g. an aborted jet) or one extended over the disk. *Athena+* will reverberation-map hundreds of AGN and many black hole binary systems (and also neutron star binaries), exploiting the large effective area, angular resolution, spectral resolution, timing capabilities, soft energy response and uninterrupted long exposures provided by the mission and its orbit. This will enable detailed exploration and calibration of a wide range of effects in the brighter AGN, in particular in objects with jets or winds where it may reveal the important acceleration zone. The behaviour of the inner accretion flow where the energy release occurs will be revealed and, from a large sample of both massive and stellar mass black

holes, show how and why the gravitational energy is split between the disk emission, the hot corona, the fast outflow and a jet, the last three presumed to be magnetically powered.

Reverberation signals also allow measurement of the black hole spin. In fact, *Athena+* will offer multiple probes of spin, such as disk continuum fitting and aperiodic variability in black hole binaries. These measurements in bright sources will allow calibration of the most common method, which exploits the shape of the broad iron $K\alpha$ line (Fabian et al. 2000). This will then open up these measurements at cosmologically significant distances ($z=1-2$), with the iron line region redshifted to the 1-3 keV range where *Athena+* has enormous effective area. This could enable a unique probe of the black hole evolution, via the determination of the spin distribution in AGN, which encodes information about whether the black holes grew, mainly from accretion or mergers (Berti & Volonteri 2008). In Table 2, we summarize the key issues for the hot Universe.

Table 2: The Energetic Universe: key issues and key observations.

The Energetic Universe: how do black holes grow and influence the Universe?	
Key issue	<i>Athena+</i> key observation
Formation and early growth of supermassive black holes	
Determine the nature of the seeds of high redshift ($z>6$) SMBH, which processes dominated their early growth, and the influence of accreting SMBH on the formation of galaxies in the early Universe.	Accreting SMBH, even in obscured environments, will be detected out to the highest redshifts through their X-ray emission in multi-tiered WFI X-ray surveys. The most obscured objects will be unveiled by targeted X-IFU spectroscopy revealing strong reflected iron lines.
Trace the first generation of stars to understand cosmic re-ionization, the formation of the first seed black holes, and the dissemination of the first metals.	X-IFU measurements of metal abundance patterns for a variety of ions (e.g., S, Si, Fe) for at least 10 medium-bright gamma-ray burst X-ray afterglows per year with H equivalent column densities as small as 10^{21} cm^{-2} and gas metallicities as low as 1% of solar.
Obscured accretion and galaxy formation	
Find the physical conditions under which SMBH grew at the epoch when most of the accretion and star formation in the Universe occurred ($z\sim 1-4$).	Perform a complete census of AGN out to $z\sim 3$, including those that reside inside a Compton-thick environment. This will be achieved via WFI surveys, where strong iron lines will be the signposts of heavily obscured AGN.
Galaxy-scale feedback	
Understand how accretion disks around black holes launch winds and outflows and determine how much energy these carry.	Use X-IFU to fully characterize ejecta, by measuring ionization state, density, temperature, abundances, velocities and geometry of absorption and emission features produced by the winds and outflows in tens of nearby AGN.
Understand the significance of AGN outflows in determining the build-up of galaxies at the epoch when most stars in present day galaxies formed.	X-IFU observations of nearby AGN/ULIRGs/starbursts will probe the interactions of AGN- and starburst-driven outflows with the ISM, and will provide a local template for understanding AGN feedback at higher redshift.
Understand how the energy and metals are accelerated in galactic winds and outflows and are deposited in the circum-galactic medium. Determine whether the baryons and metals missing in galaxies since $z\sim 3$ reside in such extended hot envelopes.	Use X-IFU to directly map galactic haloes in nearby galaxies to characterize warm and hot gas outflows around starburst, ULIRG and AGN galaxies. Measure gas mass deposited, mechanical energy and chemical abundances to model baryon and metal loss in galaxies across cosmic time.
The physics of accretion	
Determine the relationship between the accretion disk around black holes and its hot electron plasma. Understand the interplay of the disk/corona system with matter ejected in the form of winds and outflows.	Perform time-resolved X-ray spectroscopy of X-ray binaries and AGN out to significant redshifts ($z\sim 1$), so that time lags between different spectral components can be found and the transfer function measured. These measurements will then determine the geometry of the disk/corona system, key to understanding how jets and winds are launched.
Infer whether accretion or mergers drive the growth of SMBH across cosmic time.	Measure black hole spins through reverberation, timing, time-resolved spectroscopy and average spectral methods. Use spectroscopy to perform a survey of SMBH spins out to $z\sim 1-2$ and compare with predictions from merger and accretion models.

2.3. Science enabled by the *Athena+* observatory capabilities

The **Hot and Energetic Universe** includes almost all known astrophysical objects, from the closest planets to the most distant quasars and gamma-ray bursts. The instrument suite required to achieve the science goals described above provides *Athena+* with unprecedented observatory capabilities, enabling new science to be performed for a wide range of objects, of great interest to the whole astronomical community. Here, we provide a non exhaustive list of scientific issues which could be addressed by *Athena+*, together with a short description of the *Athena+* breakthrough observations to be performed (Table 3).

Table 3: Science enabled by the observatory capabilities of *Athena+*

Key issue	<i>Athena+</i> key observation
Planets	
Establish how planetary magnetospheres and exospheres, and comets, respond to the interaction with the solar wind, in a global way that in situ observations cannot offer.	First detailed spectral mapping of Jupiter's X-ray emission, of the Io Plasma Torus, of Mars' exosphere and of X-rays from comets. Fluorescence spectra of Galilean Satellites for surface composition analysis.
Exoplanets	
Extend exoplanet research to X-rays to explore the magnetic interplay between stars and planets.	Measurements of X-ray spectral variability over the activity cycle of the host star and over the planet's orbital period.
Stellar physics	
Assess the mass loss rates of high velocity chemically-enriched material from massive stars to understand the role they play in the feedback processes on Galactic scales.	Time-resolved X-IFU spectroscopy of single and binary massive stars to characterize the large scale structures in their winds and assess their mass-loss rates.
Understand how high energy irradiation of disks during the formation and early evolution of low and intermediate-mass stars affects disk evolution and eventually planetary system formation	Time resolved X-IFU spectroscopy of the brightest objects to explore the accretion process variability and the modulation due to accretion stream shadowing, and constrain with Doppler line shifts the bulk velocity of accreting material.
Supernovae	
Understanding the physics of core collapse and type Ia supernova remnants, quantifying the level of asymmetry in the explosion mechanism, the production of heavy elements, and their impact on the galactic environment.	First detailed 3D mapping of the hot ejected material in the line of sight (velocity, temperature, ionization state and composition) to determine to the full geometry and properties of the different layers of shocked ejecta.
Stellar endpoints	
How does this mass loss from disk winds influence the binary evolution and impact the interstellar medium?	Perform multiple X-IFU observations on time-scales shorter than the wind time variability, measuring outflow changing velocities and ionisation states.
Extending the measurements of mass and radius of neutron stars to isolated millisecond pulsars and faint quiescent neutron star binaries	Waveform fitting of X-ray pulses from isolated millisecond pulsars and modeling of atmospheric emission from globular cluster sources.
Sgr A*	
Understand flare production in Sgr A*, the origin of the quiescent emission, and set constraints on the past AGN activity of Sgr A*.	X-IFU observations along with multi-wavelength coverage to measure the ionization process and physical properties of the plasma during the flaring and quiescent states.
Interstellar dust	
Understand the chemical composition of interstellar dust.	X-IFU observation of extended X-ray absorption features.
Interstellar medium	
Determine the chemical composition of the hot gas of the interstellar medium, as a tracer of stellar activity in our and other galaxies.	X-IFU spectrum of the hottest emission and absorption components of the ionized gas characterized by e.g. OVII, OVIII, NeIX

3. SYNERGIES AND DISCOVERY SPACE

In the late 2020s, pre-eminent facilities operating at other wavelengths will include LOFAR, SKA, ALMA, GAIA, JWST, E-ELT, LSST and CTA. *Athena+* has capabilities unrivalled by any other planned high-energy mission,

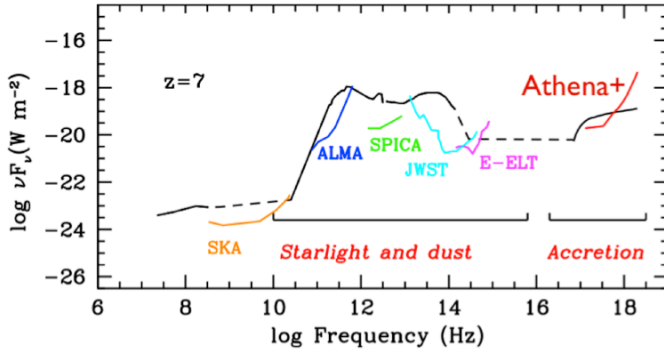


Figure 9: Spectral Energy Distribution (SED) of a high redshift, obscured AGN at $z=7$. The average SED, adapted from Lusso et al. 2011, is shown as the solid line. In terms of bolometric luminosity, this object is representative of what will be detected in *Athena+* medium-deep surveys (e.g. 300 ks exposures, red line). The 3σ sensitivities (for a ~ 40 ks exposure) of SKA, ALMA, SPICA, JWST and E-ELT are also shown, as labeled.

complementing those facilities in providing the essential panchromatic view needed to understand a wide range of astrophysical phenomena. Figure 9 provides one illustration of this complementarity, focusing on the study of the first galaxies in cosmic history, emphasizing the excellent match between the sensitivity offered by *Athena+* and the anticipated capabilities at other wavelengths. This is necessary to understand the complex nature and phenomenology of the vast majority of astrophysical sources, where high energy phenomena are present and X-ray observations therefore provide unique information.

Athena+ will deliver a rich return of serendipitous discoveries by opening up new discovery space, enabled by a major step forward in observational capabilities. Historically, high energy astrophysics missions have been very effective in this respect. Some

examples include the discovery of vast amounts of hot baryonic gas trapped in the potential wells of galaxy clusters and the clear demonstration of the existence of growing super-massive black holes in many galaxy cores previously believed to be normal systems. How *Athena+* will explore its discovery space cannot be stated in advance, but the observatory nature of the mission ensures that it will be driven by cutting-edge science questions. As an example, its unprecedented sensitivity will reveal populations of X-ray sources never seen before, in particular some of the first galaxy groups at $z \sim 2.5-3$, predecessors of today's galaxy clusters.

4. THE ATHENA+ MISSION CONCEPT

Achieving the ambitious goals set out in this White Paper requires an X-ray observatory-class mission delivering a major leap forward in high-energy observational capabilities. Thanks to its revolutionary optics technology and the most advanced X-ray instrumentation, the *Athena+* mission, outlined below, will deliver superior wide field X-ray imaging, timing and imaging spectroscopy capabilities, far beyond those of any existing or approved future facilities. Like *XMM-Newton* today, *Athena+* will play a central role in astrophysical investigations in the next decade. No other observatory-class X-ray facility is programmed for that timeframe, and therefore *Athena+* will provide our only view of **the Hot and Energetic Universe**, leaving a major legacy for the future. The *Athena+* mission has an exceptionally mature heritage based on extensive studies and developments by ESA and the member states for *Athena* (and IXO). Compared with *Athena*, the *Athena+* concept incorporates important enhancements, including a doubling of the effective area (to 2m^2); an improvement in the angular resolution by a factor ~ 2 (to $5''$) and quadrupling of the fields of view of both the WFI and X-IFU, yet representing a realistic evolution in performance for a mission to fly in 2028.

Mapping the dynamics and chemical composition of hot gas in diffuse sources requires high spectral resolution imaging (2.5 eV resolution) with low background; this also optimizes the sensitivity for weak absorption and emission features needed for WHIM studies or for faint point source characterisation. An angular resolution of 5 arcsec is required to disentangle small structures of clusters and groups and, in combination with a large area, provide high resolution spectra, even for faint sources. This angular resolution, combined with the mirror effective area and large field of view (40 arcmin) of the WFI provides the detection sensitivity (limiting flux of $10^{-17} \text{ erg/cm}^2/\text{s}$ 0.5-2 keV band) required to detect AGNs at $z > 6$ within a reasonable survey time. The science requirements and enabling technologies are summarized in Table 4.

Table 4: Key parameters and requirements of the *Athena+* mission. The enabling technology is indicated

Parameter	Requirements	Enabling technology/comments
Effective area	2 m ² @ 1 keV (goal 2.5 m ²) 0.25 m ² @ 6 keV (goal 0.3 m ²)	Silicon Pore Optics developed by ESA. Single telescope: 3 m outer diameter, 12 m fixed focal length
Angular Resolution	5" (goal 3") on-axis 10" at 25' radius	Detailed analysis of error budget confirms that a performance of 5" HEW is feasible.
Energy range	0.3-12 keV	Grazing incidence optics & detectors
Instrument field of view (diameter)	<i>Wide-Field Imager: (WFI):</i> 40' (goal 50')	Large area DEPFET Active Pixel Sensors
	<i>X-ray Integral Field Unit: (X-IFU):</i> 5' (goal 7')	Large array of multiplexed Transition Edge Sensors (TES) with 250 micron pixels
Spectral Resolution	WFI: <150 eV @ 6 keV	Large area DEPFET Active Pixel Sensors
	X-IFU: 2.5 eV @ 6 keV (goal 1.5 eV @ 1 keV)	Inner array (10"x10") optimized for goal resolution at low energy (50 micron pixels).
Count Rate capability	> 1 Crab ³ (WFI)	Central chip for high count rates without pile-up and with micro-second time resolution
	10 mCrab, point source (X-IFU)	Filters and beam diffuser enable higher count rate capability with reduced spectral resolution
	1 Crab (30% throughput)	
TOO response	4 hours (goal 2 hours) for 50% of time	Slew times <2 hours feasible; total response time dependent on ground system issues

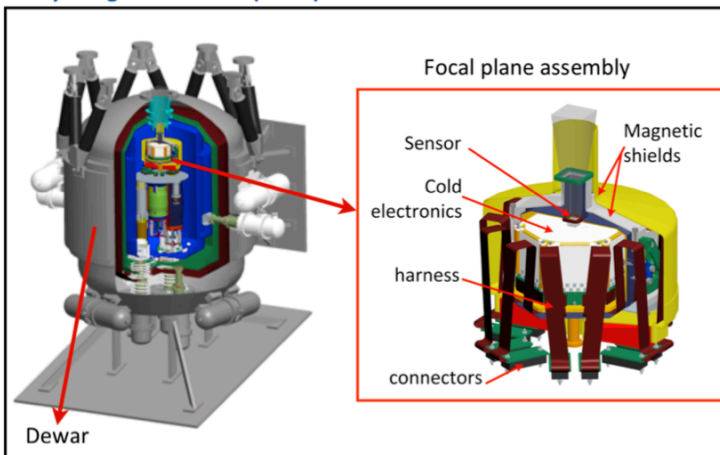
4.1. Science payload

The strawman *Athena+* payload comprises three key elements:

- A single X-ray telescope with a focal length of 12m and an unprecedented effective area (2 m² at 1 keV). The X-ray telescope employs Silicon Pore Optics (SPO), an innovative technology that has been pioneered in Europe over the last decade mostly with ESA support. SPO is a highly modular concept, based on a set of compact individual mirror modules, which has an excellent effective area-to-mass ratio and can achieve high angular resolution (<5").
- The X-ray Integral Field Unit (X-IFU), an advanced actively shielded X-ray microcalorimeter spectrometer for high-resolution imaging, utilizing Transition Edge Sensors cooled to 50 mK.
- The Wide Field Imager (WFI), a Silicon active pixel sensor camera with a large field of view, high count-rate capability and moderate resolution spectroscopic capability.

The two instruments (in Figure 10) can be moved in and out of the focal plane by an interchange mechanism, which is a simplified version of the IXO one. Key characteristics of the instruments are also listed in Table 3.

X-ray Integral Field Unit (X-IFU)



Wide-Field Image (WFI)

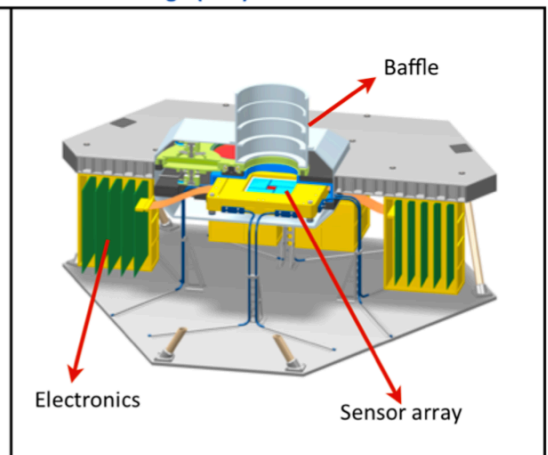


Figure 10: The *Athena+* science instruments. *Left:* Design drawing of the X-IFU showing the Dewar and a zoom on the focal plane assembly. *Right:* Design drawing of the WFI.

³ 1 Crab corresponds to a flux of $2.4 \cdot 10^{-9}$ ergs/s/cm² (2-10 keV).

4.2. Athena+ performance

The *Athena+* telescope delivers a throughput a factor ~ 10 larger than *XMM-Newton* and almost a factor 100 larger than *Astro-H* at low energies (and more than a factor 10 larger at high energies), coupled with major improvements in focal plane instrumentation, including the use of a large format microcalorimeter to provide high-resolution spectroscopic imaging and an advanced Si-sensor to provide wide-field imaging with spectroscopic capability and a combination of high time resolution and count-rate capability. We have selected a few key comparisons to illustrate *Athena+* performance in Figure 11.

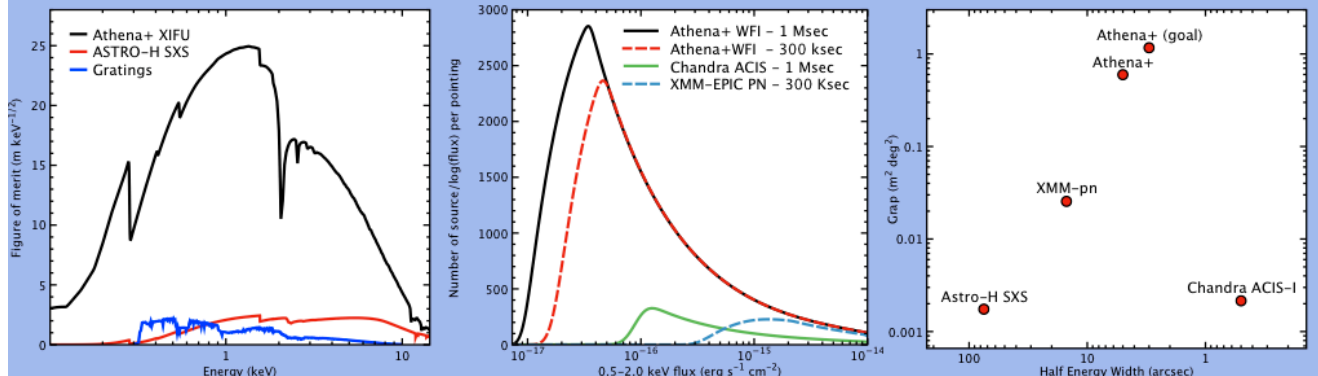


Figure 11: *Athena+* scientific performance. **Left:** Figure of merit for weak spectral line detection of X-ray high-spectral resolution spectrometers, derived from the number of counts per independent spectral bin. The gratings line represents the best of the current *XMM-Newton* or *Chandra* grating at each energy. **Centre:** Number of sources per logarithmic flux interval expected in single *Athena+* WFI pointings at high galactic latitudes compared to *Chandra* and *XMM-Newton*. **Right:** Grasp of operational and planned missions as a function of angular resolution. Grasp is defined as the product of effective area at 1 keV and the instrument field of view.

4.3. Mission profile

Preliminary industrial designs for the *Athena+* spacecraft are shown in Figure 12. Like *Athena*, it is a conventional design retaining much heritage from *XMM-Newton*. Considerations of observing efficiency and thermal stability favour an L2 orbit reached by Ariane V. The initial industrial assessments performed for this white paper indicate very safe margins for the mass and power budgets.

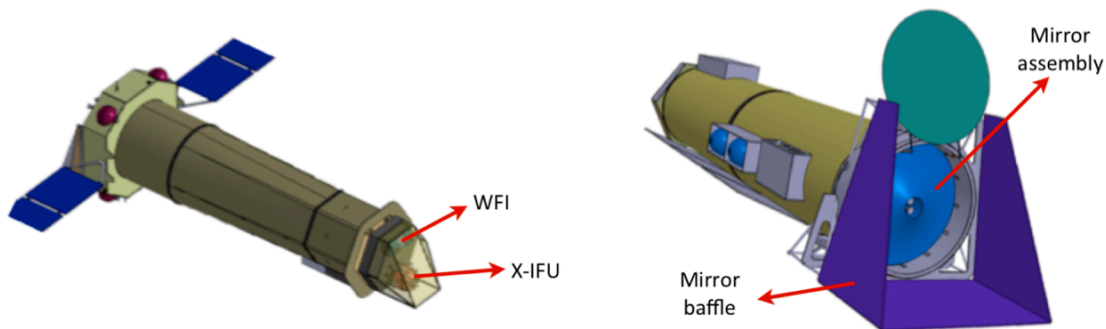


Figure 12: *Athena+* spacecraft designs. **Left:** *Athena+* Astrium-UK satellite designs provided along the preparation of this white paper. **Right:** Same for Thales-Alenia. For the Astrium design the interchangeable instruments are shown at the bottom, where the optics module is at the top and the solar panels will be unfolded. For the Thales-Alenia design the solar panels are body mounted and the optics, with its unfolded cover and sunshield are visible.

Mission and science operations are conventional with community based instrument and science data centre teams providing further support to ESA, as required. *Athena+* is an observatory whose program will be largely driven by calls for proposals from the scientific community, but may be complemented by key programs for science goals requiring large time investments. A nominal mission lifetime of 5 years would allow the core science goals set out in this White Paper to be achieved, while preserving a large fraction of the available time for broad based science programs.

Strong international interest in the mission has been expressed, based on the earlier collaborations with Japan (JAXA) and the US (NASA/GSFC, NIST) for *IXO* and *Athena*. These contributions could potentially reduce the costs to ESA and/or the ESA Member States, but *Athena+* can be implemented independently by Europe alone.

5. REFERENCES

Working Groups and Supporting Papers: As part of the process of developing this White Paper we set up a number of Working Groups to look in detail at various science areas within the theme; much of the work of these groups is of course reflected in this document. The more detailed reports from the working groups provide more extensive discussion of the science topics and additional simulations and references. These reports, which we refer to as supporting papers, are available at <http://www.the-athena-x-ray-observatory.eu>.

- Aird, J. et al., 2010, *NRAS*, 401, 2531
 Begelman M., et al., 2006, *MNRAS*, 370, 289
 Berti, E & Volonteri, M., *ApJ*, 684, 822
 Birzan, L., et al., 2008, *ApJ*, 686, 859
 Blandford, R. D., McKee, C. F., 1982, *ApJ*, 255, 419
 Blandford, R. D., Znajek, R. L., 1977, *MNRAS*, 179, 433
 Bohringer, H., Werner, N., 2010, *A&ARv*, 18, 127
 Cen, R. & Ostriker, J.P. 2006, *ApJ*, 650, 560
 Crawford, C. S., et al., 1999, *MNRAS*, 306, 857
 Croton D.J., et al., 2006, *MNRAS*, 365, 11
 Di Matteo T, Springel V, Hernquist L. 2005. *Nature* 433:604–7
 Ettori, S., et al., 2004, *MNRAS*, 354, 111
 Fabian, 2012, *ARAA*, 50, 455
 Fabian, A. C., Zoghbi, A., Ross, R. R., et al., 2009, *Nature*, 459, 540
 Fabian, A. C., et al., 2000, *PASP*, 112, 1145
 Fabian, A.C., et al., 2003, *MNRAS*, 344L, 43
 Giodini et al., 2013, *SSRvin press*, arXiv:1305.3286
 Gobat et al., 2011, *A&A*, 526, 133
 Heger, A. & Woosley, S.E. 2010, *ApJ* 724, 341
 King A., 2005, *ApJ*, 635, L121
 King, A.R., & Pounds, K.A., 2003, *MNRAS*, 345, 657
 Kormendy J., Ho, L. C., 2013, *ARA&A* in press, arXiv:1304.7762
 Kravtsov, A., Borgani, S., 2012, *ARA&A*, 50, 353
 Li Y., et al., 2007, *ApJ*, 665, 187
 McNamara, B.R., Nulsen, P.E.J., 2007, *ARA&A*, 45, 117
 Nandra, K. et al., 2007, *MNRAS*, 382, 194
 Nicastro, F. et al. 2013, *ApJ*, in press (arXiv1210.7177)
 de Plaa, J., et al., 2007, *A&A*, 465, 345
 Proga, D., Stone, J.M., & Kallman, T.R., 2000, *ApJ*, 543, 686
 Reiprich, T.H et al., *SSRv*, 2013, in press (arXiv:1303.3286)
 Schindler & Diaferio, 2008, *SSRv*, 134, 363
 Short, C. J, et al., 2010, *MNRAS*, 408, 2213
 Shull, J.M., et al. 2012, *ApJ*, 759, 23
 Stella, L., 1990, *Nature*, 344, 747
 Vazza, F., et al., M., 2011, *A&A*, 529, A17
 Volonteri, M. & Begelman, M, 2010, *MNRAS*, 409, 1022
 Wagner, A.Y., et al. 2013, *ApJ*, 763, L18
 Werner et al. 2008, *SSRv*, 134, 337