5.5 SIR Infrared Mapping Spectrometer
PI: U. Keller (MPAE - D); Co-Is: D, UK, IRL, + ESA

Science and Technology Objectives:
- Near infrared point-spectrometer, NIR mapping of lunar southern hemisphere (300 m resolution), angular spectroscopy of the surface features
- Flight qualification of a miniature monolithic grating spectrometer derived from commercial device (Zeiss)
- 0.9±2.4 μm, 256 channels, resolution 6 nm/ pixel.
- Correlation to AMIE 950 nm filter

![SIR infrared spectrometer](image)

Fig. 12: SIR infrared spectrometer

- Lunar Spectral mapping:
  - Discrimination of pyroxenes and olivine
  - Olivine from mantle: crustal differentiation/evolution
  - SPA exposed materials from mantle
  - SIR highest spatial resolution 300 m of units on central peaks, walls, rims and ejecta blankets of large impact craters, giving stratigraphy of lunar crust

![Reflectance spectra of the main lunar minerals](image)

Fig. 13: Reflectance spectra of the main lunar minerals

5.6 SMART-1 AMIE Multicolour Camera
PI: J.L. Josset, Co-Is: F, I, FI, NL, ESA

Science and Technology Objectives:
- Miniature camera for medium/high-res. multi-spectral imaging (27 m Moon South pole)
- Extension of data-set Apollo/Clementine
- Support to laser-link, OBAN, RSIS; aligned to SIR.

Main Experiment features:
- 5.3° FOV, 1024 x 1024 Si-CCD, 3 fixed wide-band filters (0.75±0.96 μm) + panchromatic + laser-link
- High-density CCD electronics. & Micro-DPU
- Packaged 3-D interconnect technology
- Shielded Off The Shelf components.
- 1.8 kg (Opt.Head 400 gr), 9W

![AMIE Laserlink experiment](image)

Fig. 14: AMIE Laserlink experiment
6. LUNAR–A

Lunar–A will be launched in mid-2004, on an M-V rocket, from Kagashima Space Center. It is expected to arrive at the Moon in mid-2005. Its objectives include:
- Study of lunar interior
- Seismometers and Heat flow probes
- Two penetrators (near and far side)
- Study of deep moonquakes (see Fig. 17)
- Mapping topographic data with B&W camera with emphasis on the near terminator

Fig. 16: View of Lunar–A with deployed 2 penetrators

Core Radius = 400 km

Fig. 17: Principle of Lunar–A probe of the Moon core

7. SELENE

SELENE, to be launched in 2005, covers 3 broad science objectives with a suite of instruments:
- Elemental and mineralogical composition:
  - X-ray spectrometer
  - Gamma Ray spectrometer
  - Multiband imager and Spectral Profiler
- Surface and subsurface structure/ tectonics:
  - Terrain Camera & Laser Altimeter
  - Lunar Radar Sounder
  - Gravity Field (VLBI and data relay)
- Lunar environment studies:
  - Lunar Magnetometer- Plasma Imager
  - Charged Particle Spectrometer and Plasma analyser
  - Radio Science S and X

Fig. 18: View of SELENE

Fig. 19: Schematics of SELENE subsystems

Fig. 20: SELENE mission profile
8. KEY LUNAR SCIENCE ISSUES

Fundamental science topics can be addressed by lunar exploration:

8.1 Formation and Evolution of Planets
- Understanding how rocky planets form and evolve
- Chemical constraints on Earth-Moon origin
- Signatures of accretional processes in inner planets
- South Pole Aitken Basin and large impact basins
- Evolution of Earth/Moon system
- Impacts: giant bombardment in the inner solar system

8.2 Comparative Geophysical Processes
- Volcanism, tectonics, cratering, erosion,
- Deposition of ices and volatiles
- Geophysics and Geochemistry

8.3 The Moon as collector of extraterrestrial samples
- Regolith Sample of the solar wind history
- Samples of ice cometary deposits in the last 10^9 years
- Samples from the Early and Evolving Earth
- Samples from Venus, Mars and asteroids

8.4 Preparing for Future Lunar Exploration
- Survey of lunar resources
  (minerals, volatiles, lighting)
- High resolution studies for landing sites/outposts
- Coordination between lunar missions.
- Environment studies in support of human exploration

Fig. 21: Phased International Lunar Exploration (ILEWG)

9. INTERNATIONAL LUNAR EXPLORATION

9.1 ILEWG

ILEWG, the International Lunar Exploration Working Group organized the ICEUM International Conferences on Exploration & Utilisation of the Moon (Beatenberg 94, Kyoto 96, Moscow 98, ESTEC 00) to consult the community on the definition of recommendations for the future.

Also scientists, engineers and exploration experts have discussed their priorities e.g. at COSPAR (Washington 92, Hamburg 94, Nagoya 98, Warsaw 00, Houston 02) and at EGS lunar sessions (Vienna 97, Nice 98, The Hague 99, Nice 00 – 04) [2,3,4,5]

10.2 ESA Lunar Exploration Studies

ESA has conducted several studies for lunar missions in the past, including a study for a Lunar Polar Orbiter (POLO) in the 80’s. At the beginning of the 90’s, a large consultation of the community led to the survey of possible science and social benefits for a renewed lunar exploration (cf ESA Report: Missions to the Moon [1]). A strategy for progressive exploration in 4 phases (precursor missions, landers, resource utilisation and deployment of large infrastructures, human permanent presence) was proposed by ESA, and agreed with other space agencies coordinated by ILEWG, the International Lunar Exploration Working Group, after the Beatenberg International Lunar Workshop [2].
Between 1994 and 1996 ESA studied a scientific lunar mission (MORO) as a contender for a medium cost mission. Technical studies (1994-1996) were also conducted on a lunar lander LEDA, and in 1996-1998 on the Euromoon study to land near the lunar south pole peak of quasi-eternal light. In parallel a series of key technologies for future lunar and planetary exploration were developed. ESA’s Long Term Space Policy Committee made recommendations for the future of European Space, including the emplacement of a permanent lunar base, and robotic precursors.

Other synergies for lunar exploration include:
- Preparation for Bepi Colombo
- Synergies with other science planetary missions (Rosetta, Mars express)
- ESA Advanced Technologies (Propulsion, Landing, Robotics, Telepresence…)

9.3 ICEUM4 ESTEC 2000, 10-14 July

http://solarsystem.estec.esa.nl/Moon2000/ilew4_4_frame.htm
The highlights from this symposium included [6]:
- Young Lunar Explorers Special Session
- Science of the Moon: Clementine, Prospector
- Key science issues
- Technology activities /Future Missions to the Moon:
  - SMART-1, Lunar-A, SELENE
  - Moon testbed for robotic outposts & telepresence
- Infrastructure, resources, expansion in solar system
- Recommendations and Space Agencies plans:
  - synergies with Mars and solar system exploration
  - new approaches and long term perspectives
- Foundation of Lunar Explorers Society
http://lunarexplorer.org/

9.4 International Lunar Missions under study
Several missions are under study, with very tentative launch dates:

2005? Ice Breaker Moon Rover
  US commercially sponsored lander/rover
2006? LUNARSAT
  educational mission by young lunar explorers
2006 Indian Lunar Mission
2008 Chinese Lunar Mission
  US Discovery Lunar Proposals
2007 SELENE B
  Soft Landing Technology Mission
2009 SELENE II
  Lunar Global Net/ Rover

10. TOOLS FOR LUNAR EXPLORATION

10.1 Tools for Science of and on the Moon
Technologies have to be further developed e.g.:
- Remote sensing miniaturised instruments
- Surface geophysical and geochemistry package
- Instrument deployment and robotic arm
- Close mobility, nano-rover, sampling, drilling
- Sample finder and collector
- Regional mobility: rover, navigation
- Resource utilisation, outpost installation
- Life sciences laboratories

10.2 Tools for Science from the Moon
Tools needed for autonomous robotic telescopes, and for the deployment of large telescopes:
- Deep Surveys Lunar transit telescope
- Dark Matter Lensing Telescope
- Near Earth Object telescopes
- Submm- IR telescopes in dark cold sites
- Hypertelescope interferometers for exoplanet studies
- Very Low Frequency astronomy on limb/far side sites
- SETI telescopes

10.3 The Moon Testbed for Exploration Technologies
The Moon can be used to demonstrate new technologies, and system level engineering e.g. for:
- Robotic laboratory
  - Mecha-electronics-sensors
  - Tele control, Telepresence, Virtual reality
  - Autonomy and Navigation
  - Artificially intelligent robots
- In-Situ Utilisation of lunar resources
- Regolith, Oxygen, glasses, metals utilisation
- Long term: He3 extraction
- Establishment of permanent lunar infrastructure
- Environmental protection aspects
- Support to human expansion to the Moon and beyond

Fig. 22: Brainstorming of Young Lunar Explorers
11. THE MOON AS A STEP TO THE SOLAR SYSTEM

The following is a list of aspects where science and technology demonstration on the Moon, will prepare for the expansion of human activities in the solar system:

11.1 Moon as a test bed for solar system exploration:
- Moon-Mars science synergies
- Instrument technologies
- Robotic outposts
- Tele-presence, Virtual reality
- Deployment of large infrastructures
- Earth-Moon L1 libration point for transfer
- Coordination humans and robots
- Medical aspects
- Biospheres on the Moon
- Human expansion in solar system

11.2 Astrobiology and Life sciences lab on the Moon:
- Analysis of organics from extraterrestrial samples
- Bacteria and extremes of life
- Survival, replication, mutation and evolution
- Extraterrestrial botanics: Growing plants on the Moon
- Animals: physiology and ethology on another planet
- Closed Ecological Life Support Systems,
- Greenhouses and Food

11.3 Expanding Life & Humans on the Moon:

A Lunar Exploration Roadmap can be given in an broad historical perspective:

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965</td>
<td>First organisms (Luna, Ranger)</td>
</tr>
<tr>
<td>1969</td>
<td>First humans (Apollo)</td>
</tr>
<tr>
<td>1994</td>
<td>Robotic precursors in orbit</td>
</tr>
<tr>
<td>1994</td>
<td>(Clementine, Prospector)</td>
</tr>
<tr>
<td>2003</td>
<td>Penetrators and Landers (SMART-1, SELENE)</td>
</tr>
<tr>
<td>2010</td>
<td>Virtual telepresence</td>
</tr>
<tr>
<td>2012</td>
<td>Evolving life on the Moon</td>
</tr>
<tr>
<td>2015</td>
<td>Sample returns</td>
</tr>
<tr>
<td>2020</td>
<td>Ecosystem experiments</td>
</tr>
<tr>
<td>2030</td>
<td>Plants, animals</td>
</tr>
<tr>
<td>2030</td>
<td>Resource utilization</td>
</tr>
<tr>
<td>2040</td>
<td>Robotic village</td>
</tr>
<tr>
<td>2060</td>
<td>Short crew missions</td>
</tr>
</tbody>
</table>

11.4 Human aspects of lunar exploration:
- Architecture design and operations of lunar base
- Man/robotics synergies, Life support systems
- Low gravity physiology laboratory, Telemedicine
- Psychology, Social and Multi-cultural Laboratory
- Infrastructures: communication, transport,
- Construction, exploitation
- Commercial development

Fig. 23 An advanced lunar base concept

11.5 Elements for Human Moon/Mars Exploration

New technologies and systems must be developed for future Human Exploration of the Moon and Mars:

- Advanced Launch /access to space
- Orbital Infrastructure
- Transport/ communication
- Habitable Descent / Ascent Vehicle
- Surface Power Generation
- In-Situ Fuel Production
- Robotic outposts and rovers
- Habitation Modules
- Workshop
- Scientific Laboratories
- Greenhouse / Agriculture Module
- Medical Centre
- Pressurized Rover
- Advanced EVA Suit
- Life Support Systems

© European Space Agency • Provided by the NASA Astrophysics Data System
12. THE ESA AURORA INITIATIVE

Having reached maturity in human space-flight with the development and operation of the International Space Station (ISS), the next step for human kind will be to reach out to other planets in the solar system. They will start first as explorers and then spend extended living and working periods on lunar and planetary bases.

Precursor missions with soft and precision landing, drilling and sample return, in-situ resource utilisation will also greatly advance our technology capability. Technology spin-offs are expected in spacecraft and crew systems autonomy, communications, navigation for precision targeting to distant places, data transmission technology for large volumes of data, information technologies, non-conventional power and propulsion systems, reliable and efficient thermal control for extreme temperatures, radiation hardened electronics, "self-repairing" and adaptable software, in-situ resources utilisation, and robotics.

In line with the European long-term strategy to explore the Solar System and the Universe and to prepare for the “next step” in human space exploration, a new Programme – Aurora, has been proposed by ESA. The programme proposal, with an initial period of technology studies in 2002-2004 outlines a preparatory framework for robotic and then human exploration missions. Its focus is on Mars, the Moon and Near Earth Objects. It is characterised by a phased scenario with remote sensing first, then automated planetary in-situ reconnaissance, sample return and eventually the transportation and assembly of the necessary infrastructure for human in-situ exploration at the final destination. The scenario will be implemented in full synergy with other planetary missions planned elsewhere. Its science objectives are the search for life in the Solar System, the search for the origin of the solar system and to gain knowledge on Near Earth Objects-NEOs. Apart from its technological challenges, the programme also serves as an exciting and peaceful goal to society.

The challenge for future lunar and planetary exploration is about science, technology and innovation; it is about people and cultures and it is about finding our place in the Universe, with the active involvement and excitement of the youth in particular.

13. ACKNOWLEDGEMENTS

We thank the members of ILEWG, the participants of the ICEUM, COSPAR and EGS sessions, for the discussion of ideas that led to this paper. We thank the teams of Clementine, Lunar Prospector, SMART-1, Lunar-A and SELENE for sharing the information on their projects, and for illustrations.

14. REFERENCES

14.1 Links
International Lunar Exploration working Group: http://www.estec.esa.nl/ilewg/
ESA Science web page: http://sci.esa.int
SMART-1 page: http://sci.esa.int/smart-1/
Lunar Explorers Society: http://lunarexplorer.org
Aurora: http://www.esa.int/export/esaHS/future.html

14.2 Publications
http://solarsystem.estec.esa.nl/Moon2000/ilewg4_frame.htm

14.3 Images credits: ESA, NASA, ISAS, ILEWG

Fig. 24: A biosphere on the Moon
MISSIONS TO VENUS


(1) Max-Planck-Institut für Aeronomie, Max-Planck-Str. 2, 37191 Katiensburg-Lindau, Germany; titov@linmp.mpg.de
(2) Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109-8099
(3) Vernadsky Institute of Geochemistry and Analytical Chemistry, RAS, Kosygin str. 19, 119991 Moscow, Russia
(4) Service d’Aéronomie, Université P&M Curie, 4 place Jussieu, 75252 Paris, France
(5) Laboratory for Extraterrestrial Physics, NASA/GSFC, Greenbelt, MD 20771, USA
(6) LASP, Campus Box 392, University of Colorado, Boulder, CO 80309-0392, USA
(7) ESA/RSSD/ESTEC, SCI-DO, 2200 AG Noordwijk, The Netherlands
(8) LEISA, Observatoire de Paris-Meudon, 5, pl. J.Janssen, 92195 Meudon, France
(9) Space Research Institute (IKI), RAS, Profsojuznaja 84/32, 117810 Moscow, Russia
(10) University of Michigan, 2455 Hayward st., 1414 SRB, Ann Arbor, MI 48109-2143, USA
(11) University of Hawaii, Inst. for Astronomy, 2680 Woodlawn Dr., Honolulu, HI 96822, USA
(12) Institute of Space and Astronautical Sciences, 3-1-1, Yoshinodai, Sagamihara, Japan
(13) University of California, Inst. of Geophys. and Planet. Phys., 405 Hilgard Ave., Los Angeles, CA 90095-1567, USA
(14) University of Oxford, Clarendon Laboratory, Parks Rd., Oxford OX1 3PU, UK
(15) NASA Ames Research Center, Mail Code 245-3, Moffet Field, CA 94035-1000, USA

ABSTRACT

Venus has always been a fascinating objective for planetary studies. At the beginning of the space era Venus became one of the first targets for spacecraft missions. Our neighbour in the solar system and, in size, the twin sister of Earth, Venus was expected to be very similar to our planet. However, the first phase of Venus spacecraft exploration in 1962-1992 by the family of Soviet Venera and Vega spacecraft and US Mariner, Pioneer Venus, and Magellan missions discovered an entirely different, exotic world hidden behind a curtain of dense clouds. These studies gave us a basic knowledge of the conditions on the planet, but generated many more questions concerning the atmospheric composition, chemistry, structure, dynamics, surface-atmosphere interactions, atmospheric and geological evolution, and the plasma environment. Despite all of this exploration by more than 20 spacecraft, the “morning star” still remains a mysterious world. But for more than a decade Venus has been a “forgotten” planet with no new missions featuring in the plans of the world space agencies. Now we are witnessing the revival of interest in this planet: the Venus Orbiter mission is approved in Japan, Venus Express – a European orbiter mission - has successfully passed the selection procedure in ESA, and several Venus Discovery proposals are knocking at the doors of NASA. The paper presents an exciting story of Venus spacecraft exploration, summarizes open scientific problems, and builds a bridge to the future missions.

1. VENUS BEFORE THE SPACE ERA

Venus is the brightest celestial object and its apparent motions were extensively observed in ancient times. Early telescopic observations by Huygens showed Venus as an extended disc with no evidence of features, leading him to conclude that all the light we see is probably scattered by an atmosphere [1]. In 1761 Lomonosov observed the Venus transit across the solar disc and saw a bright crescent around the planet that indicated that Venus has an atmosphere “equal to, if not greater than our earthly sphere” [2]. A breakthrough in Venus studies became possible in the 20-th century when spectroscopic observations gave the CO₂ and H₂O abundances at the cloud top (see [3] and references therein). The very low water vapour abundance ruled out the possibility that the clouds were composed of water droplets or ice crystals. Later measurements of the polarization phase curve implied sulfuric acid as the main constituent of the upper clouds. A further surprise was delivered by the early microwave studies that discovered a very high brightness temperature of ~600K, in remarkable contrast to ~240K measured in the thermal infrared range. The observations collected before the space era so poorly constrained the models that the conceptions
of Venus’ environment ranged from a planet with conditions very similar to those on the Earth, to a hell-like world with lakes of molten lead and acids on the surface.

2. EARLY MISSIONS TO VENUS

Venus was the first planet to which the earthlings decided to send a spacecraft, the Soviet Venera-1 in 1961. However this launch as well as two successive attempts failed. The first spacecraft that flew by Venus was the American Mariner-2 in 1962. It was the first successful planetary mission in the history of solar system exploration. Later, in 1967 Soviet Venera-4 reached the planet Venus and transmitted data directly from inside its atmosphere. Venera-4 was the first successful atmospheric probe. Since then and until the mid 80s, spacecraft were sent to Venus at almost every launch opportunity, i.e. each 19 months. Table 1 contains a chronology of the launches and scientific milestones of the missions to Venus (see also [4]). In the 60s-70s the Mariner programme of Venus exploration consisted of three fly-bys (Mariner-2, 5, and 10). Soviet Venus missions at that time were focused on the development of descent probe technology and in situ studies of the atmosphere, the clouds, and the surface. Both programmes were very successful due to the evolutionary strategy that implied development of spacecraft and payload “families” with step-by-step modifications between the launches. Venera-7 was the first spacecraft that safely reached the surface, in 1972. The high point of Venus exploration was in 1978 when a flotilla of seven spacecraft (Venera-11, 12, and Pioneer-Venus orbiter and four descent probes) visited the planet. Pioneer Venus Orbiter was the longest mission at Venus so far. It studied the upper and middle atmosphere for about 14 years. In 1985 the first balloons were injected into the atmosphere of another planet as a part of the Soviet VEGA mission to Venus and comet Halley [5]. An important milestone in the exploration of Venus’ surface was the US Magellan mission in 1990-94 [6,7] that completed the radar mapping of the Venus surface started by the Pioneer-Venus and Venera-15, 16 orbiters. Then, in the mid 1980s, Venus became the “forgotten planet” in a sense that no new missions to the planet were planned.

<table>
<thead>
<tr>
<th>Mission</th>
<th>Arrival at Venus</th>
<th>Type</th>
<th>Mission results and “firsts”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mariner-2</td>
<td>14.12.1962</td>
<td>Fly-by</td>
<td>First fly-by; Approach: 34,833 km; no magnetic field found.</td>
</tr>
<tr>
<td>Venera-4</td>
<td>18.10.1967</td>
<td>DP</td>
<td>First entry probe; T-P structure down to ~25 km; CO₂ and N₂ content; atmospheric dynamics by Doppler tracking; H-corona observed in L-alpha</td>
</tr>
<tr>
<td>Mariner-5</td>
<td>19.10.1967</td>
<td>Fly-by</td>
<td>Approach 4,100 km; radio occultation sounding of the atmosphere.</td>
</tr>
<tr>
<td>Venera-5</td>
<td>16.05.1969</td>
<td>DP</td>
<td>Similar to those of Venera-4.</td>
</tr>
<tr>
<td>Venera-6</td>
<td>17.05.1969</td>
<td>DP</td>
<td>Similar to those of Venera-4.</td>
</tr>
<tr>
<td>Venera-7</td>
<td>15.12.1970</td>
<td>DP</td>
<td>First soft landing; T-P structure down to 0 km.</td>
</tr>
<tr>
<td>Venera-8</td>
<td>22.07.1972</td>
<td>DP</td>
<td>T-P structure down to 0 km, detection of light under clouds; surface composition by GRS.</td>
</tr>
<tr>
<td>Mariner-10</td>
<td>05.02.1975</td>
<td>Fly-by</td>
<td>Approach: 5,700 km; cloud imaging and measurements of their motions.</td>
</tr>
<tr>
<td>Venera-9</td>
<td>22.10.1975</td>
<td>Orbiter</td>
<td>First orbiter; nightglow spectroscopy, plasma environment.</td>
</tr>
<tr>
<td>Venera-10</td>
<td>25.10.1975</td>
<td>Orbiter</td>
<td>First surface panorama; T-P structure, cloud structure by nephelometer and SP; reflectivity of surface by SP.</td>
</tr>
<tr>
<td>Pioneer Venus-1</td>
<td>04.12.1978</td>
<td>Orbiter</td>
<td>The longest mission at Venus (~14 years); Long term studies of the upper atmosphere and plasma environment; imaging; IR study of the mesospheric temperature structure and cloud tops.</td>
</tr>
<tr>
<td>Pioneer Venus-2</td>
<td>09.12.1978</td>
<td>4 DP</td>
<td>T-P structure down to ~12 km; Atmospheric composition; cloud structure and particle properties, radiative energy balance.</td>
</tr>
<tr>
<td>Venera-11</td>
<td>25.12.1978</td>
<td>Fly-by</td>
<td>Scattered solar radiation by SP; Atmospheric composition by GC, MS, and HS; cloud structure by nephelometer; search for lightning, atmospheric dynamics by Doppler tracking.</td>
</tr>
<tr>
<td>Venera-12</td>
<td>21.12.1978</td>
<td>Fly-by</td>
<td>Scattered solar radiation by SP; Atmospheric composition by GC, MS, and HS; cloud structure by nephelometer; search for lightning, atmospheric dynamics by Doppler tracking.</td>
</tr>
<tr>
<td>Venera-13</td>
<td>01.03.1982</td>
<td>Fly-by</td>
<td>Colour panoramas, surface elemental analysis by X-RFS; scattered solar radiation by SP; Atmospheric composition by GC, MS, and HS; cloud structure by nephelometer; search for lightning, atmospheric dynamics by Doppler tracking.</td>
</tr>
<tr>
<td>Venera-14</td>
<td>05.03.1982</td>
<td>Fly-by</td>
<td>Colour panoramas, surface elemental analysis by X-RFS; scattered solar radiation by SP; Atmospheric composition by GC, MS, and HS; cloud structure by nephelometer; search for lightning, atmospheric dynamics by Doppler tracking.</td>
</tr>
<tr>
<td>Venera-15</td>
<td>10.10.1983</td>
<td>Orbiter</td>
<td>Radar mapping of the surface; high resolution thermal emission spectroscopy of the mesosphere and cloud tops.</td>
</tr>
<tr>
<td>Venera-16</td>
<td>14.10.1983</td>
<td>Orbiter</td>
<td>Radar mapping of the surface.</td>
</tr>
<tr>
<td>VEGA-1</td>
<td>11.06.1985</td>
<td>Fly-by</td>
<td>First balloons at ~55 km; T measurements, dynamics by Doppler tracking; photometry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Balloon</td>
<td>DP: first precise T-P profile down the surface; in situ measurements of cloud composition</td>
</tr>
</tbody>
</table>

© European Space Agency • Provided by the NASA Astrophysics Data System
3. EXPERIMENTAL TOOLS

The missions to Venus used a variety of remote sensing and in situ techniques. Their combination increased the reliability of the results. However in some cases different tools gave significantly diverging results due to the extreme conditions on the planet. For instance, the first measurements of abundance of some chemically important species strongly varied from one experiment to another. Only later observations and re-analysis of the old data removed the controversies. The mass limitation intrinsic to all space missions was, in the case of Venus descent probes, worsened by the very short life time on the surface. The maximum duration of surface operations was about 2 hours in the case of the Venera-13 lander. The VEGA mission showed the usefulness of balloons for Venus robotic exploration. New techniques that have recently become available for planetary studies are spectro-imaging in the near infrared range, and microwave sounding. The first provides access to the lower atmosphere and even to the surface of Venus by using near infrared transparency “windows” discovered in the mid 1980s. Radiation from the lower atmosphere leaks to space in several narrow spectral intervals between 1 and 2.5 μm and can be measured on the night side. Such measurements can be used to study the composition and structure of the lower atmosphere, the dynamics of the cloud layer, and surface properties thus providing an excellent tool to study the deep atmosphere from orbit. Microwave sounding provides extremely high spectral resolution (up to 10^4) - that is high enough to resolve individual spectral lines, combined with very high sensitivity. Microwave groundbased observations have already proved their efficiency, and similar observations from orbit will soon become a powerful tool to study the composition, structure, and dynamics of the Venus atmosphere.

The role of groundbased observations for Venus studies decreased significantly after the beginning of space era. However it is still not negligible and sometimes groundbased observations bring discoveries that significantly change the whole strategy of Venus exploration. An excellent example is the discovery of the emission from the Venus nightside in the near IR spectral windows [18]. Groundbased observations also provide important support for the spacecraft missions by tracking the descent probes and balloons to obtain very important results on the atmospheric dynamics.

4. CURRENT KNOWLEDGE AND REMAINING PROBLEMS

4.1 Atmospheric structure

The temperature structure of the Venus atmosphere was extensively studied by various techniques onboard the orbiters and descent probes. Below about 40 km the temperature profile is virtually constant all over the planet. The equator to pole and day-night differences do not exceed 5K (Fig. 1). The greenhouse effect produced by CO₂, some trace gases (H₂O, SO₂), and cloud particles maintains the surface temperature at ~735K. Without it Venus would have had a surface temperature closer to that of the Earth. The temperature structure inside the cloud layer and especially in the mesosphere (60-100) km shows significant latitudinal variability which is not well understood. The thermospheric temperature above 120 km shows strong day-night variability, with remarkably low night-time temperatures. The most important unexplored issues in this field are the following. 1) Detailed study of the

![Fig. 1. Structure of the Venus (solid lines) and Earth (dashes) atmospheres.](image)
latitude and time variability of the mesospheric temperature structure and its correlation with the structure of the cloud tops, especially in the polar regions; 2). Measurements of the temperature gradient in the lower atmosphere to determine stability regions; 3). Role of various trace gases in the greenhouse effect and in the cloud chemistry.

4.2 Cloud layer
Venus is completely shrouded by clouds. The cloud system occupies the altitude range between -80 km and 40 km (Fig. 2), an absolute record among the terrestrial planets. Venus' clouds are not as dense as the terrestrial ones: visibility is about 1 km. The upper cloud consists of sulfuric acid particles a few microns in diameter with number densities of about 100-1000 cm$^{-3}$. Although polarimetric and spectroscopic observations gave strong evidence of sulfuric acid composition there are at least two unsolved problems in the cloud particle composition and chemistry. The first one is the nature of the unknown ultraviolet absorber that is located in the upper cloud and gives Venus a yellowish colour. This absorber is a very important agent in the atmospheric heat balance since about half of the solar flux received by Venus is deposited in the upper cloud due to its presence. The second problem is related to the composition of the main cloud deck. Experiments on the PV Large Probe led to the conclusion that cloud particles at -55-50 km are probably crystals, ruling out a sulfuric acid composition. Moreover, in situ analysis on the Venera-13,-14 descent probes found large amount of chlorine and phosphorus compounds present in the particulate matter. Recent imaging in the near IR spectral windows by Galileo also showed significant spatial inhomogeneities in the cloud opacity that implies complex dynamics of the deep cloud layer which is not understood at all.

4.3 Atmospheric composition and chemistry
The main atmospheric gases on Venus are carbon dioxide (-96.5%) and nitrogen (-3%). The total amount of minor constituents is less than 1% but these gases are responsible for very complicated chemistry of the atmosphere, cloud layer, and interactions with the surface. The abundance of the main trace gases (SO$_2$, COS, CO, H$_2$O) is shown in Fig. 3. These data and their accuracy do not constrain existing models of the atmospheric chemistry very well [8]. Moreover, the measurements of other minor species that could be of importance in the chemical cycles are still missing. The lower atmosphere of Venus is thought to be in thermochemical equilibrium with the surface [9]. At least the observed carbon dioxide pressure is in good agreement with the model of a calcite-wollastonite buffer. However the observed SO$_2$ abundance is about an order of magnitude higher than that expected from thermochemical equilibrium. In this case, the replenishment of sulfur dioxide would require additional sources, like active volcanism, for instance. The sulfuric acid clouds are probably produced by photochemical reactions between sulfur dioxide and water vapour at the cloud tops. Any future progress in our understanding of the Venus chemical cycles is closely related to our ability to obtain more precise measurements of the abundances of key trace gases and their variations, especially in the lower atmosphere. It would also require the study of cloud chemistry and the surface mineralogy and oxidation state.
4.4 Dynamics of the atmosphere
Two types of global circulation patterns have been observed in the Venus atmosphere [10, 11]. The mesosphere and the troposphere (0-100 km) are involved in zonal retrograde superrotation with wind speeds reaching a maximum of ~120 m/s at the cloud tops (Fig. 4). This regime is well described by the balance of centrifugal forces and pressure gradients although the mechanism that maintains it is poorly understood. Above 100km in the thermosphere the day-night temperature difference produces solar-anti-solar flow which was expected to be much stronger than that observed. Obviously we do not understand the decelerating mechanisms. The most important open issues are related to the vertical momentum transport, the role of Hadley cell (or cells) and in the meridional circulation, coupling between the thermospheric and mesospheric dynamics, and the nature of the polar vortices. Of key importance in resolving these problems are coherent observations of the atmospheric dynamics and temperature fields at different altitudes, including characterization of winds and stability in the lower atmosphere.

![Fig. 4. Circulation regimes of the Venus atmosphere (left) and vertical profile of the zonal wind (right).](image)

4.5 Energy balance
About 80% of incident solar radiation is scattered back to space by the cloud layer and thick atmosphere, only 10% reaches the surface, and about 10% is deposited in the upper cloud due to the presence of the unknown UV absorber [12]. Atmospheric gases like CO₂, H₂O, SO₂ and also cloud particles produce a strong greenhouse effect (~500K) due to their ability to absorb thermal radiation (Fig. 1) while partially transmitting solar radiation to the surface. The precise role of the various greenhouse agents and cloud opacity variations are major issues for future observations and numerical modeling.

4.6 Plasma environment
Venus has no magnetic field so the solar wind directly interacts with the atmosphere forming complex and highly dynamic features like the ionopause, tail rays, and holes (Fig. 5). Despite long-term observations by PV Orbiter the variability of the upper atmosphere and ionosphere of Venus is not completely understood. The processes of interaction with the solar wind are of great importance for the evolution of the atmosphere since they control the rate of escape [13, 14, 15].

![Fig. 5. Schematic representation of the Venus plasma environment (from Brace and Kliore 1991).](image)

4.7 Surface of Venus
Radar mapping by Pioneer-Venus, Venera-15,16, and Magellan showed that ~80% of the surface is occupied by the lowlands - smooth bedrocks covered with basalt debris (Fig. 6). Other landforms include tesserae (ridged highlands), volcanoes, and impact craters. The main features of the Venus surface are uniform distribution of the impact craters, absence of plate.

![Fig. 6. Venus topography map from Magellan with landing sites of descent probes (top). Venera-13 panorama of the surface (bottom).](image)
tectonics, and very low weathering rate [16, 17]. The Radar observations imply that the Venus surface is very young and about <500 Mye... global resurfacing. The main open questions in the physics of the Venus surface are the mineralogy and the history of resurfacing and volcanism.

4.8 Origin and evolution of the atmosphere
The relative abundance of noble gases and their isotopes is the main observational clue to the origin and evolution of the planets. Comparison of the noble gas composition of the three terrestrial planets helps to determine the role of different processes in the formation of the inner solar system, such as accretion of gas and planetesimals, outgassing from interiors, cometary supply, and catastrophic events. The noble gas abundance on Venus is much higher than that on the Earth and Mars and much more solar-like. However, measurements of isotopes of heavy gases like krypton and xenon on Venus are still missing. A very intriguing problem in Venus evolution is the depletion of water on the planet. The D/H ratio on Venus was found to be about a factor of 150 greater than that on the Earth, which implies the presence of large amounts of water on the planet in the past. Mechanisms of atomic escape and fractionation that resulted in significant enrichment in deuterium are still to be studied.

4.9 Uniqueness of Venus
Early studies showed that in many respects Venus is a unique planet on which various phenomena could be studied from first principles. 1). Venus is a natural laboratory for the study of thermochemistry, surface-atmosphere interactions, and radiative transfer at high pressures and temperatures. 2). Venus gives us an example of atmospheric dynamics on a slowly rotating planet. 3). Venus provides an opportunity to study the plasma environment and its interaction with the solar wind on a non-magnetic planet. 4). Venus has the youngest surface among terrestrial planets and very low weathering rate that implies that the footprints of geological processes are well preserved.

5. STRATEGY OF FUTURE VENUS EXPLORATION AND PROPOSED MISSIONS
The first phase of Venus exploration by the Venera, Pioneer Venus, VEGA, and Magellan missions established a basic description of the physical and chemical conditions prevailing in the atmosphere, near-planetary environment, and at the surface. These studies solved some problems that appeared before the space era. At the same time, they raised many questions about physical processes sustaining these conditions, most of which remain unsolved. The fundamental mysteries of Venus are related to the global atmospheric circulation, the atmospheric chemical composition and its variations, the surface-atmosphere physical and chemical interactions including volcanism, the physics and chemistry of the cloud layer, the thermal balance and role of trace gases in the greenhouse effect, the origin and evolution of the atmosphere, and the plasma environment and its interaction with the solar wind. Beyond the specific case of Venus, resolving these issues is of crucial importance in a comparative planetology context and notably for understanding the long-term climatic evolution processes on Earth. Table 2 shows key open questions in the physics of Venus and required critical measurements that should be carried out by future missions. Some of them can be implemented immediately, while more complicated missions like sample return or long-lived balloons or probes would require significant technological developments. The great number of unsolved problems encourages the planetary community to propose missions to Venus. Table 3 contains the list of missions proposed recently in Europe, USA, and Japan and their scientific objectives. For the time being only Planet-C – a Japanese orbiter mission to Venus in 2007, focused on the study of atmospheric dynamics - is approved.

6. CONCLUDING REMARKS
About three decades of Venus exploration by Soviet and American spacecraft brought us a basic understanding of the conditions on the planet. At the same time they raised many fundamental questions concerning the processes responsible for the observed phenomena. The main mysteries of Venus science fall in several categories: global atmospheric circulation, chemistry of the atmosphere and cloud layer, thermal balance and role of gases and clouds in greenhouse effect, origin and evolution of the atmosphere, plasma environment and its interaction with the solar wind, and geological history of the planet. Resolving these issues is of crucial importance in a comparative planetology context. In recent years planetary community has formulated a strategy for future Venus exploration. Several missions are being proposed in Europe, USA, and Japan. Planet-C – the Japanese orbiter mission in 2007 that focuses on the study of Venus meteorology has recently been approved.
Table 2. Open questions in the Venus physics and critical measurements

<table>
<thead>
<tr>
<th>Open questions</th>
<th>Critical measurements</th>
<th>Missions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Atmospheric composition</strong></td>
<td>• Mapping the composition and cloud properties</td>
<td>Venus Express</td>
</tr>
<tr>
<td>• Abundance of trace gases</td>
<td>• In situ gas analysis</td>
<td>VESAT</td>
</tr>
<tr>
<td>• Chemical cycles</td>
<td>• O₃ fugacity</td>
<td>VESPER</td>
</tr>
<tr>
<td>• Surface-atmosphere interaction</td>
<td>• Surface composition</td>
<td>Morning Star</td>
</tr>
<tr>
<td>• Stability of the atmosphere</td>
<td></td>
<td>VAMP</td>
</tr>
<tr>
<td><strong>Atmospheric circulation</strong></td>
<td>• Tracking cloud features in UV and IR</td>
<td>Planet C</td>
</tr>
<tr>
<td>• Mechanism of superrotation</td>
<td>• Direct wind measurements in the mesosphere and thermosphere</td>
<td>Venus Express</td>
</tr>
<tr>
<td>• Thermospheric circulation</td>
<td>• Nightglow observations</td>
<td>VESAT</td>
</tr>
<tr>
<td>• Wave phenomena (polar dipole)</td>
<td>• 3-D temperature field</td>
<td>VESPER</td>
</tr>
<tr>
<td></td>
<td><strong>Descent probes/balloons</strong></td>
<td>Morning Star</td>
</tr>
<tr>
<td></td>
<td>• VLBI tracking of DP and balloons</td>
<td>Venus Multiprobe mission</td>
</tr>
<tr>
<td></td>
<td>• Wind field in the lower atmosphere</td>
<td></td>
</tr>
<tr>
<td><strong>Radiative balance</strong></td>
<td>• Radiative fluxes</td>
<td>Planet C, VESAT</td>
</tr>
<tr>
<td>• Role of trace gases in the greenhouse</td>
<td>• Atmospheric temperatures</td>
<td>Venus Express</td>
</tr>
<tr>
<td></td>
<td><strong>Orbiter/descent probes</strong></td>
<td>Morning Star</td>
</tr>
<tr>
<td><strong>Origin and evolution of the atmosphere</strong></td>
<td>• Mass spectrometry</td>
<td>VAMP</td>
</tr>
<tr>
<td></td>
<td>• Fluxes of neutral atoms</td>
<td>Morning Star</td>
</tr>
<tr>
<td></td>
<td><strong>Descent probes</strong></td>
<td>Venus Express</td>
</tr>
<tr>
<td></td>
<td>• Mass-spectrometry of isotopes</td>
<td></td>
</tr>
<tr>
<td><strong>Plasma environment and solar wind interactions</strong></td>
<td>• Spectroscopy and in situ measurements of ions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Field measurements</td>
<td>Venus Express</td>
</tr>
<tr>
<td></td>
<td>• Radar sounding</td>
<td></td>
</tr>
<tr>
<td><strong>Surface, geology, and interiors</strong></td>
<td>• Subsurface radar sounding</td>
<td>Venus Express</td>
</tr>
<tr>
<td></td>
<td>• Radar imaging with ~50 m resolution</td>
<td>VESAT</td>
</tr>
<tr>
<td></td>
<td>• Surface imaging in 1 micron “window”</td>
<td>VAMP</td>
</tr>
<tr>
<td></td>
<td><strong>Descent probes</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Surface composition analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Seismometry</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Surface imaging</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Proposed missions to Venus and their scientific goals.

<table>
<thead>
<tr>
<th>Mission</th>
<th>Type</th>
<th>Agency</th>
<th>Science goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venus Express</td>
<td>Orbiter</td>
<td>ESA</td>
<td>Structure and composition of the atmosphere and clouds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Atmospheric dynamics and radiative balance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plasma environment and escape processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Subsurface sounding and surface properties</td>
</tr>
<tr>
<td>Planet-C(*)</td>
<td>Orbiter</td>
<td>ISAS(Japan)</td>
<td>Atmospheric dynamics and plasma investigation (optional)</td>
</tr>
<tr>
<td>VESAT</td>
<td>Orbiter</td>
<td>NASA</td>
<td>Global dynamics, chemistry, surface</td>
</tr>
<tr>
<td>VESPER</td>
<td>Orbiter</td>
<td>NASA</td>
<td>Global chemistry and dynamics</td>
</tr>
<tr>
<td>Morning Star</td>
<td>Orbiter &amp; DP</td>
<td>NASA</td>
<td>Chemistry, evolution, thermal balance</td>
</tr>
<tr>
<td>Multiprobe mission</td>
<td>16 DP</td>
<td>NASA</td>
<td>Structure and dynamics</td>
</tr>
<tr>
<td>VAMP</td>
<td>DP</td>
<td>NASA</td>
<td>Evolution and chemistry</td>
</tr>
</tbody>
</table>

(*) The only mission approved so far.
7. REFERENCES
MISSIONS TO MARS
A.F. Chicarro
and
The Science Team,
European Space Agency, Space Science Department, ESTEC/Code SCI-SB, Postbus 299, 2200 AG Noordwijk, The Netherlands (email: agustin.chicarro@esa.int).

ABSTRACT

This presentation started with a historical perspective of the astronomical discovery of Mars and followed by an overview of previous missions to Mars by the United States and the Soviet Union. Recently launched missions, such as Nozomi, Mars Global Surveyor and Mars Odyssey were addressed in more detailed, as well as a few other missions soon to be launched. Among these, Mars Express is particularly relevant as the first European mission towards the red planet, and the talk concentrated on it, including both the Mars Express orbiter spacecraft and the Beagle-2 lander to be launched in 2003.

The European Space Agency and the scientific community have performed concept and feasibility studies for more than ten years on potential future European missions to the red planet (Marsnet, InterMarsnet), focusing on a network of surface stations complemented by an orbiter, a concept which is being implemented by the CNES-led Netlander mission to be launched in 2005. Before that, however, the ESA Mars Express mission includes an orbiter spacecraft and a small lander module named Beagle-2 in remembrance of Darwin’s ship Beagle. The mission, to be launched in May-June 2003 by a Russian Soyuz rocket, will recover some of the lost scientific objectives of both the Russian Mars-96 mission and the ESA InterMarsnet study, following the recommendations of the International Mars Exploration Working Group (IMEWG) after the failure of Mars-96, and also the endorsement of ESA’s Advisory Bodies that Mars Express be included in the Science Programme of the Agency.

The specific scientific objectives of the Mars Express orbiter are: global high-resolution imaging with 10 m resolution and imaging of selected areas at 2 m/pixel, global IR mineralogical mapping, global atmospheric circulation study and mapping of the atmospheric composition, sounding of the subsurface structure down to the permafrost, study of the interaction of the atmosphere with the surface and with the interplanetary medium as well as radio science. The goals of the Beagle-2 lander are: geology, geochemistry, meteorology and exobiology of the landing site.

The scientific payload on the Mars Express orbiter includes a Super/High-Resolution Stereo Colour Imager (HRSC), an IR Mineralogical Mapping Spectrometer (OMEGA), a Planetary Fourier Spectrometer (PFPS), a Subsurface-Sounding Radar Altimeter (MARSIS), an Energetic Neutral Atoms Analyser (ASPERA), an UV and IR Atmospheric Spectrometer (SPICAM) and a Radio Science Experiment (MaRS). The Beagle-2 lander includes a suite of imaging instruments, organic and inorganic chemical analysis, robotic sampling devices and meteorological sensors (see Table).

The Mars Express mission will address the issue of astrobiology on Mars both directly and indirectly. The majority of instruments on the orbiter will look for indications of favourable conditions to the existence of life, either at present or during the planet’s past, and in particular for traces of liquid, solid or gaseous water. Therefore, the HRSC camera will take pictures of ancient riverbeds, the OMEGA spectrometer will look for minerals with OH- radicals formed in the presence of water, the MARSIS radar will look for subsurface ice and liquid water, the PFPS and SPICAM spectrometers will analyse water vapour in the atmosphere, and finally ASPERA and MaRS will study neutral atom escape from the atmosphere, in particular O2 coming from water and carbonates. The instruments on Beagle-2 will also look for the presence of water in the soil, rocks and the atmosphere, but will also try to find traces of life with direct measurements, such as presence of methane (CH4) indicative of extinct life and a larger amount of the light C12 isotope compared to the heavier C13, which would even indicate the existence of extinct life. Since NASA’s Viking mission in 1976, it is the first time that the exhaustive search for life is so central to a space mission to Mars.

Current design estimates allow for an orbiter scientific payload of about 110 kg and 65 kg total lander mass (at launch) compatible with the approved mission scenario. The Beagle-2 lander was selected due to its innovative scientific goals and challenging payload. Beagle-2 will deploy a sophisticated robotic-sampling arm, which could manipulate different types of tools and retrieve samples to be analyzed by the geochemical instruments mounted on the lander platform. One of the tools to be deployed by the arm is a ‘mole’ capable of subsurface
sampling to reach soil unaffected by solar-UV radiation, another is a corer/grinder to reach the rock under the weathering varnish.

A Soyuz-Fregat launcher will inject a total of about 1200 kg into Mars transfer orbit in early June 2003, which is the most favorable launch opportunity to Mars in terms of mass in the foreseeable future. The Mars Express orbiter is 3-axis stabilized and will be placed in an elliptical martian orbit (250 x 10142 km) of 86.35 degrees inclination and 6.75 hours period, which has been optimized for communications with Beagle-2, the Netlanders, as well as NASA landers or rovers to be launched both in 2003 and 2005. The Beagle-2 lander module will be independently targeted from separate arriving hyperbolic trajectory, enter and descend through the martian atmosphere in about 5 min, and land with an impact velocity <40 m/sec and an error landing ellipse of 100 x 20 km. A preliminary Beagle-2 landing site has been selected in the Isidis Planitia area (10.6°N, 270°W). The nominal mission lifetime of one martian year (687 days) for the orbiter investigations will be extended by another martian year for lander relay communications and to complete global coverage. The Beagle-2 lander lifetime will be of a few months.

ESA provides the launcher, the orbiter and the operations, while the Beagle-2 lander is delivered by an UK-led consortium of space organizations. The orbiter instruments are provided by scientific institutions through their own funding. In addition to relaying the data from the Beagle-2 lander to Earth, Mars Express will also service landers and rovers from other agencies during its nominal/extended lifetime. The ground segment includes the ESA station at Perth, Australia, and the mission operations centre at ESOC. The Mars Express mission is now in Phase-C/D, with Astrium (formerly Matra Marconi Space) in Toulouse, France, as its Prime Contractor and involving a large number of European companies.

International collaboration, either through the participation in instrument hardware or through scientific data analysis is very much valued to diversify the scope and enhance the scientific return of the mission, such as NASA’s major contribution to the subsurface-sounding radar, and the use of its DSN for increased science data downloading and critical manoeuvres. Also, arriving at Mars at the very end of 2003, Mars Express will be followed by the Japanese Nozomi spacecraft a few days later. Both missions are highly complementary in terms of orbits and scientific investigations; Nozomi focusing on the study of the upper atmosphere of Mars as well as the interaction of the solar wind with the ionosphere from a highly elliptic equatorial orbit. Close cooperation, including scientific data exchange and analysis, is foreseen by the Nozomi and Mars Express teams within a joint ESA-ISAS programme of Mars exploration.

For more details on the Mars Express mission and its Beagle-2 lander:
http://sci.esa.int/marsexpress/
and http://www.beagle2.com/

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Instruments</th>
<th>Principal Investigators</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbiter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRSC</td>
<td>Super/High-Resolution Stereo Colour Imager</td>
<td>G. Neukum</td>
<td>D, F, RU, US, FI, I, UK</td>
</tr>
<tr>
<td>OMEGA</td>
<td>IR Mineralogical Mapping Spectrometer</td>
<td>J.P. Bibring</td>
<td>F, I, RU</td>
</tr>
<tr>
<td>PFS</td>
<td>Atmospheric Fourier Spectrometer</td>
<td>V. Formisano</td>
<td>I, RU, PL, D, F, E, US</td>
</tr>
<tr>
<td>MARSIS</td>
<td>Subsurface-Sounding Radar/Altimeter</td>
<td>G. Picardi &amp; J. Plaut</td>
<td>I, US, D, CH, UK, DK</td>
</tr>
<tr>
<td>SPICAM</td>
<td>UV and IR Atmospheric Spectrometer</td>
<td>J.L. Bertaux</td>
<td>F, B, RU, US</td>
</tr>
<tr>
<td>MaRS</td>
<td>Radio Science Experiment</td>
<td>M. Paetzold</td>
<td>D, F, US, A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Landers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BEAGLE-2</td>
<td>Suite of imaging instruments, organic and inorganic chemical analysis, robotic sampling devices and meteo sensors</td>
<td>C. Pillinger &amp; M. Sims</td>
<td>UK, D, F, HK, CH</td>
</tr>
</tbody>
</table>
Fig. 1. Mars Express orbiter in polar orbit around Mars.

Fig. 2. Deployed Beagle-2 lander on the surface of Mars.
MISSIONS TO MERCURY

Réjean Grard, Harry Laakso, Håkan Svedhem

Research and Scientific Support Department, European Space Agency
ESTEC, P.O. Box 299, 2200 AG Noordwijk, The Netherlands
rejean.grard@esa.int - harri.laakso@esa.int - hakan.svedhem@esa.int

ABSTRACT

Mercury is a poorly known planet. It is difficult to observe from Earth and to explore with spacecraft, due to its proximity to the Sun. Only the NASA probe Mariner 10 caught a few glimpses of Mercury during three flybys, more than 27 years ago. Still, this planet is an interesting and important object because it belongs, like our own Earth, to the family of the terrestrial planets. After reviewing what we know about Mercury and recapitulating the major findings of Mariner 10, we present the two missions, Messenger and BepiColombo, which will perform the first systematic exploration of this forgotten planet in 2009 and 2014, respectively.

1. INTRODUCTION

As the inner member of the planetary system (Fig. 1), Mercury plays an important role in constraining and testing theories of planetary formation. A better knowledge of the terrestrial planets is the key to understanding how conditions to support life have met in the Solar System and how the habitability of Earth will evolve. Terrestrial-like objects orbiting other stars are not accessible and the Solar System is the only laboratory where we can test models that will prove or disprove that such bodies may exist elsewhere in the Galaxy.

Due to its proximity to the Sun, Mercury is the least known of the terrestrial planets. From the Earth surface, Mercury can be observed only shortly before sunrise or after sunset, when the sky is not very dark. Earth-orbiting telescopes usually cannot point close to the Sun and target Mercury. Sending a planetary orbiter to this planet is not easy due to the large gravitational potential of the Sun and the harsh thermal environment. Mercury was visited only once by Mariner 10 which performed three flybys of the planet in 1974-1975. After 27 years, the Mariner 10 data have been fully exploited and have raised new questions that form the goals of the Messenger and BepiColombo missions which will be launched within the coming decade.

We first sum up the known features of Mercury and formulate the scientific objectives of the future missions. We then recapitulate the main characteristics of Mariner 10. We finally give an overview of Messenger and BepiColombo, and outline their synergy and complementarity. More information and related references can be found in a special issue of Planetary and Space Science dedicated to these missions and to the planet Mercury (Volume 49, No 14-15, 2002).

Fig. 1. The terrestrial planets. They share a common origin and each of them must be explored to understand all four.

2. THE PLANET MERCURY

Mercury was known to be an odd planet, even before the space age. With a radius of 2440 km, Mercury is slightly larger than our Moon but its density is out of line with those of other terrestrial planets (Fig. 2). Mercury revolves around the Sun in almost 88 days and rotates around its own axis in 2/3 of that time; as a result, one solar day equals two solar years on Mercury. Its orbit is very eccentric and its distance to the Sun varies from 0.3 to nearly 0.5 AU.


© European Space Agency • Provided by the NASA Astrophysics Data System
Mariner 10 imaged less than 50% of Mercury's surface, at a resolution comparable to that with which the Moon is observed from Earth. It discovered a planet poked with craters like the Moon (Fig. 3) but with distinctive features, such as the lobate scarps which may have recorded a global contraction of 1-2 km in radius during Mercury's thermal history. The NASA probe returned also images of the Caloris Basin which has a diameter of 1300 km and was caused by an impact of such a magnitude that it disturbed the terrain at the antipode (Fig. 4).

Before Mariner 10, it was assumed that Mercury's core was frozen and could not support any dynamo. Contrary to expectations, the planet was found to have a small but significant magnetic field, about one hundredth that of the Earth at the equator (Fig. 5). It was confirmed that Mercury, like the Moon, was surrounded by a rarefied atmosphere (exosphere), which can be observed from Earth. It was also discovered that the interaction of the planetary magnetic field with the solar wind led to the formation a small magnetosphere somewhat similar to that which surrounds the Earth. More recently Earth-based radar images might have revealed the existence of water ice deposits in permanently shadowed craters near the poles.

These are the questions that form the central rationale for future missions to Mercury:
- What will be found on the unseen side of the planet?
- How did the planet form and evolved geologically?
- Why is Mercury's density so high?
- What is its internal structure?
- Does it possess a liquid outer core?
- What is the origin of the magnetic field?
- What is the chemical composition of the surface?
- Is there any water ice in the polar regions?
- What are the constituents of the exosphere?
- How does the magnetic field interact with the solar wind?

Other objectives go beyond the exploration of Mercury as a planet and take advantage of its proximity to the Sun to test Einstein's theory of relativity. Mercury is also a vantage point for detecting Near Earth Objects, which are possible Earth impactors; these asteroids cannot be easily observed because their orbit lie, for the most part, within that of our planet.
Mariner 10 was launched in 1973 in a solar orbit and approached Mercury three times, on 29 March 1974 (altitude 707 km), 21 September 1974 (distance 53 000 km) and 16 March 1975 (altitude 327 km). The spacecraft had a mass of 526 kg and carried a scientific payload consisting of 2 cameras, ultraviolet and infrared imagers, a magnetometer, a plasma experiment and a high energy particle detector (Fig. 6).

3.2. Messenger

Messenger is part of the Discovery Program and was selected by NASA in July 1999. This mission is led by the Carnegie Institution of Washington and the John Hopkins University. The spacecraft will be launched by a Delta rocket, in March or May 2004, and arrive at Mercury in April 2009. The spacecraft has a dry mass of 476 kg. It is three-axis stabilized and protected from the direct solar illumination by a sunshade (Fig. 7). It carries a payload of 40 kg mostly dedicated to remote sensing. The interplanetary transit includes two Venus and two Mercury flybys. The chemical propulsion module carries 600 kg of propellant for deep space manoeuvre and orbit insertion. Messenger will return 61.8 Gb of information in a nominal life time of 1 year.

3. THE MISSIONS

3.1. Mariner 10

Fig. 6. The NASA Mariner 10 probe.

BepiColombo

BepiColombo is an approved mission in the scientific programme of ESA. It is conducted in cooperation with ISAS (Japan) and, possibly, another space agency. This mission is named after an Italian planetologist Giuseppe (Bepi) Colombo (1920-1984), who explained the 2:3 resonance of Mercury’s rotation and orbit periods, and suggested to NASA how to send Mariner 10 in an orbit that would enable the probe to achieve three flybys of Mercury. The spacecraft consists of several scientific modules: two orbiters and, possibly, one lander. These elements are undergoing definition studies and their final configurations might significantly differ from those illustrated by Fig. 8. The scientific payloads have not been selected yet.
Fig. 8. The three elements of the BepiColombo spacecraft: MPO, MMO and MSE.

The baseline scenario assumes that the three spacecraft elements are distributed over two composites simultaneously launched by two Soyuz-Fregat in January 2011. The interplanetary transits employ, like that of Messenger, two Venus and two Mercury flybys, and are supported by solar electric propulsion modules which are jettisoned upon arrival at Mercury. The transit times are typically 2.5 years. They could increase by 1 year if a lunar flyby is needed to accommodate larger launch masses, thus entailing arrivals in June 2014. Orbit insertions are achieved with chemical propulsion modules. ESA carries the responsibility for the fabrication of the propulsion systems, the interplanetary operations and the insertions into orbit.

**Mercury Planetary Orbiter (MPO)**

MPO is developed by ESA. This module is three-axis stabilized and points downwards, to facilitate the observation of the surface. The spacecraft has a mass of the order of 400 kg, after separation from the chemical propulsion module, and carries 70 kg of payload. An articulated one-meter diameter dish antenna permits the transmission of a data volume larger than 1000 Gb during a nominal life time of one year.

**Mercury Magnetospheric Orbiter (MMO)**

MMO will be supplied by Japan. This orbiter has a mass of 200 kg, including 25 kg of instruments. Dedicated to the study of the magnetospheric environment, the spacecraft is spin stabilized to enable the deployment of long electric wire antennas and to facilitate the azimuthal scan of particle detectors. The data volume returned in one year is close to 150 Gb.

**Mercury Surface Elements (MSE)**

MSE will make a detailed exploration of the landing site and provide observations that will serve as ground-truth measurements for the orbiters. The lander would have a mass of 40-50 kg and carry a payload of a few kg. The data volume relayed to Earth via MPO and MMO, in a nominal life time of one week, would amount to about 100 Mb. The provision of MSE is currently negotiated with Russia and will possibly be included in the mission baseline at the completion of the definition study.

4. **COMPARING THE FUTURE MISSIONS**

4.1. **The Payloads**

The payload of Messenger and the model payloads of the Bepicolombo orbiters are compared in Table 1, which reveals that the core instrumentations of MPO and Messenger are similar. Imaging in the visible is performed with a dual camera. The mineralogical composition of the surface is surveyed with infrared and ultraviolet spectrometers. UV spectrometry is also applied to the identification of the volatiles which constitute the exosphere. The elemental composition is mapped with X, gamma and neutron spectrometers. The morphology of the planet is studied with a laser altimeter. Owing to larger resources, more instruments might be carried by MPO, such as a telescope for Near Earth Objects.

Radioscience is based on the measurements of the propagation time and Doppler shift of electromagnetic signals, such as those used in the telecommunication links, and contributes information to gravimetry and fundamental physics. Thanks to the availability of signals with frequencies in the X and Ka bands, it is anticipated that MPO should achieve a better accuracy than Messenger, which uses the X band exclusively. Combining radioscience measurements with altimetry data, on Messenger, or with high resolution images, on MPO, will provide information about the rotational state of the planet, inclination and libration, and tell us whether the outer core of the planet is molten, thereby identifying the possible source of the magnetic field.

Messenger also includes a magnetometer, and an energetic particle and plasma analyzer for solar wind and magnetospheric measurements, but MMO is unrivaled in terms of field and particle instrumentation. The mass at the Scientific payload carried by the three BepiColombo elements is twice as large as that carried by Messenger. The overall data volume returned during the nominal life time is 18 times larger for BepiColombo than for Messenger.
Table 1. Payloads of the Mercury orbiters. The instruments marked with an asterisk are not part of the MPO core payload but may be accommodated if the resources are available.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>MPO</th>
<th>MMO</th>
<th>Messenger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow-angle camera</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wide-angle camera</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>VIS/IR spectrometer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UV/VIS spectrometer</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-ray spectrometer</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma-ray spectrometer</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutron spectrometer</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radio Science</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser altimeter</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetometer</td>
<td>X*</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dust mass spectrometer</td>
<td>X*</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Near-Earth Objects Telescope</td>
<td>X*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrared radiometer</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energetic neutral particle analyser</td>
<td>X*</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Energetic particle &amp; plasma analyser</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ion mass spectrum analyser</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electron spectrum analyser</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Electromagnetic field</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma wave instrument</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search coil</td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

The configuration of Messenger's orbit enables the spacecraft to approach the planet at 200 km and biases the domain of its investigations towards the northern hemisphere. MPO, on the other hand, does not approach the planet at altitudes less than 400 km, but covers 80% of the surface at an altitude lower than that of Messenger.

4.2. The Orbits

Fig. 9 illustrates the trajectories of the three orbiters:
- Messenger revolves around Mercury in 12 h. Its altitude varies between 400 and 15 193 km. The latitude of the perihelion is 60° and the orbit inclination 80°.
- MPO is dedicated to remote sensing and orbits on a polar, low altitude and low ellipticity trajectory, with a period of 2.3 h. The perihelion and apohem are located in the equatorial plane, at altitudes of 400 and 1 500 km, respectively.
- Like the orbit of MPO, that of MMO is polar, has a perihelion at an altitude of 400 km and a line of apsides in the equatorial plane. The eccentricity of the orbit enables the spacecraft to explore the magnetotail of the Mercury up to an altitude of 11 817 km, i.e. at a planetocentric distance of almost 6 R_M. The orbital periods of MPO and MMO are resonant in the ratio 4:1.

Fig. 9. The orbits of the Messenger and BepiColombo orbits (Inclinations: MPO and MMO, 90°; Messenger, 80°).

5. CONCLUSION

The BepiColombo and Messenger missions are complementary and mutually beneficial. They form together the most logical and best approach to the exploration of Mercury. Messenger, acting as a precursor, will identify targets of interest which will be explored in greater detail by BepiColombo. The Messenger findings will be particularly useful in planning the operations of BepiColombo, thus enabling the latter to make an optimal use of its global surface coverage, uniform spatial resolution, longer integration times and larger telemetry capability. Messenger could also play a critical role in the timely selection of a landing site for MSE, that minimizes landing risk while maximizing the scientific return from such a site.
THE DARWIN MISSION

Malcolm Fridlund

ESA Research and Scientific Support Department
ESTEC, P.O. Box 299, NL-2200AG Noordwijk, The Netherlands

Tel:+31-71-565-4768/Fax:+31-71-565-4690
e-mail: Malcolm.Fridlund@esa.int

ABSTRACT
The Infra Red Space Interferometer Darwin is an integral part of ESA’s Cosmic Vision 2020 plan, intended for a launch towards the middle of next decade. It has been the subject of a feasibility study and is now undergoing technological development. The scientific scope is aimed towards developing a system that could carry out the search for, and characterisation of Earth-like planets orbiting other stars. A secondary objective is to carry out imaging of astrophysical objects with unprecedented spatial resolution.

The implementation of Darwin is based on the new technique of ‘nulling interferometry‘, in the mid-IR and becomes the culmination of a decade of technology- and science precursor missions. Darwin is also foreseen to be carried out in an international context.

1. INTRODUCTION
The Infra Red Space Interferometer Darwin is a major element in the Cosmic Vision 2020 program of the European Space Agency. Darwin has the explicit purpose of detecting other Earth-like worlds, analyze their characteristics, determine the composition of their atmospheres and – investigate their capability to sustain life as we know it. As a secondary objective, it will also provide interferometric imaging of astrophysical objects in the wavelength range of the thermal infrared with unprecedented resolution. The closing years of the 20th century have allowed us, for the first time, to seriously discuss interferometric instruments deployed in space. With the express purpose of achieving unprecedented spatial resolution, these missions will lead to new astrophysics. Especially – and this is the greatest challenge – we expect to be able to carry out the first detailed study of terrestrial exoplanets (defined as planets similar to our own Earth as what concerns size and mass, and orbiting other stars than our Sun). The detection and study of the latter promises to usher in a new era in science and will affect a broad spectrum of science and technology. We can now confidently expect the first results from space-based interferometers within 10 years. Sophisticated instruments will follow in short order.

The European Space Agency identified interferometry from space as an important topic already in the original Horizon 2000 plan in the early eighties. In this scheme, it was classified as a green dream, i.e. a topic considered to be worthy to pursue scientifically but technologically immature, and thus a candidate for the future stages of European space exploration. In this context, a number of conferences/workshops were held, the purposes of which were to explore and define the scientific cases in preparation for the next generation of missions. Although the general topic of high-resolution imaging was considered as the overall scientific driver for developing an interferometric capability in space, the search for terrestrial exoplanets did figure prominently already in this beginning phase.

When, in the beginning of the 1990s, an external survey committee made its recommendations for the extension of the Horizon 2000 program (H2000+), they responded to technology developments worldwide, by identifying interferometry from space as a cornerstone candidate for the new program.

Within the context of interferometry, three topics were identified for further study:

- Astrometry.
- The search for terrestrial exoplanets, including the characterization of their properties and atmospheres and the possible detection of biospheres through remote sensing.
- Astrophysical imaging at a spatial resolution 1 to 2 orders of magnitude better than that foreseen to be achievable with the Next Generation Space Telescope (NGST). This would lead to light being shed on topics as varied as the formation of stars and planetary systems and the observation of galaxies at very high values of z.
Of these topics, the first was considered to be (relatively) 'simpler' in implementation, and resulted in the GAIA proposal to the European Space Agency and a pre-phase A study has recently been carried out by ESA’s science directorate. As a result of this study, GAIA is, however, no longer an interferometric mission.

The Infra Red Space Interferometer Darwin was studied by the European Space Agency (ESA), at system level, between 1997 and 2000. The study, carried out by Alcatel space division in Cannes, France has focused on developing a system that could carry out two main scientific objectives:
- The detection and characterisation of Earth-like planets orbiting other stars.
- The imaging of astrophysical objects with unprecedented spatial resolution.

The most challenging of these objectives consist of the recording of infrared spectra of Terrestrial exo-planets that could detect signs of biological activity at distances up to 20 pc. In order to do this, the Darwin project is constructed around the new technique of 'nulling interferometry', which exploits the wave nature of light to extinguish the light from a bright object (the central star in this case). At the same time the light from a nearby source (the planet) is enhanced. The contrast between planets and stars being the least in the infrared wavelength region, that has been chosen for this mission. The result of the study was presented to the community in September of 2000 in Paris, France.

A more or less identical activity has been taking place in the United States within the context of NASA’s ORIGINS program. This program has as its goal to answer questions like:
- How did the first galaxies form?
- How do stars and planetary systems form?
- Are there any planets outside our solar system that are capable of sustaining life?
- How did life originate on Earth?
- Is there life (however primitive or evolved) outside our solar system?

It can be clearly seen that the priorities for the two agencies are rather similar and form a solid base for collaboration on a number of programs.

2. EXO-PLANETS
The last 7 years have seen the detection of planets beyond our solar system finally becoming a fait accompli. The technique utilized so far has been indirect. It relies on measuring the reflex motion of the parent star, with respect to the common center of mass of the star-planet system. The planet causes the star to undergo a reflex motion around the star-planet center-of-mass, which leads to a perturbation of three (in principle) measurable variables:
- The radial velocity (a small periodic Doppler shift in the radial velocity of the absorption lines in the stellar spectrum which need to be measured over a relatively long period of time – typically at least three times the planetary orbital period).
- The astrometric position.
- The arrival times of electromagnetic radiation.

The last measurable was detected first. In the timing of radio signals from a number of pulsars, several Earth-size planets have been detected. These objects have presumably been formed out of the debris of the supernova explosion generating the pulsar. They are located at large distances, and the study of them will very likely not shed much light on the principles guiding the origin and evolution of the Earth.

The first detection of a planet outside the Solar System and using the radial velocity method was reported by Mayor & Queloz (1995) for the Solar type star 51 Peg. This was quickly followed by the Lick group who reported planets around 70 Vir and 47 Uma (Marcy & Butler, 1996; Butler & Marcy, 1996). Several groups have now altogether put 99 planets (as of July 1 2002) into the catalogue of known exo-planets. About 3000 Solar type stars are currently being monitored by one or another of about ten different programs initiated after the first successes.

It appears unlikely, however, that this method could be used to infer the presence of Earth-type planets. The current precision in radial velocity is ~ 3 ms\(^{-1}\), while the Earth at 1 AU from the Sun would require 0.1 ms\(^{-1}\). Acoustic pressure-mode oscillations in solar-type stars have amplitudes of 0.5 – 1 ms\(^{-1}\), and will thus make it essentially impossible to detect a deflection of an Earth simile.

The planets so far found with the radial velocity method are objects more akin to the planet Jupiter than something like our own Earth. This is of course due to the selection effect that biases searches for the objects causing the greatest radial velocity deflection which thus have large masses and short orbital periods. Consequently, as time passes, longer and longer periods are being picked up, and there are now a number of confirmed planets in orbits with periods of several years. A few of the objects are most likely more like so-called Brown Dwarfs. The radial velocity method only provides us with the minimum mass and the orbital radius, and there is no way to determine the

© European Space Agency • Provided by the NASA Astrophysics Data System
planets absolute mass, actual size or composition – without other data. Two planets orbiting the star HD168443 have, however, minimum masses of 7 & 17 $M_{\text{Jup}}$ respectively, and the bigger one has thus passed already the limit for what is usually classified as a brown dwarf ($-12 M_{\text{Jup}}$). Thus, a question raised sometimes concern the issue if the detected objects actually could be stars (brown dwarfs) in orbits seen face on (and consequently significantly more massive). A number of facts speak against such an hypothesis:

- A plot of the number of detected objects as a function of $M_{\sin(i)}$ increases exponentially towards $M_{\sin(i)} < 1 M_{\text{Jup}}$. An unknown process would have to bias our detection towards only picking up systems seen face on.

- One planet, HD209458b, has been found to transit its star (see below). This object is definitely in an edge on orbit and its mass can be shown to be below that of $1 M_{\text{Jup}}$. It is thus a bona fide planet.

In the case of the observed occultation, the planet orbits the star HD 209458 every 3.52 days at a distance of about 0.05 AU. The occultation lasts about 2.5 hours and from this observation, the inclination is found to be 87.1°. Charbonneau et al (2000) derive a planetary mass of 0.63 $M_{\text{Jup}}$ and a planetary radius of 1.27 $R_{\text{Jup}}$. This can be done since the orbital radius is well known from the radial velocity measurements, and the stellar radius is known to good accuracy from stellar evolution theory. The actual shape of the light curve during the occultation (Mazeh et al. 2000), constrain the planetary radius and orbital inclination to a very high degree. The average density of the planet turns out to be only half of that of the major gaseous giant planets in our own Solar System immediately ruling out the possibility of a rocky, terrestrial body since such a planet would be significantly smaller than 1.27 $R_{\text{Jup}}$. The planet is thus a gas giant. Being physically larger than Jupiter but with a lower mass, is caused by its proximity to its primary which heats it to a surface temperature of 1200 K. Such temperatures would, however, only affect the outer 1% of the planet, and the large diameter immediately says something about its evolution. As pointed out by Lunine (2001), what is happening is that the flux from the star retards the cooling of the planetary interior. A giant planet formed in isolation would cool in a brief time ($\sim 10^6$ yr), and thus it also shrinks rapidly from its original distended state. For a planet in very close proximity to a star such as is the case for HD209458b, the atmospheric temperature profile is flattened and the rate by which heat can be transported outwards from the interior is reduced and the contraction will be retarded. Detailed models (Burrows et al. 2000) show excellent agreement with the planetary radius at its current age of 4 – 7 billion years (the age being determined from stellar evolution theory). It can also be shown in these models, that the planet must have arrived at an orbital radius of $\sim 0.05$ AU within at most a few tens of millions of years after formation. Otherwise it would take longer than the present age of the Universe for the external heat to diffuse inwards far enough to expand the radius to the observed value. The observation of a single occultation have thus shown us that the so called 'hot Jupiters' either form in place or migrate inwards within at most $\sim$ a few $x 10^7$ years (Lunine 2001; Burrows et al 2000).

Another indirect method, is to obtain astrometric data and thus track a star's path across the sky, measuring the wobble introduced by the rotation around the common center of mass of the star-planet system. The European Space Agency's GAIA mission promise large statistical surveys of massive planets. For further information on exo-planets we refer to the excellent article by Perryman (2000).

3. DIRECT DETECTION OF TERRESTRIAL EXO-PLANETS

All of the above-mentioned methods will continue to refine our knowledge about planetary systems. Unfortunately they are restricted towards the indirect detection of relatively massive planets. Space missions (e.g. COROT, Eddington, Kepler) designed for studying occultations will eventually pick up lower (Earth-size) mass bodies, but relatively little information will be gathered in this fashion. The most important datum will be the frequency of Earth-like planets, which will allow for a proper design of direct detection experiments such as Darwin. The above-mentioned missions are thus to be considered as true precursors to cornerstone level missions like Darwin.

The main problem involved in the direct detection of an exo-planet of a size comparable to our own Earth – and located at a similar distance from its own star – is mainly one involving contrast and dynamical range. The central star (the Sun, the primary) outshines the planet in the visual wavelength range by a factor of at least $10^5$. This problem is alleviated by going to the mid-infrared where the planets thermal emission peaks (a terrestrial planet is in this case defined as one of roughly the same size as the Earth and of the same surface temperature – ca 270 K – thus having its peak emission at 10 $\mu$m). Even at these wavelengths, the contrast is more than a factor of $10^5$. The star and planet will be very near each other on the sky, and we need to devise a way of extinguishing the light from the star. Different coronographic methods (in space) have been evaluated, and albeit having the capability of achieving the scientific objective of detecting the exoplanet, these methods do not lend themselves to a
large enough search space (unless the telescope is extremely large). Lately it has been suggested to fly coronographic systems operating in the visual wavelength range in space. A 5 m to a 10 m diameter monolithic telescope could then suffice. A 10 m class coronograph operating in the visual could detect and study the Earth around 100 – 150 of the closest Solar type stars. As a contrast, the Darwin model mission nulling interferometer described below, operating in the mid-infrared could detect and study the Earth in orbit around more than 500 stars in the Solar neighborhood.

Two other methods that have been suggested are ‘nulling interferometry’ and ‘densified pupil’ techniques. Of these, the ‘nulling interferometry’ technique was selected for the Darwin study because of its relative simplicity and maturity (at the time).

3.1 Nulling Interferometry

Nulling interferometry can simply be described by considering two apertures, separated by a baseline D. One now point the two telescopes towards the same star, and connects the light output of the two units. If one then has the optical path lengths of both apertures the same, the amplitudes of the electromagnetic radiation will interfere. This is interferometry in the classical sense producing a set of dark and bright bands – so called fringes. If we instead now make the light from one of the telescopes arrive at the site of beam combination with an added phase shift, π, the light along the optical axis will instead interfere destructively (the dark fringe appear ‘on top’ of the star). At the same time, waves arriving from a small angle, θ, away will interfere constructively. This separation will depend on the distance between the two telescopes. If we now assume that we have a star, which is orbited by a planet located at an angle θ away, we can extinguish the light from the primary and isolate the planetary light. The contrast between the star and planet is now restricted to light leaks from the ‘central null’ – due to imperfections in the optics and jitter of mechanical components. By using more telescopes it is possible to create a more complex transmission pattern. The actual pattern depends on the number of apertures, and the geometrical configuration. In the Darwin configuration we have 6 telescopes in a hexagonal pattern, and with all telescopes equidistant from a central beam combiner. Then the pattern is roughly doughnut shaped. It is now possible to ‘tune’ the array to each individual star that is observed, such that the transmission ring is located on top of the so called habitable zone (see below). The signal also need to be modulated, in order to separate out an eventual planetary signal from any background (such as exo-zodiacal light – dust in the target system radiating as a blackbody at ~ 300 K) or to discriminate between different combinations of planets in the observed system. This can be performed either by the rotation of the array of telescopes, switching between different combinations of apertures or by a combination of both. The latter option is chosen as the baseline for Darwin.

4. BIOMARKERS

A major goal of the mission is not only to detect terrestrial exoplanets, but also to investigate if the conditions on the planet in question would allow life as we know it to exist and indeed if it already exist. In order to do so we need to define what life is, and how life as we understand it interact with its environment in an observable way. Life contains information; Life is self-replicating; Life evolves; and Life influences its environment. We are attempting to detect these attributes of life, through remote sensing, at interstellar distances. At first this goal seems hopeless. It has been found, however, that the simultaneous detection of Water – H₂O – and molecular Oxygen and at a temperature of about 300 K is a clear indication of life as we know it. This is because oxygen is one of the most reactive substances there is. If all life on the Earth were removed suddenly – bacteria, green plants etc – all of the free oxygen in the Earth’s atmosphere would disappear in the geologically short time of 4 million years. The atmosphere of the Earth is out of equilibrium as is evidenced by a comparison with models or with the situation in the other terrestrial planets in the Solar system – Mars and Venus. This dis-equilibrium is caused by the living things on our planet. Previous to life being dominated by Oxygen
generating species, the atmosphere of the early Earth was out of equilibrium by Methane, CH₄.

The criterion on temperature will define a habitable zone around each individual star. Strictly speaking, the surface temperature of a planet will depend not only on the energy input from the primary, but also on the atmospheric pressure and composition. Our own Earth, for instance, would be significantly colder without its greenhouse effect, caused by CO₂ and CH₄. Since we a priori have no idea about the presence and/or composition of any eventual atmospheres, we will have to use this criterion with some care in the individual case.

We mentioned above that we wish to observe in the mid-IR in order to make the detection problem more tractable because of a lessened contrast. Within the wavelength range of 5 μm to 20 μm we find important absorption features of Water, Carbon dioxide and Ozone. Ozone has been shown to be a good tracer of Oxygen since its absorption feature in the spectrum show a logarithmic dependence on the abundance of molecular Oxygen. The downside of using Ozone as a tracer is that it is easily saturated. Very early in the history of life on our planet the O₂ saturated, and since then it has been possible to – remotely – determine the level of biological activity on Earth. The actual content of O₂ during the first half of the history of living things on our planet was significantly lower than it is today.

With a spectral resolution, \( \Delta \lambda / \lambda = 20 - 40 \), we can detect all of the features mentioned above. We have to remember, that this refers to our Earth as it is today. The oxygen level in our atmosphere have been relatively high only for the last 20–30 % of the Earth’s history. Although, it is now more or less generally accepted that life arose on the Earth immediately after the era of bombardment, i.e. 3.8 x 10⁹ years ago, it remained in the sea, and at a relatively simple level until just about 600 million years ago. This was the time of the so-called Cambrian explosion, when the phyla we can see today had their origins. Life until then was dominated by Methane producing species and if we want to define our ‘remote sensing’ criterion such that life is indicated by a disturbing of the equilibrium of a terrestrial planets atmosphere, we need to also take this evolutionary aspect into account.

Figure 2: Cross-section of one of the telescope flyers showing the optical train and the thermal structure

5. THE MODEL MISSION
The current mission scenario is to be taken as a model that can fulfill the stated scientific objectives, but which does not include aspects related to international collaboration with e.g. NASA (see below). It consist of six 1.5 m telescopes, which transmit their input beams to a beam combiner satellite. The individual telescopes – each mounted on a separate spacecraft – are kept to within the required precision of their relative positions by utilizing a metrology system, which includes laser metrology, radio frequency goniometry, and the tracking of interferometric fringes from a bright guide star. In the case of the planet finding part of the mission this latter requirement is not going to be a problem, since the search for exo-planets will be carried out around relatively nearby – and thus relatively bright – stars. The primary in each system can thus be used as a guide star for both pointing and fringe tracking for the adjustment of positions. Final adjustment of path lengths and the required phase delay is then introduced in the beam combiner satellite, where detection is also carried out. The system operates between wavelengths of \( \approx 5 \mu m \) and \( \approx 20 \mu m \), and require passive cooling of all optical components to below 40 K in order to reduce the thermal background. Because of this latter requirement on temperature, all power generating functions (computing, transmitting of data to Earth, laser metrology) is carried out from a non-cooled communications satellite. All 8 spacecraft are deployed and kept flying in formation in an orbit around the L2 Sun - Earth Lagrangian point. Launch is foreseen with an Ariane-5, and the Darwin parameters are well within the envelope of that launcher. In the planet finding mode, Darwin utilizes baselines between 40 and 250 m.

A separate beam combiner table on the central spacecraft is used for ‘normal’ Michelson interferometry in an imaging mode. In imaging mode, baselines up to 1 km are foreseen. In this mode, objects
such as star forming regions, Active Galactic Nuclei, the core of our own Galaxy, Black Holes and very early galaxy formation, will be studied at unprecedented spatial resolution.

6. TECHNOLOGY DEVELOPMENT AND IMPLEMENTATION

An ambitious technology program has been initiated in order to develop the required items such as the nulling interferometry, metrology, and formation flying. Development of the required optical components such as polarisers, beam splitters and achromatic phase shifters are also being initiated.

The following is a summary of the current elements in the Darwin technology program:

**Metrology:**

1) RF ranging – radio frequency position system used for deployment, coarse acquisition
2) Laser metrology – high precision optical metrology (HPOM) for fine acquisition
3) Fringe sensor – interferometric fringe tracking for the control loop of the Attitude and Orbital Control System

**Actuators:**

1) FEEP – Field Emission Electric Propulsion – μN and mN thrusters with μN(Hz^1/2) noise
2) μN cold-gas thrusters

**Control software:**

1) Deployment
2) Optical Path Difference Control (using fringe sensor and/or HPOM)
3) Formation flying command and control

**Optical Components:**

1) Achromatic phase-shifter
2) Optical delay line (SMART-2 verification)
3) Wavefront filter (monomode fibre)
4) Amplitude / Polarization matching units
5) Integrated optics at 5 – 18 μm wavelength

Further programs have been initiated in the area of detectors, coolers, interferometric configurations, beam combination and data post processing.

Several precursor missions and projects are planned in order to implement Darwin in a timely and efficient manner. The SMART-2 mission is planned for 2006. Being primarily intended as a test flight for the LISA gravitational wave detector mission, it nevertheless lend itself efficiently for a formation flying qualification flight, since it need a two component space craft anyhow. On this flight, software and hardware required for formation flying, as well as the metrology is planned for test. Accordingly:

1) Formation flying
   a. Deployment
   b. Collision avoidance, FDIR
   c. Radio Frequency Metrology
2) Precision formation flying
   a. Along one axis (only two spacecraft)
   b. Accuracy (about 1 micron)
   c. High precision optical – RF metrology
3) Optical Delay line (ODL)
   a. ODL used to find and track fringes
   b. Micron propulsion used to find fringes
4) Trade-off between Field-Emission Electric Propulsion (FEEP) and micron cold-gas propulsion systems

Figure 3: Artist's view of SMART-2

© European Space Agency • Provided by the NASA Astrophysics Data System
Nulling interferometry itself will first be tested on the ground. Currently, the objective of two industrial breadboard designs, it is intended to carry out the Ground based European Nulling Interferometry Experiment (GENIE) in 2006, based on the laboratory work and in collaboration between ESA and the European Southern Observatory (ESO).

The rationale for GENIE is the following:

1) Gain technological experience
2) Precursor science:
   a. Exo-zodiacal disks
   b. Study target systems.
3) Unique and valuable science:
   a. AGN’s, star formation discs, brown dwarfs, etc
   b. Spectroscopy of ‘hot Jupiters’?

A final consideration is to provide European scientists with a nulling IR interferometer.

Another test flight (SMART-3) could be flown in 2009. This mission will be the objective of an assessment study during 2003, but could consist of a free-flying two-telescope interferometer with a separate beam combiner. The important element of SMART-3 is to carry out a representative test of the complete Darwin mission (e.g. wide-band nulling interferometry at representative wavelengths, avionics, hardware elements). The Darwin mission could itself then be following in 2014.

Darwin, because of its ambitious scope and technology requirements lead itself very well to international collaboration. As mentioned above, NASA have a similar mission as the goal of the ORIGINS program. Scientific collaboration and discussions have already been initiated with this NASA’s Terrestrial Planet Finder (TPF) with the ultimate goal of attempting a joint Darwin-TPF mission. During the next 3 years leading up to 2006, both TPF and Darwin will be carrying out scientific and technological studies and development work – separately, but with joint membership on the science teams. Both the European and the NASA teams will be tasked to arrive at a final architecture could be implemented in a joint scenario.

7. CONCLUSION

The Darwin model mission represent today a design that could answer one of the longest lasting scientific questions – that of if we are alone in the Universe. The feasibility study has demonstrated the maturity of the existing technology, as well as providing a development plan that could improve the efficiency of the mission. It is also the current conclusion that to carry out the scientific plan as defined above, free-

Figure 4: Artist’s impression of the Darwin array orbiting at L2

8. REFERENCES

Lunine J.I., 2001, PNAS, 98, 809
Mayor M., Queloz D., 1995, Nature 378, 355
Session 2
Earth as a Planet
Chair: E.A. Herland
EARTH’S IONOSPHERE AND MAGNETOSPHERE

Harri Laakso
ESA Space Science Department, Noordwijk, The Netherlands

ABSTRACT

This paper gives a brief overview of some of the key areas in the Earth’s ionosphere and magnetosphere. Although we know these regions and their connections to the solar radiation and solar wind activities already quite well, there are still plenty of unresolved problems. The fundamental goal of the solar-terrestrial physics is to predict space weather around the Earth and to understand how the sun and its variability affect life and human technology/society on the Earth. In addition knowledge achieved in detailed Sun-Earth connection studies are useful for a variety of problems related to the other solar system bodies and astrophysical objects because many of the complex plasma processes that are common in the universe can be investigated in great detail in the near-Earth space.

1. INTRODUCTION

The Sun may seem to be far from the Earth, but actually the Earth and the other solar system bodies are immersed in the extended atmosphere of the Sun. The planetary atmospheres are constantly forced by the solar radiation, the solar wind and a variety of solar disturbances, such as coronal mass ejections (CMEs), solar flares and fast streams of the coronal holes [Song et al., 2001].

1.1 Solar wind and interplanetary magnetic field

The solar wind is a primary source of energy and momentum to the Earth’s magnetosphere and ionosphere. A particularly important variable is the interplanetary magnetic field (IMF) which is drawn from the sun by the expanding solar wind. A key factor for the terrestrial plasma environment appears to be the sign of the IMF Bz component (north-south). When the IMF is pointing southward, i.e. IMF Bz < 0, the solar wind mass, energy, and momentum are effectively transported into the magnetosphere, causing strong magnetic disturbances around the Earth (see section 3.7).

An important property of the collisionless plasmas, such as the solar wind, is that the charged particles, electrons and ions, are bound to a specific magnetic field line or flux tube (a bundle of field lines). Consequently, wherever the cold plasma moves, it will transport magnetic field lines with it. In the case of the solar wind, the field lines are attached to the sun, and therefore they form a spiral field line structure in the heliosphere due to the solar rotation and the radial motion of the solar wind.

1.2 Solar Disturbances

After the launches of the SOHO (Solar & Heliospheric Observatory; see http://sci.esa.int/home/soho/) and Ulysses (http://sci.esa.int/home/ulysses/) satellites, our knowledge about the structure and dynamics of the Sun has significantly improved. In particular, SOHO continuously monitors the Sun and provides us with important information about solar disturbances while they are heading towards the Earth.

Figure 1. CME detected by the SOHO satellite. CMEs are the main cause of magnetic storms around the Earth and other planets. The front of the CME is a shock that accelerates protons to relativistic energies.

Figure 1 presents an example of a coronal mass ejection detected by SOHO. After leaving the solar atmosphere, it expands and can cover 180 degree of the Earth’s orbit. Therefore the probability that it hits a planet like the Earth is large. The front of the CME structure is an interplanetary shock that continuously accelerates particles up to relativistic energies [Reames, 1999]. They are one of the main sources of Solar Proton Events that are an im-
important issue in space weather, particularly for astronaut safety and satellite operations.

1.3 Sun-Earth connection

Figure 2 summarizes the main elements in the Sun-Earth connection (SEC) chain. The emphasis of the solar-terrestrial physics is to study the various links in this chain. It is clear that monitoring the solar and solar wind conditions plays a key role when the dynamics and structure of the magnetosphere and ionosphere are investigated. The NASA ACE satellite (http://helios.gsfc.nasa.gov/ace/) monitors the solar wind conditions at the L1 point and can give us an early (some 20-60 minutes) warnings of coming CMEs before they hit the Earth’s

![Sun-Earth connection chain](image)

Figure 2. Sun-Earth connection chain.

The ionosphere and magnetosphere respond differently to the solar energy inputs. The solar wind is the driver of the magnetospheric dynamics whereas the solar radiation is the major energy source of the atmosphere-ionosphere system. However, in the nightside atmosphere, particle precipitation from the magnetosphere, dissipation of electric currents in the ionosphere, and magnetospheric electric fields driven by the solar wind make all fundamental impacts on the ionospheric structure and dynamics. The energy inputs from these sources are deposited mainly in the high-latitude ionosphere and can result in strong disturbances; for example, in the thermosphere neutral winds driven by fast plasma drifts can well exceed winds driven by the normal solar radiation pressure.

2. IONOSPHERE

The ionosphere, the ionized part of the upper atmosphere, was discovered in the turn of the 20th century, by means of radio wave studies [Chamberlain and Hunten, 1987]. The conducting ionosphere reflects radio signals and allows us to transmit radio signals beyond the horizon.

The monitoring of the ionosphere is important as the ionosphere disturbances can cause detrimental effects on our technologies. For instance, the ionosphere can support very strong electric currents that can further induce damaging electric currents in ground-based technological systems [Pirjola, 2002]. The high-quality, phase-sensitive radio transmission between satellites and ground, needed e.g. by the GPS system, is necessary but is occasionally strongly disturbed by ionospheric irregularities [Bhattacharyya and Basu, 2002].

The purpose of this section is to give a quick overview to the structure and dynamics of the ionosphere that lies at 60-1000 km altitude range. The ionosphere is highly variable both temporally and spatially. Although we have quite a few ionospheric models available, none of them can accurately predict the ionospheric characteristics [Staschewicz, 1995]. A significant improvement here would require a large increase in ionospheric observations.

The ionosphere is monitored from ground, using radars, magnetometers and optical instruments, but also using space-borne instrumentation on rockets and low-altitude satellites. Notice that the low-altitude part of the ionosphere, below 200 km altitude, cannot be accessed with satellites due to strong atmospheric drag.

2.1 Ionospheric formation and structure

The ionosphere is formed by ionizing solar UV and X-ray radiation. Near 60-70 km altitude, the atmospheric density has decreased to the level where the atmosphere can become partially ionized when an ionizing source exists. Figure 3 shows a typical altitude variation of electron density together with the penetration depths of different solar photons. Due to the altitude variation of the thermosphere’s composition and the absorption characteristics of the atmosphere, the ionosphere appears as a number of layers or regions, called D-, E-, and F-region.

![Altitude profile of electron density in the dayside ionosphere](image)

Figure 3. Altitude profile of electron density in the dayside ionosphere. The main ionizing sources of the atmosphere are solar UV- and X-ray photons [courtesy: J.H. Yee, APL, Johns Hopkins University].
The D-region exists approximately at 60-90 km altitude, is wholly of solar origin, and is caused mainly by X-rays but also by Ly-α that ionizes NO molecules. The region exists only during daytime and disappears immediately after the sunset. A typical density is in the range 10^4-10^5 cm^-3, and the main ions are NO^+, water cluster ions, negative ions (O^-; O_2^-). This region has a complicated chemistry and is least understood because of observational difficulties to explore this region (satellites cannot orbit at this low altitude, and the D region reflects radar signals very poorly). However, the D region affects radio transmission because radio waves must cross the region twice while reflecting from the higher regions so its better understanding is highly desirable.

The E region, which appears between 90-140 km altitude and has a peak density near 110 km, is the key reflector of radio signals. This layer is caused mainly by EUV radiation below 110 nm, and the major ions are O^+ and NO^+ (see Figure 4). A typical density is 10^5 cm^-3. In the nightside electron bombardment and in all local time sectors meteor bombarding can locally and temporarily cause significant increases in the E-region density; these structures are called sporadic E. The meteor bombardment can deposit a large amount of metal (Fe, Mg, Ca) atoms/molecules and ions in the E region.

![Figure 5. Precipitation of electrons into the atmosphere.](image)

The F region fills the largest volume of the ionosphere. It is divided into two parts: F1- and F2-region, where F1 exists only in the daytime. The F1 layer, located between 140-210 km altitude, is caused by Ly-α and EUV radiation. The main ions are O^+ and NO^+. The ionization in the F2 layer, located above 210 km, is also caused by Ly-α photons, but this region appears due to decreasing recombination processes rather than due to a peak in ionization rate. Here the main ion is O^+ but also some N^+ ions exist (Figure 4). This region has the highest density, usually of the order of 10^6 cm^-3, exists at all local times, and its altitude has a significant diurnal variation, rising above 400 km in the nightside sector.

2.2 Particle precipitation and auroral emission

Electron bombardment is an important source of ionization particularly in the nightside ionosphere at auroral latitudes, where large fluxes of energetic (1-10 keV) electrons stream from the magnetosphere into the atmosphere, causing significant ionization particularly in the E region. The low-energy electrons ionize the molecules at higher altitudes, in the F region. The resulting plasma densities can often be an order of magnitude higher than those caused by photoionization on the dayside. Electron precipitation also causes bremsstrahlung X-ray emission, molecule dissociation, and atom/molecule excitation, where the last processes are the cause of the auroral emission.

![Figure 4. Example of daytime ionospheric and atmospheric composition based on mass spectrometer measurements [Johnson, 1969].](image)

![Figure 6. Global auroral oval images taken by Dynamic Explorer I (courtesy: L. Frank, University of Iowa)](image)
2.3 Ionospheric conductivities

The electric conductivity in a collisional magnetized plasma, e.g., in the ionosphere, is a complex parameter whose magnitude is a function of many variables and has a directional dependence. The conductivity along the magnetic field $B$ is usually very large except at low altitudes (near or below 100 km) (see Figure 7). The conductivity in the perpendicular plane to the magnetic field is defined by two terms, Pedersen conductivity parallel to the electric field $E$ and Hall conductivity parallel to $E \times B$ direction (for more detail, see e.g. Kelley [1989]). Figure 7 shows typical values for these conductivities in the nightside and dayside under moderate magnetic activities. Real values can differ quite significantly from these profiles. One can see that in the E region the strong currents are primarily Hall currents whereas in the F region the currents are mainly Pedersen currents. Figure 8 shows model calculations for the Pedersen and Hall conductivity in the E region. The Hall conductivity in the auroral oval is comparable to the dayside conductivity.

![Figure 7. Typical altitude profiles of the conductivities for moderate magnetic activity levels on the dayside (larger values) and nightside.](image)

![Figure 8. Pedersen and Hall conductivity in the E region [courtesy: Weimer, Mission Research Corporation].](image)

2.4 Ionospheric currents

The electric currents in the ionosphere can be very large. Solar radiation pressure produces thermospheric neutral winds that drive (via neutral-ion/electron collisions) the dayside electric current system (see Figure 9). The strongest current in the system occurs at the equator near noon; this is called the equatorial electrojet that can be several 100 kA.

![Figure 9. Major electric current systems in the ionosphere.](image)

The largest ionospheric electric currents, however, flow in the auroral region on both hemispheres. During magnetic storms these auroral electrojets can rise to very high levels (several millions of Amperes) and can cause severe disturbances on the ground-based power lines and pipelines [Pirjola, 2002]. There are two electrojets, the westward electrojet on the postmidnight sector and the eastward electrojet on the premidnight sector. These electrojets are driven by the magnetospheric processes via the field-aligned current system. Part of the electrojet energy is deposited into the ionosphere via Joule heating and this energy drives fast plasma drifts that cause fast neutral winds via ion-neutral collisions.

The ionosphere-thermosphere coupling is an important, but still largely unknown issue. In the dayside there are good examples of how neutral winds of the thermosphere drive the motions of ions and electrons of the ionosphere. On the other hand, there are good examples of the opposite processes where strong energy inputs from the magnetosphere drive fast neutral winds at velocities of up to 1000 m/s [Zhang and Shepherd, 2000].

2.5 Ionospheric irregularities

The ionosphere appears as a highly irregular region, particularly in the auroral region, due to large magnetospheric energy inputs. Also, the equatorial ionosphere is known from its large disturbances for over six decades [Booker and Wells, 1938]. The equatorial $F$ region often evolves from a smooth one-dimensional function of alti-
tude in the daytime to a disturbed region of inhomogene-
ities at a number of scale lengths on the nightside. This
phenomenon, called spread-F or equatorial bubbles (Fig-
ure 10), has been extensively studied with the aid of
ground-based radars, rocket-borne probes, and satellite
probes [Laakso et al., 1997, and references therein]. The
events are strongest in the pre-midnight sector, but they
are frequently detected also in the post-midnight sector
[Palmroth et al., 2000].

![Figure 10. Equatorial spread-F detected by the Ji-
camarca radar. The upwelling structures cause scintilla-
tion and are detrimental to phase-sensitive signals.](image)

High-frequency radio waves propagate through these
irregularities, but unfortunately not without some effects.
Small-scale irregularities cause random amplitude fading
and changes in phases. This is called scintillation that can
seriously disturb communication between the satellites
and ground, and that is particularly detrimental to phase-
sensitive signals (e.g., GPS system). At high latitudes,
magnetic storms can last for a few days during which the
communication system may suffer from partial or total
blackout for several hours. On the other hand, in the
equatorial region, the occurrence of spread-F irregulari-
ties that is well documented and often observed, cannot
yet be predicted. To improve the situation here, the US
Air Force is going to launch its C/NOFS satellite in 2003
in order to collect more observations on this phenomenon
and therefore to improve forecasts on it.

3. MAGNETOSPHERE

The magnetosphere is defined as the region where the
Earth’s magnetic field controls the motion of charged
particles. The structure and dynamics of the magneto-
sphere have been widely investigated with in-situ instru-
mentation on numerous satellites for forty years [Stern,
1995].

The Earth’s internal magnetic field is approximately a
dipolar field, but due to solar wind effects it is deformed,
into a comet-like structure with a long magnetotail that
can extend to several 100 Re (Figure 11). On the dayside,
due to the solar wind pressure, the magnetosphere ex-
tends only to about 10 Re distance.

![Figure 11. Main regions of the Earth’s magnetosphere.](image)

3.1 Bow shock and magnetosheath

The solar wind is a supersonic (~400 km/s) stream, and
when an obstacle is placed in the stream, a steady shock
wave is formed in front of the obstacle [Slavin and Hol-
zer, 1981]. In collisionless plasmas, the reference speed
for shocks is not the sound speed but the Alfvén speed
that is defined as \( \frac{B}{\sqrt{\mu_0}} \), where \( B \) is the magnetic field
and \( \rho \) is the mass density of the plasma.

At the Earth, the average bow shock location is about 14
Re at the subsolar point, but this distance is highly vari-
able depending primarily upon the distance of the mag-
etopause (see section 3.2). Behind the bow shock, the
solar wind streams at a subsonic speed; this region is
called the magnetosheath.

At the bow shock the solar wind plasma is heated and
decelerated before it is diverted around the magneto-
sphere. The bow shock is also a source of particle accel-
eration and wave generation. Characteristics of the shock
depends strongly on the orientation of the IMF [Horbury
et al., 2002]. The bow shock and interplanetary shocks in
general are one of the main areas of the new Cluster
mission (see http://sci.esa.int/home/cluster/).

The shocks are believed to be a common and important
phenomenon in the collisionless plasmas of the universe,
and therefore a good understanding of the Earth’s bow
shock helps us explain similar astrophysical situations.

3.2 Magnetopause and cusp

The magnetopause is the outer boundary of the magneto-
sphere, and this boundary is formed at a location where
the magnetosheath dynamic pressure equals to the Earth's magnetic field pressure. Therefore the Earth's magnetic field is somewhat compressed on the dayside. The average magnetopause location is about 10 Re at the subsolar point and about 15 Re at the dawn-dusk meridian. During high solar wind pressure, the magnetopause can move close to the Earth so that geostationary satellites (at 6.6 Re) on the dayside can move into the interplanetary space.

In the magnetosphere, characteristics of the plasma environment are quite different from that in the magnetosheath, and particularly the magnetic field orientation and plasma density can change drastically at the magnetopause. In the solar wind a typical plasma density at 1 AU is 5-10 cm\(^{-3}\), in the magnetosheath it increases often by a factor of 10, but in the magnetosphere near the magnetopause the density is only in the 0.01-1 cm\(^{-3}\) range. Therefore, the magnetosphere can be regarded as a magnetic bubble, with a thin and sharp boundary, in the dense solar wind.

When magnetosheath particles hit the magnetopause, they gyrate quickly in the boundary before returning into the magnetosheath. As the direction of gyromotion is different for electrons and ions, an electric current system is generated on the magnetopause (Figure 12). This is one of the main electric current systems of the magnetosphere and is part of the magnetotail current system that maintains the structure of the long magnetotail (see section 3.3).

![Figure 12. Solar wind electrons and ions gyrate in the magnetopause and cause an electric current system to flow from dawn to dusk.](image)

Reconnection is one of main energy conversion processes in the universe. It happens in the Sun's atmosphere, causing flares, and in many astrophysical phenomena (see the proceedings of Reconnection in Space Plasmas, ESA SP-285, ESTEC, Noordwijk, 1989), but it also takes place in the magnetosphere. In the magnetotail reconnec-
tion process, the stored magnetic energy is transferred into particle energy, which results in fast plasma jets along magnetic field lines into the polar ionosphere and into the downstream tail (see Figure 15) [Oieroset et al., 2001].

![Diagram of magnetic reconnection in the magnetotail](image1)

**Figure 15.** Magnetic reconnection in the magnetotail, as detected by the Wind satellite [Oieroset et al., 2001].

### 3.4 Inner magnetosphere

The inner magnetosphere (Figure 16) is filled by two plasma populations, the radiation belts and the plasmasphere, that do not interact much with each other.

![Diagram of inner magnetosphere](image2)

**Figure 16.** Inner magnetosphere

The plasmasphere, a doughnut-shaped region of the inner magnetosphere, is composed of cold (~1 eV) ions and electrons that have escaped from the ionosphere. Within the plasmasphere, the charged particles are frozen to the magnetic fields and are forced to corotate with the Earth. The plasma density decreases constantly from the ionosphere until a sudden density decline occurs at the plasmapause, which is the outer boundary of the plasmasphere. There is no boundary between the ionosphere and the plasmasphere; sometimes the boundary is defined to be a location where the O⁺ dominated plasma changes into He⁺ dominated plasma, which happens near 1000 km altitude. The plasmapause usually exists between 3-7 Re distance at the equator; the distance depends strongly on local time and geomagnetic activity in the magnetosphere [Laakso et al., 2002].

The radiation belts consist of energetic (several 100 keV - several MeV) electrons and ions that are trapped to the magnetic field lines. The dipole field lines form a magnetic bottle where the particles are confined for a long time; the decay time of such trapped particles is several days. The trapped particles do three basic motions: gyration around field lines, bouncing between the mirror points of the bottle, and drifting around the Earth (Figure 17). The drift motion has different directions for electrons and ions and causes therefore an electric current, called the ring current.

![Diagram of motion of energetic trapped particles](image3)

**Figure 17.** Motion of energetic trapped particles on dipole field lines.

Our knowledge about the dynamics of the radiation belt has significantly improved during the past decade because of long-term studies of the relativistic particles monitored by the SAMPEX and POLAR satellites [e.g., Baker et al., 1999; Kanekal et al., 1999]. Figure 18 shows an example of the variation of the relativistic (2–6 MeV) electrons in seven years. Large enhancements in the electron flux are common and appear frequently, with no relationship to a solar cycle. They occur during magnetic storms that, on the other hand, tend to happen quite constantly primarily as a consequence of CME and other solar disturbances.

### 3.5 Solar wind – magnetosphere coupling

One of the most important issues in space plasma physics is the interaction of two colliding plasma regions, and how energy, momentum and mass are transferred between the two regions. Such an interaction can be studied in detail at the Earth’s plasmapause. This is one the main issues for the new multi-point measurements by the Cluster mission.

As in the magnetotail, it is believed that reconnection is the main process that causes the transport of energy, momentum, and mass from the solar wind into the magnetosphere. It was proposed by Dungey already in 1961...
Figure 18. Relativistic (2-6MeV) electron flux measured by SAMPEX in years 1992-1999. Left axis shows the L shell of the satellite, bottom axis the number of days. [courtesy: S. Kanekal, LASP, Boulder].

Figure 19. Flow chart for magnetospheric/ionospheric coupling and production of aurora [Akasofu, 1991].

but only recently we have received strong observational support to this process [Phan et al., 2000; Mozer et al., 2002]. The main requirement for reconnection is the opposite directions of the IMF and the Earth’s magnetic field. Notice that magnetic disturbances occur when the IMF points southward (whereas the Earth’s magnetic field always points northward at the equator).

The magnetosphere and solar wind interact in many other ways, too. One of the most common phenomenon is the Kelvin-Helmholtz instability that can occur at an interface between two streaming plasmas [Southwood, 1979]. As a result, large-scale waves (cf. waves on the surface of water) are triggered on the magnetopause, sending compressional waves towards the Earth and causing a number of phenomena [Laakso et al., 1998].

An important consequence of the solar wind – magnetosphere interaction is the existence of a large-scale flow pattern of charged particles in the magnetosphere which is driven by the electric field caused by that interaction. These electric fields map into the ionosphere along magnetic field lines and drive the motion of charged particles in the ionosphere and the motion of neutral particles via collisions (see section 3.6).

3.6 Magnetosphere – ionosphere coupling

The understanding of the M-I coupling is important as the conditions of the upper atmosphere are clearly affected by magnetospheric conditions. Figure 19 summarizes the main processes related to the generation of neutral winds and auroral emission. There are two fundamental magnetospheric energy inputs into the ionosphere-thermosphere system: particle precipitation and dissipation of ionospheric current systems. Both processes peak in the auroral oval and cause heating in the upper atmosphere. Due to strong electric and magnetic fields, plasmas tend to drift and at the same time force neutrals to move in the same direction. During strong magnetic activities, this motion can well exceed the normal thermospheric winds driven by the solar UV energy input into the dayside thermosphere.

Monitoring of the ionosphere help us understand magnetospheric conditions. Basically the whole magnetosphere is mapped along magnetic field lines into the ionosphere, and particle precipitation causing auroral emission carries information about magnetospheric particles and dynamics. Ground-based magnetic field observations are used to determine the auroral electrojets that are driven by field-aligned currents of the magnetosphere.

3.7 Storms and substorms

The magnetosphere-ionosphere system exhibits frequently a variety of disturbances during which strong electric cur-
rents and plasma flows, particle precipitation and energization of particles up to relativistic energies can occur.

Basically the main events are substorms and storms. They have some similarities but also quite a few different characteristics and origins. Substorms are ultimately driven by the southward turning of the IMF (dayside reconnection) that feeds energy into the magnetosphere, and approximately after half an hour the magnetotail is driven into an unstable state that results in the occurrence of substorm onset (via nightside reconnection). The most visible consequence of a substorm is a vivid auroral display. The whole sequence lasts for a few hours and can be repeated several times a day.

Figure 20 summaries the main energy transfer processes during a substorm sequence. The energy is fed into the system by the solar wind when the IMF is pointing southward. The energy is either stored in the magnetotail or is used for driving the convection of the magnetospheric particles. This convection pattern maps into the polar ionosphere where the energy is slowly dissipated (causing heating of the upper atmosphere). The storage phase of the energy input in the magnetotail ends when the tail becomes into an unstable situation and a magnetic reconnection occurs, resulting in explosive dissipation of energy in the ionosphere; part of the released tail energy returns to the solar wind [Baker et al., 1984].

![Figure 20. Main elements of the substorm sequence (Baker et al., 1984)](image)

Storms are primarily caused by CME’s [Richardson et al., 2002] and may reoccur at 27 day intervals if the source of CME is a coronal hole at the Sun’s equator. Although these events are also associated with strong auroral display, the major consequences are related to relativistic particle injections (detrimental to satellites and astronauts) and strong auroral electrojets that can be several millions of Amperes that can cause large current peaks in the ground-based technological systems.

Figure 21 summaries time scales and the sequence of events during a magnetic storm. The release of a CME is associated with strong X-ray and radio noise emission that affect directly the dayside ionosphere. The first, nearly relativistic protons that are constantly accelerated by the shock front of the CME arrive within a few tens of minutes. The CME arrives at the Earth distance within a day which starts the magnetic storm in the magnetosphere.

![Figure 21. Time scales for magnetospheric and ionospheric observations during a CME release (courtesy: M. A. Shea, Geophysics Directorate, Philips Lab.)](image)

4. SUMMARY

During the past few decades, the importance of solar-terrestrial physics has increased since a number of failures have occurred in human technologies due to severe space weather conditions (see articles in Song et al. [2001]). Figure 22 summaries some of the major space weather issues. In order to perform forecasts on space weather and prevent detrimental effects on our technological systems, major improvements in understanding of the STP phenomena are needed. This requires more investments on basic research in the STP area as well as a large set of satellites and ground-based instruments for monitoring the sun, solar wind, magnetosphere and ionosphere (cf. a number of meteorological observations needed for weather forecasts).

5. REFERENCES


Figure 22. Space weather effects on technological systems [Lanzerotti, 2001].


SPACEBORNE HYPRETELESCOPE CONTROLLED BY SOLAR SAILS

O. Lardièrè, A. Labeyrie

LISE CNRS 04870 Saint-Michel-l’Observatoire, France, Email: lardiere@obs-hp.fr, labeyrie@obs-hp.fr

ABSTRACT/RESUME

Snapshot imaging of exoplanets seems feasible with multi-apertures interferometric arrays according the hypertelescope principle. A space hypertelescope can be formed from a vast constellation of ultra-light optical elements positioned by small solar sails along the large primary sphere. Our calculations show that 1m² of solar reflective surface is sufficient to control a mass of 1.3 kg in geo-stationary orbit. However, L2 orbits are preferable for hypertelescopes larger than 200m, in order to avoid tidal forces between elements and slow drifts of the whole constellation. A 10kg-spacecraft is the heavier limit to have a reasonable solar sail size, but 1kg would be preferable. This mass constraint seems compatible with recent advances in MOEMS and ultra-light mirrors.

1. INTRODUCTION

Astronomical interferometry has undergone a spectacular development during the last decades. Current trends announce generations of highly diluted instruments capable of producing direct high-resolution images. Hypertelescopes may be defined as multi-element interferometers using a densified exit pupil which provide an interesting contrast gain for coronagraphy and exoplanet searching [1][2].

2. HYPRETELESCOPE IN SPACE

Two versions of kilometric ground-based hypertelescope are under study: CARLINA-type [3] and OVLA-type [4]. These ground-based hypertelescope versions both require a suitable site and rather complex mechanisms and adaptive optics to compensate the Earth rotation and the atmospheric turbulence. The real future of large hypertelescopes certainly lies in space where optical baselines approaching several hundreds kilometres may become feasible. A first-generation 100m-baseline hypertelescope has been proposed to ESA [5] and to NASA for the Terrestrial Planet Finder mission. It can use a flotilla of dozens of small ultra-lightweight free-elements deployed in the form of a large diluted mosaic spherical mirror (Fig. 1). Reflected beams are recombined in a focal station carried by another free-element. Many focal stations can be used independently. A polychromatic laser beacon located at the centre of the primary sphere provides the fine metrology of the mirror elements.

3. SPACECRAFT DESIGN AND CONTROL

Since the early TRIO proposal [6] of a large formation flying of telescopes in space for interferometry, solar sails seems well suited to provide the continuous force for sky scanning and the exquisitely delicate forces needed for maintaining the element position at the sub-wavelength scale (60 nm is required for coronagraphy in N band). Stronger thrusters, such as ion thrusters, can perhaps be preferred for the faster, but less accurate and less frequent, positioning of the focal stations, as required to quickly acquire a new object after completing an exposure. It is however unclear whether the expected 30°K temperature of the mirrors is compatible with the contaminating exhaust from ion rockets. Moreover, solar sails have an infinite autonomy and are not pollutant for mirror surfaces.

Fig. 2. Ultra lightweight free-flyer element of a space hypertelescope with its solar sails and the spherical mirror attached to it by an angle bracket.

Fig. 3. Example of positioning control along the sunward axis (Z). In absence of correction, only half of the radiation net force is compensated in order to be able to generate relative acceleration between different elements in the both directions.

The solar sails can be shaped like a hat made of three off-axis paraboloidal mirrors (Fig. 2). Three small tiltable mirrors, located at each focus, give the force and torque required to control the position and attitude of the free-flyer, according to all six degrees of freedom. An example of relative positioning along the sunward axis is shown on Fig. 3.

Pointing and phasing are assured by a star tracker receiving light from a laser beacon in a separate free-flyer located at the curvature center of the mirror array. The free-flyer control is ensured as long as solar sails face the sun. Self-realigning is possible if the center of mass is located towards the apex. Slow oscillations (T=2hrs) arise, but they can be damped passively with mechanical, electrostatic (variable capacitor), or electromagnetic (galvanometer) dissipater.

Table 1 shows the position accuracy and accelerations required for pointing and phasing a 100-m, f/3, hypertelescope in geo-stationary orbit, and the corresponding physical characteristics of the solar sails for a 1.3-kg element.

A full sky scanning can be achieved in 6 months by a slow and continuous rotation of the whole satellite constellation around the curvature centre and the solar anti-solar axis.

Tidal forces are strong at geo-stationary orbits and limit the size of the array to 200m. Lagrangian L2 orbits, or artificial L2 orbits, i.e. the modified Lagrangian equilibrium points achievable with radiation pressure [7] are preferable for larger hypertelescopes and for avoiding a slow drifts of the whole constellation.

Our calculations show that 7.5m² of solar reflective surface is sufficient to control the attitude and the position of an optical element having a mass of 10 kg [8]. Recent advances micro-optoelectromechanical systems and ultra-light mirrors (SiC, composite, stretched membrane, etc.) seem compatible with this mass constraint. However, heavier free-flyers remain acceptable up to perhaps 20 kg, if their larger sail can fit in launchers such as Ariane 5. Deployable solar sails are also possible, but presumably less reliable, and have not been considered here.

To increase the solar sail efficiency we can collimate the reflected solar beams using an afocal "solar telescope". A "Fresnel" flat sail is interesting for compactness (Fig. 4). Omni directional (+/-90°) tiltable deflectors using thin plan double sided mirror or μ-shutters with reflective coating seems also possible to increase the angular beam deflection range (Fig. 5).

Table 1. Position and acceleration requirements in geo-stationary orbit, and corresponding expected spacecraft characteristics:

<table>
<thead>
<tr>
<th>Relative position accuracy</th>
<th>Spacecraft total mass (goal)</th>
<th>Inertia moment</th>
<th>Solar sails surface</th>
<th>Spherical mirror element diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 nm</td>
<td>1.3 kg</td>
<td>0.05 kg.m²</td>
<td>1 m²</td>
<td>0.2-0.6 m</td>
</tr>
</tbody>
</table>

Tidal forces are strong at geo-stationary orbits and limit the size of the array to 200m. Lagrangian L2 orbits, or artificial L2 orbits, i.e. the modified Lagrangian equilibrium points achievable with radiation pressure [7] are preferable for larger hypertelescopes and for avoiding a slow drifts of the whole constellation.

A full sky scanning can be achieved in 6 months by a slow and continuous rotation of the whole satellite constellation around the curvature centre and the solar anti-solar axis.

Tidal forces are strong at geo-stationary orbits and limit the size of the array to 200m. Lagrangian L2 orbits, or artificial L2 orbits, i.e. the modified Lagrangian equilibrium points achievable with radiation pressure [7] are preferable for larger hypertelescopes and for avoiding a slow drifts of the whole constellation.

Our calculations show that 7.5m² of solar reflective surface is sufficient to control the attitude and the position of an optical element having a mass of 10 kg [8]. Recent advances micro-optoelectromechanical systems and ultra-light mirrors (SiC, composite, stretched membrane, etc.) seem compatible with this mass constraint. However, heavier free-flyers remain acceptable up to perhaps 20 kg, if their larger sail can fit in launchers such as Ariane 5. Deployable solar sails are also possible, but presumably less reliable, and have not been considered here.

To increase the solar sail efficiency we can collimate the reflected solar beams using an afocal "solar telescope". A "Fresnel" flat sail is interesting for compactness (Fig. 4). Omni directional (+/-90°) tiltable deflectors using thin plan double sided mirror or μ-shutters with reflective coating seems also possible to increase the angular beam deflection range (Fig. 5).

Fig 4. Compact and efficient collimator using a "Fresnel" solar sail.
4. CONCLUSION

From the mission point of view, the nano-satellite constellation philosophy is a recent approach which involves small, low-cost elements ensuring a high level of global reliability through their number and interchangeability.

Here, the 30cm mirror elements, spaced tens of meters apart, are of interest if they can be produced and controlled at low cost. Ultra lightweight hardware components are needed.

Spaceborne hyperscope concept gives a unique application for small non-deployable solar sail which are now technically feasible [9]. Molding techniques with composite materials, or sheet-metal forming techniques such as nickel electroforming, are obvious candidates for the small solar sail, the figure accuracy of which is arc-minutes. The small tiltable light deflectors can use electrostatic mini- or micro-mirror techniques. The attitude and position stabilisation loop can use coarse error signals from a radio-telemeter or local GPS-type system, to be tested in NASA's Nanosatellite Constellation Trailblazer mission.

Lastly, to complement the ESA SMART-2 mission, we propose a set of 2 ultra-light geo-stationary satellites equipped of solar sails, laser ranging and star trackers to validate the control of a formation flying by solar radiation pressure, and to prepare future space hyperscope missions.

5. REFERENCES


SMART-1 SCIENCE EXPERIMENTS CO-ORDINATION


(1) European Space Agency, ESTEC, Keplerlaan 1, NL-2201 AZ, Noordwijk, The Netherlands,
Email: dalmeida@rssl.esa.int, Bernard.Foing@esa.int

ABSTRACT

SMART-1 is the first European Space Agency mission to the Moon [1] [2] [5], due for launch in the first months of 2003. Its primary goal is to test new technologies for space navigation and science. In its science experiments, SMART-1 will include new, very compact experiments. This paper aims to demonstrate some of the science experiment operations foreseen for the mission. We describe the SMART-1 mission, its orbit and example scenarios for imaging specific targets (such as Tycho and Copernicus craters).

1. THE SMART-1 MISSION

1.1 Overview

SMART-1’s primary objective is to test and validate a new electric propulsion engine for potential use on larger ESA Cornerstone missions. However, the SMART-1 spacecraft will also carry a number of scientific instruments and experiments for use en-route to, and in orbit about, the Moon.

1.2 Experiments

The payload comprises several instruments and experiments: the Asteroid-Moon Micro Imager Experiment (AMIE), the Demonstration Compact Imaging X-ray Spectrometer (D-CIXS), the X-ray Solar Monitor (XSM), the SMART-1 Near-Infrared Spectrometer (SIR), the Electric Propulsion Diagnostic Package (EPDP), the Spacecraft Potential, Electron and Dust Experiment (SPED), the Deep Space X/Ka Band TTC Experiment (KaTE), the Radio Science Investigation for SMART-1 (RSIS).

1.3 Objectives

The majority of scientific objectives will be achieved from lunar orbit. The AMIE multispectral high-resolution camera will mainly aim to image the lunar south pole and map the southern regions of the Moon. D-CIXS will look for the spatial distribution of major lunar rock types on the sunlit side of the Moon, and the X-ray emission from the impact of solar wind electrons on the lunar night-side. SIR will gather data to study the mineralogy of the lunar surface. During this phase, the RSIS experiment will also take place, using AMIE images and the high accuracy tracking provided by KaTE to measure the lunar libration.

1.4 Mission Phases

The SMART-1 mission has three main phases. Immediately after launch there will be a commissioning phase during which all systems are tested and routine operational sequences are sent to the spacecraft to check that everything is running according to specifications. Commissioning will be followed by a long, fifteen month cruise phase that will take the spacecraft to the Moon. During this phase, all the engine tests will be performed in order to validate the propulsion system. Also during cruise some science data acquisition will be possible during the periods that SMART-1 is not thrusting. The third phase, lunar operations, starts at Moon arrival. Most of the scientific data will be acquired during this period.

1.5 Mission Co-ordination

Also in the frame of the mission a co-ordinated utilisation of the experiments is envisaged. The Science and Technology Operations Co-Ordination (STOC) will be in charge of this task. The STOC will, in light of the capabilities of the experiments, advise experiment teams to study specific features of the Moon simultaneously. The data acquired this way, when cross-checked, will be able to deliver improved scientific return. The first test analyses performed to date in preparation of STOC activities were related to the AMIE high-resolution camera. The camera was used to obtain different filter information while completing simulated imaging of the well known craters Tycho and Copernicus. Also, some evaluations were made on other AMIE science objectives, as well as simultaneous AMIE/SIR operation for combined science.

2. LUNAR ORBIT

SMART-1 will be inserted in a Polar Orbit, with pericentre near the lunar south pole. The perilune distance will vary during the mission (see Fig. 1).
3. THE AMIE MICRO IMAGER (AMIE)[1][2]

The study performed was based on the SMART-1 imager. AMIE is a compact imaging system. It acquires 1024x1024 pixel images with a field of view of 5.3 degrees by 5.3 degrees.

A filter was set in front of the optics, divided into different wavelength zones, as can be seen in Fig. 2.

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>No Filter</td>
</tr>
<tr>
<td>915</td>
<td></td>
</tr>
<tr>
<td>960</td>
<td>9</td>
</tr>
<tr>
<td>847</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 2. AMIE filters

The filter layout accommodates the two different camera orientations with respect to the spacecraft movement.

This way it is possible to obtain multi-wavelength information using pictures taken of the same area through each of the filters.

4. OBTAINING MULTI FILTER INFORMATION FOR COPERNICUS AND TYCHO CRATERS

4.1 Orbit Selection

Using a software tool called Project Test Bed (PTB) [3], a program that simulates the SMART-1 mission with the trajectory parameters given by the European Space Operations Centre (ESOC), it is possible to select a site of interest in the Moon orbit as well as obtain spacecraft parameters, such as the height and spacecraft velocity in relation to the Moon.

For this scenario, two consecutive orbits were selected that pass by two well-known fresh impact craters, Copernicus and Tycho (Fig. 3). The perilune is at 55 degrees south and 165 degrees east, and the height is 317 km.

4.2 Orbit and scenario parameters

In order to plan how to cover a crater it is necessary to know what is the projection of the AMIE field of view on the Moon.

This is calculated in Eq. 1, using simple trigonometry:

\[
\tan(5.3/2) = \frac{m}{d}
\]

\[
\Rightarrow m = 2 \times d \times \tan(2.65)
\]

Where \(m\) is the footprint size in one dimension and \(d\) is the distance from SMART-1 to the Moon.

Knowing the orbit height, \(d\), from PTB, as it is shown in Fig. 3:

Fig. 3. Plot of the orbit time versus the height of the spacecraft. The colours refer also to Fig. 5

it is possible, using Eq. 1, to calculate the field of view size projection on the Moon for the all scenario time. The result is shown in Fig. 4:

Fig. 4. AMIE field of view projection size. Near perilune it is possible to have 30 Km FOV and hence 30 m/pixel resolution
4.3 Operations Examples

Fresh impact craters such as Copernicus and Tycho are primary targets for colour imaging.

4.3.1 Copernicus Crater

Copernicus crater is situated at 20 degrees west, 9.6 degrees north and has 94 km diameter. In this simulated orbit, when SMART-1 passes by Copernicus it is 4946 km above the lunar surface. At this height the field of view of one of the filters of the AMIE camera projected on the Moon is a rectangle of 228 km by 114 km. Therefore, three pictures are necessary to obtain multi-filter information on this crater, as illustrated in Fig. 6, only with relevant colour filters from Fig. 2. The area covered on the sequence can be seen in Fig 5.

![Fig. 5. Projection of two consecutive SMART-1 orbits on the Moon](image)

Fig. 6. Copernicus picturing sequence. Time between pictures: 8 minutes (see 4.3.3)

Some context imaging is also added, including some of the ejecta. The resolution would be about 440 m/pixel.

4.3.2 Tycho Crater

Tycho has 102 Km diameter and is situated 43.4 degrees south and 11.1 degrees west. As SMART-1 travels south, as can be seen in Fig. 4, the spacecraft gets closer to the Moon. In the simulation, at Tycho latitude SMART-1 is 1399 km above the Moon. This obviously provides an improved resolution for AMIE. One filter projection in this case is a rectangle of 64 km by 32 km, with a resolution of 126 m/pixel. A sequence of four pictures is needed to cover Tycho in this case, as can be seen in Fig. 7. The area covered on the sequence can be seen in Fig 5.
4.3.3 Time step between pictures

To calculate the time step between pictures a simple calculation was needed. The Spacecraft velocity is simulated with PTB (Fig 8).

\[ V_{FP} = \frac{R_m}{R_m + R_{SC}} V_{SC} \] (2)

where \( V_{FP} \) is the foot print velocity, \( R_m \) is the radius of the Moon, \( R_{SC} \) is the distance from SMART-1 to the Moon and \( V_{SC} \) is the velocity of the spacecraft.

Having the velocity of the spacecraft (Fig. 9)

It is possible using Eq. 2, obtain the foot print velocity for the entire simulation (Fig. 10).

Fig. 10. SMART-1 foot print velocity

Knowing the size of one filter of the AMIE camera projected on the Moon, and the velocity of the projection it is easy to calculate the time step.

Copernicus:

Size of the filter: 228 km x 114 km
Foot print velocity: 226 m/s

\[ TimeStep = \frac{114000}{226} = 504 s = 8.4 \text{ min} \] (3)

Tycho:

Size of the filter: 64 km x 32 km
Foot print velocity: 861 m/s

\[ TimeStep = \frac{32000}{861} = 37 s \] (4)

5. FURTHER STUDIES

Some more studies were performed, but need improvement, and work is continuing to this end.

5.1 Global Coverage

For global coverage, it was possible to deduce that using the “No Filter” area of the AMIE camera at high northern latitudes, a picture has about 450 km side. This is more than 10 degrees in Moon latitude coordinates. Therefore global coverage is easily feasible. However for southern latitudes whenever the spacecraft is below 500 km there are some gaps in the coverage along the spacecraft movement due to download time constraints from AMIE to the...
spacecraft mass memory [4]. Exact latitudes at which gaps occur vary as the orbit and perilune fluctuate. Considering the coverage obtained by multiple passages, it is possible to see in Fig. 5 that there is a gap between two passages at Tycho latitude. This was corroborated by a new software tool, Mapping and Planning Payload Science (MAPPS), as can be seen in Fig. 11.

In yellow we have the projection of the field of view of one of the colour filters. In green is the SMART-1 trail. In the simulation the spacecraft rotates near the equator, therefore the field of view passes from the left of the trail to the right.

The distortion of the field of view is due to the Moon orthographic projection.

It is clear that for northern latitudes as SMART-1 is far from the Moon there is overlapping of contiguous fields of view. For southern latitudes between 0 and 70 degrees there is a gap up to 30%. Further studies should be developed in order to create strategies to overcome this problem.

5.2 Crater Counting

Statistical studies of the Moon, based on crater counting are possible if pictures of 50 km by 50 km are collected with 50 m resolution or 100 km by 100 km, with 100 m resolution.

In this simulation those criteria are achieved when SMART-1 is 50 degrees or more south.

5.3 Combining with SMART-1 Near Infrared spectrometer (SIR)

Copernicus and Tycho are also interesting targets for infrared spectrometry. It is then envisaged that while AMIE is imaging, SIR can be collecting spectra. This way, also context imaging is obtained for SIR. This is possible because, nevertheless SIR is always sending data through the CAN bus, not allowing its use by AMIE. The camera has a memory that holds 4 images, what is the maximum needed to picture these targets. Once the measurements are made and SIR stops collecting data, AMIE is able to send the pictures to the mass memory.

5.4 South Pole

Very high-resolution images are foreseen for the South Pole. It is possible to get to resolutions of about 50 m/pixel.

Longer exposure images are also foreseen to see the bottom of craters, illuminated by rim straylight, and possible ice deposits, as well as high resolution images for illumination and landing site studies.

Further, detailed studies will cover the South Pole-Aitken basin.

6. CONCLUSION

When, after a long journey, SMART-1 arrives at the Moon, there is the need to have a science data gathering campaign, well studied and planned in order to maximize the output of the mission.

The studies presented in this paper were done in this framework. If we maximize the amount of operations procedures at the Moon, we can be ready for any scenario that we may encounter.

Therefore, this is just a first step. Much work is still envisaged, with a great amount of discussion between experiment teams, the Science and Technology Working Team (STWT) and the project scientist team.

This is needed to ensure the success of the mission.
REFERENCES

1. Foing et al. 32nd LPSC, ESA’s SMART-1 Mission to the Moon, abstract LPI 32, 1787, 2001

ACKNOWLEDGMENTS

We thank the SMART-1 project & AMIE team

Credits:

Lunar Stripe image (courtesy of USGS)
Lunar crater images (courtesy of NASA)
TARGET SELECTION FOR THE HIGH RESOLUTION STEREO CAMERA (HRSC) ON THE ESA MARS EXPRESS MISSION

Ernst Hauber(1), Gerhard Neukum(2), Jarmo Korteniemi(1,3), and the HRSC Target Selection Team(1)

(1)German Aerospace Center, Rutherfordstr. 2, 12489 Berlin, Germany, Email: Ernst.Hauber@dlr.de
(2)Freie Universität Berlin, Malteserstr. 74-100, 12249 Berlin, Germany, Email: gneukum@zedat.fu-berlin.de
(3)University of Oulu, P.O. Box 3000, 90014 University of Oulu, Finland, Email: Jarmo.Korteniemi@dlr.de

ABSTRACT

The High Resolution Stereo Camera (HRSC) onboard the European ESA Mars Express (MEX) mission will map > 50% of the Martian surface in stereo and colour with a resolution of ≤15 m/pixel. In order to optimise the scientific return of the instrument, the preparation of a detailed list of surface targets and their specific scientific interest together with ancillary information is mandatory. We describe the organisation of the list of >1400 individual targets, the parameters specified for each target, and how the list will be used in operations planning. Finally, we outline possible further applications of the list for upcoming Mars missions like the Mars Reconnaissance Orbiter.

1. INTRODUCTION

The High Resolution Stereo Camera (HRSC) (Fig. 1) is a multiple line pushbroom camera instrument [1,2] to be launched onboard Mars Express (MEX) in mid-2003 [3,4]. Nine superimposed image tracks are acquired nearly simultaneously (along-track) by 9 CCD line sensors (each with >5000 pixels) mounted in parallel and behind one single optics. At a periapsis height of 250 km, the swath width is about 52 km and the resolution is ~10 m/pixel. The HRSC will cover ≥ 50% of the Martian surface at a spatial resolution of ≤15 m/pixel, in stereo, four colours, and at five phase angles. More than 70% of the surface can be observed at a spatial resolution of ≤30 m/pixel. The scientific output is significantly extended by using an additional external Super Resolution Channel (SRC). The SRC is based on the instrument development for the Rosetta Lander and is mounted on the MEX spacecraft below the HRSC stereo scanner in a common honeycomb structure in order to minimise interfaces with the spacecraft. It is a framing device and uses an interline CCD detector to cope with the very short exposure times. The 1 m focal length telescope provides a spatial resolution of 2.3 m/pixel at an altitude of 250 km. A large international co-investigator team will participate in the scientific analysis of the HRSC/SRC images.

The HRSC Target Selection Team: Alexander Basilevsky, Ruslan Kuzmin, Vernadsky Institute, Moscow (Russia); Michael Carr, USGS, Menlo Park (USA); Philippe Masson, Francois Costard, Nicolas Mangold, Université Paris-Sud, Orsay (France); Ronald Greeley, David Nelson, Arizona State University, Tempe (USA); James Head, Harald Hiesinger, Brown University, Providence (USA); Boris Ivanov, Institute for Dynamics of Geospheres, Moscow (Russia); Ralf Jaumann, Jürgen Oberst, Harald Hoffmann, Julia Lanz, Stephan van Gasselt, Dennis Reiss, Thomas Roatsch, Klaus-Dieter Matz, Rene Lauffer, DLR, Berlin (Germany); Peter Kronberg, Technische Universität Clausthal, Clausthal-Zellerfeld (Germany); Thomas McCord, University of Hawaii, Honolulu (USA); John Murray, Open University, Milton Keynes (UK); Gian-Gabriele Ori, Università d’Annunzio, Pescara (Italy); Martín Pätzold, Universität zu Köln, Köln (Germany); Jouko Raitala, University of Oulu, Oulu (Finland); Tilman Spohn, Westfälische Wilhelms-Universität, Münster (Germany); Steven Squyres, Robert Sullivan, Cornell University, Ithaca (USA).

2. TARGET SELECTION

In order to facilitate effective operations planning and to maximise the scientific return of the HRSC/SRC instrument by fully exploiting its unique capabilities, we prepared a global list of geologic targets to be imaged during the MEX mission. The target list is organised as a spreadsheet, each line of which represents a single target. Several groups of parameters (i.e., spreadsheet columns) are specified for each target (i.e., line). The parameters and their meanings are defined in Tab. 1.


© European Space Agency • Provided by the NASA Astrophysics Data System
<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Example</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>MC-23 005</td>
<td>Map quadrangle (e.g., MC-23) and running number (e.g., 005)</td>
</tr>
<tr>
<td>Name</td>
<td>&quot;small valley&quot;</td>
<td>Short name or description of target</td>
</tr>
<tr>
<td>Class (15 columns)</td>
<td>&quot;x&quot; or &quot;y&quot;</td>
<td>One or several entries possible for each target: What is the geologic process which is of interest for this target?</td>
</tr>
<tr>
<td>V volanic</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>T tectonic</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>I impact</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>F fluvial</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>G glacial or polar</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>L layering</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Pr permafrost</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Eb exobiologic</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Es eolian</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Mw mass wasting</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>La lacustrine</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Ls landing site</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Te technical</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>O other</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Ts terrain sampling</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>yes or no</td>
<td>Is the target unique or not unique within a quadrangle?</td>
</tr>
<tr>
<td>Location (4 columns)</td>
<td>5.0, 8.0, 122.0,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>124.0</td>
<td>Geographical coordinates of target boundaries</td>
</tr>
<tr>
<td>Mosaic</td>
<td>6</td>
<td>Number of HRSC standard images required to cover the target</td>
</tr>
<tr>
<td>HRSC mocupixel format</td>
<td>1, 2, 3, 5, 6, 16</td>
<td>16 possible HRSC imaging modes (see Tab. 2 for details)</td>
</tr>
<tr>
<td>Priority (5 columns)</td>
<td>high medium low</td>
<td>Priority for good resolution of this/these CCD line(s) and the SRC channel</td>
</tr>
<tr>
<td>SRC imaging</td>
<td>spot, raster, or contiguous</td>
<td>Mode of SRC imaging (see text for details)</td>
</tr>
<tr>
<td>Preferred local time</td>
<td>morning, noon, evening</td>
<td>Should the target be sampled during a specific time of the (Martian) day? (e.g., dust devils occur in the afternoon, morning fog?)</td>
</tr>
<tr>
<td>Preferred illumination</td>
<td>low, medium, high</td>
<td>Elevation of sun above horizon</td>
</tr>
<tr>
<td>Preferred local season</td>
<td>spring, summer, fall, winter</td>
<td>Season in which the target should be imaged (e.g., imaging of some high latitude targets without coverage by seasonal CO₂ cap)</td>
</tr>
<tr>
<td>Multitemporal coverage</td>
<td>yes or no</td>
<td>Variable features (e.g., polar caps, dust deposits) might be imaged several times (e.g., summer/winter, before/after dust storms)</td>
</tr>
<tr>
<td>Scientific rationale</td>
<td>separate document</td>
<td>Why is it important to image this target?</td>
</tr>
<tr>
<td>Lessons learned from MGS</td>
<td>separate document</td>
<td>What did we learn from the MGS mission about this target?</td>
</tr>
<tr>
<td>Geology</td>
<td>Hr, NPl, Ael, ...</td>
<td>Target geology as mapped by [6], [7], and [8]</td>
</tr>
<tr>
<td>Viking coverage</td>
<td>(32), 60-100, r.g.b</td>
<td>(number of Viking Orbiter images), resolution range, color filters</td>
</tr>
<tr>
<td>MOC coverage</td>
<td>good, medium, or sparse</td>
<td>Availability of very high-resolution images from the Mars Orbiter Camera (MOC)</td>
</tr>
</tbody>
</table>

At the time of writing (June 2002, one year before launch), more than 1200 targets have been specified (Fig. 2). The original list has been compiled on a quadrangle-by-quadrangle basis. Each of the 30 cartographic quadrangles covering the Martian surface according to the mapping scheme of the U.S. Geologic Survey [9] was assigned to an individual scientist. Since different individuals worked on different quadrangles, there is a bias towards specific processes in some quadrangles, depending on the scientific expertise of the workers. To avoid such inhomogeneities in the final list, we will revise the complete list with a process-based approach. Scientists with a profound background with respect to specific processes, e.g., fluvial processes, will check all targets classified as fluvial, and will approve or reject them, or add new ones.

In separate documents, the scientific rationale for imaging each target is comprehensively specified.
Important findings of the MGS mission related to the targets are also described separately. In order to estimate future data rates, we determine the areal size of each target and calculate the number of standard images and the data volume required to obtain an image mosaic of the entire target.

![Histogram of targets and respective classes](image)

**Fig. 2.** Histogram of targets and respective classes (see Tab. 1 for class definition). Note that the cumulative number of targets for all classes exceeds 1200 because one target may fall into several classes.

In addition to the table, each target is graphically marked on a global, Viking-based image map (i.e., the MDIM-2 imaging model of the U.S. Geol. Survey [10]), allowing for quick visual inspection and further use in planning activities (see below). We prepared separate image base maps for each quadrangle and marked each target in the map (Fig. 3). To allow for better visualisation and later processing, the targets were marked in 16 different "layers" of the map, each layer corresponding to one of the 16 HRSC macropixel-formats (MPF) used for standard imaging (see Tab. 2).

The data volumes for the MPF’s as shown in Tab. 2 were calculated for a standard HRSC image strip. A standard image consists of 5184 pixels across track and 30,000 lines along track (3 stereo-baselines) with a data compression ratio of 1:10. The SRC may be operated either in contiguous mode, raster mode, or spot mode. In contiguous mode, a sequence of overlapping image frames will be acquired (highest data rate). The raster mode will regularly take image frames with equal exposure times and equal delays between successive images. The contiguous and raster

![Example of target map](image)

**Fig. 3.** Example of target map (shown: MC-06). Different shading of targets indicates different macropixel-formats (cf. Tab. 2).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mapping Stereo</td>
<td>1x1</td>
<td>2x2</td>
<td>8x8</td>
<td>4x4</td>
<td>183</td>
<td>possible</td>
<td>w/o Col. or Pho</td>
</tr>
<tr>
<td>2</td>
<td>Mapping Stereo Color</td>
<td>1x1</td>
<td>2x2</td>
<td>4x4</td>
<td>4x4</td>
<td>204</td>
<td>possible</td>
<td>w/o Col. or Pho</td>
</tr>
<tr>
<td>3</td>
<td>Mapping SpecPhot</td>
<td>1x1</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
<td>163</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>4</td>
<td>Mapping Large Area</td>
<td>1x1</td>
<td>4x4</td>
<td>8x8</td>
<td>4x4</td>
<td>143</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>5</td>
<td>Stereo High-Res.</td>
<td>1x1</td>
<td>1x1</td>
<td>-</td>
<td>2x2</td>
<td>381</td>
<td>possible</td>
<td>w/o Pho</td>
</tr>
<tr>
<td>6</td>
<td>Stereo Quint.</td>
<td>1x1</td>
<td>2x2</td>
<td>8x8</td>
<td>2x2</td>
<td>224</td>
<td>possible</td>
<td>w/o Col.</td>
</tr>
<tr>
<td>7</td>
<td>Stereo Triple</td>
<td>1x1</td>
<td>1x1</td>
<td>8x8</td>
<td>8x8</td>
<td>340</td>
<td>possible</td>
<td>w/o Pho + Col</td>
</tr>
<tr>
<td>8</td>
<td>Stereo Color</td>
<td>1x1</td>
<td>2x2</td>
<td>4x4</td>
<td>2x2</td>
<td>245</td>
<td>possible</td>
<td>w/o Col or Pho</td>
</tr>
<tr>
<td>8a</td>
<td>Spectrophotometry</td>
<td>1x1</td>
<td>4x4</td>
<td>2x2</td>
<td>4x4</td>
<td>245</td>
<td>possible</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>Color</td>
<td>-</td>
<td>-</td>
<td>1x1</td>
<td>-</td>
<td>435</td>
<td>possible</td>
<td>If only 3 color</td>
</tr>
<tr>
<td>10</td>
<td>Color Stereo</td>
<td>1x1</td>
<td>2x2</td>
<td>2x2</td>
<td>-</td>
<td>272</td>
<td>possible</td>
<td>no</td>
</tr>
<tr>
<td>11</td>
<td>Map. Med.-Res.</td>
<td>2x2</td>
<td>2x2</td>
<td>8x8</td>
<td>4x4</td>
<td>102</td>
<td>possible</td>
<td>w/o Col or Pho</td>
</tr>
<tr>
<td>12</td>
<td>Med.-Res. SpecPhot</td>
<td>2x2</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
<td>82</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>13</td>
<td>Med.-Res. Stereo</td>
<td>2x2</td>
<td>2x2</td>
<td>8x8</td>
<td>2x2</td>
<td>143</td>
<td>possible</td>
<td>w/o Col</td>
</tr>
<tr>
<td>14</td>
<td>Mapping Med.-Res. Large Area</td>
<td>2x2</td>
<td>4x4</td>
<td>8x8</td>
<td>4x4</td>
<td>62</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>15</td>
<td>Mapping Low-Res.</td>
<td>4x4</td>
<td>4x4</td>
<td>8x8</td>
<td>4x4</td>
<td>41</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>16</td>
<td>Limb sounding</td>
<td>1x1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>109</td>
<td>possible</td>
<td>possible</td>
</tr>
</tbody>
</table>

**Tab. 2.** The 16 main macropixel-formats (MPF) of HRSC. The figures in the columns for the different CCD lines indicate sampling density, e.g., "1x1" means that each pixel will be recorded individually along and across flight direction (highest resolution), "8x8" means that 8x8 pixels will be summed up (lowest resolution; data volume reduced by factor 64). MPF modes marked in light gray will be used most frequently to assure homogeneous imagery. See text for details on data volume and SRC imaging.
modes require one of the four HRSC signal chains for real-time data transfer. This signal chain is then not available for the data transfer from the HRSC CCD lines, thus limiting the HRSC data rate (see last column “SRC Raster/Contiguous” in Tab. 2). In the spot mode, up to 8 SRC image frames can be stored in the SRC internal memory. This memory will be read out after the HRSC CCD-lines have finished imaging. Therefore, all 4 signal chains are available for HRSC and the spot modes is possible without limiting HRSC (see column “SRC spot” in Tab. 2).

The list, including the separate documents for the scientific rationale, the lessons learned from MGS, and the target image maps, is available to the HRSC Team via a website on the internet.

3. PLANNING SOFTWARE

The target maps will be the standard input for imaging planning. As a flexible and user-friendly planning tool we developed the interactive program Mars Express Science Opportunity Analyzer (MEXSOA) within the Interactive Data Language (IDL®) environment. It allows to plot spacecraft ground tracks (Fig. 4), the orbit periapses (Fig. 5), and the HRSC swath (Fig. 6) onto image maps. Information about the planets ephemerides, the orbit of the spacecraft, and the instrument are provided to the program as SPICE kernels (Spacecraft, Planet, Instrument, C-matrix, Event, e.g., [11]).

Fig. 5. Periapsis positions of during the MEX mission in the central Valles Marineris region. Red numbers indicate orbit numbers. Bright yellow dots: sun-lit periapsis (day), dark dots: shadowed periapsis (night). Background map is MDIM-2 [10]. Note that the final orbit is still to be determined, the displayed positions correspond to a provisionary orbit (the same is valid for Fig. 4).

Fig. 4. Topographic map of Mars and MEX ground tracks of (provisional) orbits 50 - 150 (the final orbit has still to be determined). Shown are the groundtracks within ±90° true anomaly (colour of ground track as a function of true anomaly) as well as the periapsis positions (small red dots). Note that there is a pattern of 13 “building sites” spanning the entire longitude range of the planet. From each of these 13 sites we will begin to map, e.g., to acquire overlapping images for large image mosaics. In Fig. 6 we give an example of a planned imaging sequence associated to site 9 (white box).
An operator will check which targets are crossed by (or near) the ground track of the spacecraft (Fig. 6a). It is straightforward, then, to select the desired imaging area. The corresponding start and end times of imaging will be stored together with additional parameters (e.g., imaging mode, illumination conditions) for each orbit (Fig. 7). The stored data will then be converted to commanding sequences.

Fig. 6. Example of planned HRSC imaging sequences as visualized with the Mars Express Science Opportunity Analyser (MEXSOA). (a, top left) Ground track of orbit 52 (red) and HRSC swath (blue) (background: MOLA altimetry merged with MOC wide angle image map [12, 13]). The area to be imaged is marked by the dashed white area. Additional information related to the image and the Martian surface (e.g., position, illumination, etc.) is displayed simultaneously in separate windows (Fig. 7); (b, top right) After a repeat cycle of 13 orbits, the ground track of orbit 65 plots on the surface at a lateral distance of ~52 km (10% overlap at periapsis). The image acquired during orbit 52 is shown in yellow; (c, bottom) The imaging sequence for orbit 156 is shown together with previous images taken at orbits 52, 65, 78, 91, 104, 117, 130, and 143 (dashed yellow areas).

Fig. 7. Ancillary information provided by MEXSOA. (left) Surface type (to be implemented), macropixel format (MPF), imaging times, number of lines, and data volume; (top) For each (cursor) position on the map, MEXSOA displays coordinates and (orbit-dependent) illumination, required spacecraft tilt, and image resolution.
The footprints of acquired images (cf. Fig. 6b) together with ancillary data (e.g., exposure times, image quality) will be stored in a database. In the ongoing mission, the interface between MEXSOA and the database will provide necessary information in order to optimise imaging parameters by learning from previous experience (Fig. 6c).

4. FUTURE

The list of targets will be the basis of long-term and short-term science planning for the HRSC instrument. It will be updated continuously, according to new results from the ongoing MGS and 2001 Mars Odyssey missions. Due to the considerable lifetime of Mars Express (nominal plus extended mission: two Martian years or four Earth years) we will also extend or modify the target list in response to incoming HRSC and SRC data. The list may also be of interest for future orbiter missions like the Mars Reconnaissance Orbiter to be launched in 2005.

5. REFERENCES


COORDINATION OF MARS EXPRESS AND BEAGLE 2 SCIENCE OPERATIONS

R. Trautner (1), A. Chicarro (1), P. Martin (1)

(1) Research and Scientific Support Department, ESA/ESTEC, Noordwijk, The Netherlands
Email: Roland.Trautner@esa.int Tel. +31 71 565 3955 Fax: +31 71 565 4697

ABSTRACT

The Mars Express orbiter carrying the Beagle 2 lander will arrive at Mars in late 2003 [1]. Both spacecraft carry instruments designed to investigate the Martian atmosphere, surface and subsurface. The evaluation of science data from both the Mars Express orbiter and the Beagle 2 lander will benefit from a coordination of the measurements done by orbiter and lander instruments. Data acquired by both the orbiter and the lander during the early mission should be used for the refinement of the science operations planning in order to optimise the scientific return. The basic operations planning strategies for both spacecraft are recapitulated, and the approach for coordinating the science operations of orbiter and lander is explained. The relevant requirements and constraints are discussed, and the expected benefits and results of coordinated operations are highlighted.

1. REMOTE SENSING AND IN-SITU INVESTIGATIONS

The Mars Express orbiter and the Beagle 2 lander will perform a variety of measurements, generating remote sensing and in-situ data. Any measurement done from orbit may suffer from a number of problematic effects – limited resolution, atmospheric effects, few opportunities to observe a particular location, and many others. On the contrary, in-situ measurements provide the possibility to accurately determine the properties of a local environment. But unlike the remote sensing approach, the validity of the data – with respect to extrapolation to greater geometric scales – is very limited. A combination of remote sensing and in-situ investigations, as planned with the Mars Express orbiter and Beagle 2, offers the possibility to achieve global coverage on the one hand and ground truth for the space-based observations on the other hand. Many of the Orbiter and Lander instruments measure the same or similar parameters of the Martian environment. These instruments will naturally benefit from a coordination of their operations, which will improve the quality of their data products and provide an opportunity to discover new phenomena. Tables 1 and 2 illustrate possible areas of coordination by highlighting the targets of the orbiter and lander investigations.

2. MARS EXPRESS MASTER SCIENCE PLAN

The operations of the Mars Express orbiter instruments are planned on a long-term basis as a result of requirements and constraints such as science objectives, spacecraft safety and resources, operations cost, orbit maintenance planning, etc. This long-term plan is called the Master Science Plan (MSP). Its purpose is to schedule the acquisition of science data by the Mars Express spacecraft in a way that is consistent with both the scientific objectives of the mission and the resources available for that data collection [2]. Resources such as on-board data storage capacity, telemetry bandwidth, spacecraft visibility from the Perth ground station, available bandwidth between Perth and ESOC, computing power and storage capacity available at Perth and ESOC, etc. are important constraints on Mars Express science operations. Mars Express data acquisition periods are predefined for each orbit and referenced to the respective orbit pericentre. A pre-defined orbit ground track is envisaged, which allows to enter the planned science sequence at a later point in time in case of complications, while maintaining the original science planning for the continued mission. This approach allows the definition of the science operations for the Mars Express orbiter a long time in advance, and simplifies various aspects of mission operations such as orbit maintenance and planning of on-board resource utilization.

3. BEAGLE2 OPERATIONS PLANNING

The operations approach used for the Beagle2 lander differs significantly from the one used for its orbital companion. After a successful landing, Beagle2 will perform an automated sequence of activities (initial checkouts, opening of lid, deployment of solar panels, etc.) in order to reach a safe standby condition for surviving the time until initial contacts. In a first operational phase a number of initial activities are performed, such as post landing assessment, system


© European Space Agency • Provided by the NASA Astrophysics Data System
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Instrument</th>
<th>Institute</th>
<th>Science Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRSC</td>
<td>Super/High-Resolution Stereo Color Imager</td>
<td>DLR, Berlin (D)</td>
<td>Characterization of the surface structure, topography and morphology at high spatial resolution (up to 10m/pixel) / super resolution (up to 2m/pixel) (i). Characterization of terrain composition and surface physical properties at high spatial resolution (s). Characterization of atmospheric phenomena (tp, da).</td>
</tr>
<tr>
<td>OMEGA</td>
<td>IR Mineral Mapping Spectrometer</td>
<td>IAS, Orsay (F)</td>
<td>Characterization of the composition of surface materials, space and time distribution of the various classes of silicates, hydrated minerals, oxides and carbonates in soils and rocks, (s), and of ices and frosts (w), at medium resolution. Study the distribution of atmospheric CO₂, CO (a), H₂O (w) and aerosols (da).</td>
</tr>
<tr>
<td>MARSIS</td>
<td>Subsurface-Sounding Radar / Altimeter</td>
<td>Univ. of Rome, JPL (I/USA)</td>
<td>Primary objective: map the distribution of water in the upper crust of Mars (inventory, mechanisms of transport and storage, stability at the surface) (w). Subsurface geologic probing (s), surface characterization (i), ionosphere sounding.</td>
</tr>
<tr>
<td>PFS</td>
<td>Atmospheric Fourier Spectrometer</td>
<td>CNR-IFSI, Rome (I)</td>
<td>Global monitoring of the 3D temperature field in the lower atmosphere (tp); measurements of CO (a), H₂O (w), search for other atmospheric components; D/H ratio (a); atmospheric aerosols (da); atmospheric radiance balance (tp), global circulation. Surface temp. and thermal inertia (tp); restrictions of mineralogical composition of the surface layer (s); nature of surface condensates (a,w); pressure and height local determination (CO₂, altitude) for selected regions (tp).</td>
</tr>
<tr>
<td>SPICAM</td>
<td>UV and IR Atmospheric Spectrometer</td>
<td>CNRS, Verrues (F)</td>
<td>Investigate key issues about ozone (a), its coupling with H₂O, aerosols (da), atmospheric vertical temperature structure (tp), ionospheric studies. H₂O abundances and vertical profiling of H₂O (w) and aerosols (da). Ozone detection, O₃ absorption. Vertical profiles of CO₂, temperature, O₃ (a), clouds and aerosols (da).</td>
</tr>
<tr>
<td>ASPERA</td>
<td>Energetic Neut. Atoms Analyzer</td>
<td>SRI, Kiruna (SW)</td>
<td>Plasma investigations, study and imaging of atmospheric escape</td>
</tr>
<tr>
<td>MaRS</td>
<td>Radio Science Experiment</td>
<td>Univ. Köln, (D)</td>
<td>Sounding of the Martian atmosphere to derive vertical density, pressure and temperature profiles (tp), ionospheric sounding; bistatic radar experiments (s); study of gravity anomalies.</td>
</tr>
</tbody>
</table>

Tab. 1. – Mars Express orbiter payload description

Checkouts and a first acquisition of environmental data, followed by the deployment of the PAW and a first topographic landing site survey. During a second operational phase, a series of activities comprising atmospheric analysis, calibration activities, soil and rock analysis will be performed. Activities will be selected according to their scientific priority, requirements (target availability, power requirements, environmental conditions, operational requirements) and constraints (risks, environmental constraints, etc.). A third operational phase will be used to complete the full palette of scientific activities the lander is able to perform, and continue operations until the end of the Beagle2 lifetime. Contrary to the operations planning for the MEX orbiter, the science activities performed by the Beagle2 lander will be dominated by a short-term operations plan worked out in the time between radio contacts. Progress will depend on the latest status of onboard systems, achievements during previous operations, and lessons learned as work progresses.

4. COORDINATION OF SCIENCE OPERATIONS

The coordination of science operations for two spacecraft having a completely different operational approach is a complex task. Starting with inputs and requests from both orbiter and lander instrument teams, a plan for coordinated operations needs to be established, which is compatible with the long term plan used for the orbiter operations on the one hand, and fits into the operations planning for the lander on the other hand. Requirements and constraints in many areas must be taken into account in order to achieve the goal of optimising the overall science output of the Mars Express Mission. The following items describe the approach, a number of important constraints and requirements and the expected outcome of this effort.

4.1 Planning Phases

The initial planning of coordinated operations is done in parallel to the planning of both the orbiter MSP and
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Instrument</th>
<th>Institute</th>
<th>Science Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAP</td>
<td>Gas Analysis Package</td>
<td>Open University (UK)</td>
<td>Search for evidence of life past or present on Mars. Quantitative and stable isotopic measurements of gases such as H₂, N₂, O₃ and CO₂; Processing and determination of some of the Noble gases (Ne, Ar, and Xe) (a) as well as anticipated trace constituents such as CH₄, either direct gas analysis (atmospheric gases), or analysis of gases liberated / created from sample heating / chemical processing (s) (e.g. conversion of organic compounds and other forms of carbon (e.g. carbonates) to CO₂ by combustion). Processes of atmospheric evolution, circulation and cycling (a); analyze gases trapped in rocks and soils (s); low temperature geochemistry; fluid processes, organic chemistry, formation temperatures; rock ages, surface exposure duration.</td>
</tr>
<tr>
<td>ESS</td>
<td>Environmental Sensor Suite</td>
<td>Univ. Leicester, Open Univ. (UK)</td>
<td>Atmospheric temperature (tp); total accumulated radiation dose; momentum and rate of impact of aeolian transported Martian dust (da); concentration of oxidising vapours in the atmosphere (a) at a discrete time; UV flux (a,da); wind horizontal speed and direction, air temperature and pressure (tp); atmospheric density during entry / descent.</td>
</tr>
<tr>
<td>XRS</td>
<td>X-Ray Spectrometer</td>
<td>Univ. Leicester (UK)</td>
<td>Primary goal of the XRS is to determine, in-situ, the geochemical composition, and by inference, the mineralogical composition and petrological classification, of the Martian surface material at the landing site (s)</td>
</tr>
<tr>
<td>MBS</td>
<td>Mössbauer Spectrometer</td>
<td>Univ. Mainz (D)</td>
<td>Identification of Fe bearing phases; oxidation state of iron bearing minerals; identification of Fe carbonates, sulphates, nitrates etc.; determination of Fe oxides and the magnetic phase in the Martian soil; detection of nanophase and amorphous hydrothermal Fe minerals that could preserve biological materials (s).</td>
</tr>
<tr>
<td>SCS</td>
<td>Stereo Camera System</td>
<td>MSSL (UK) CSEM (CH)</td>
<td>Stereo imaging; construction of a Digital Elevation Model (DEM); panoramic imaging (i); Multi-spectral imaging of rocks and soils to determine mineralogy (s); Observations of the sun (absorption from water vapour) (w); determination of atmospheric optical density and aerosol (dust and water ice) (i,da,w); Astronomical observations of Phobos &amp; Deimos (spectral characteristics); Observations of dust properties (da); Observe optical effects due to CO₂ ice crystals. Observe transitory or seasonal changes (dune migration (s), surface frosts (a,w), clouds, haloes, dust devils (tp)).</td>
</tr>
<tr>
<td>MIC</td>
<td>Microscope</td>
<td>MPI Lindau (D)</td>
<td>Study the physical and structural properties of Martian surfaces, geophysical analysis (s); image dust and surface material particles (da); characterize samples for analytical instruments; assist chemical analysis (s); identify bio-structures in samples.</td>
</tr>
<tr>
<td>PLUTO</td>
<td>Planetary Underground Tool</td>
<td>DLR Köln (D)</td>
<td>Serve as a soil sample acquisition device for GAP; perform in-situ temperature measurements (tp) as function of time and depth in the subsurface; investigate soil mechanical properties(s).</td>
</tr>
</tbody>
</table>

Tab. 2. – Beagle2 Lander payload description

the lander baseline planning with the ultimate goal of merging the coordinated science activities into the lander and orbiter science operations plan. The basic planning sequence is shown in Figure 1. After an initial request for input to the science teams, the payload providers submit proposals for coordination of operations / expression of interest in data from instruments on the other spacecraft. The Project Scientist Team at ESTEC processes their inputs in order to establish different categories of coordination and priorities for observations according to scientific output and priority with respect to operational aspects, requirements of proposed activities and the various constraints associated with the request. The detailed operations plans are then worked out in coordination workshops in order to achieve agreement among the instrument teams and for the incorporation of the results in the orbiter and lander operations plans.
4.2 Requirements and Constraints

In order to successfully coordinate the science activities of orbiter and lander, a number of constraints has to be taken into account. There are two major categories: general constraints include arrival date at Mars and related seasonal implications, lander communication opportunities, spacecraft capabilities and commissioning periods, lander lifetime and predictability of lander operations. Individual constraints are related to the individual instruments and their characteristics such as operating conditions and other requirements that need to be fulfilled for nominal operation of an instrument. Furthermore, there is a number of requirements for coordinating operations of instruments on the two spacecraft. General requirements include the compatibility with spacecraft capabilities, compatibility with the planning of commissioning activities, an early identification of the Beagle2 landing site and suitable opportunities for coordinated science operations. Requirements on the orbiter side may include operations in parallel to other systems like the MELACOM communications package, pointing requirements for investigating the landing site, operation within the spacecraft commissioning period, and timing of operations for concurrent observations. For the lander, the requirements may include the optimisation of sensor operations during orbiter passes, proper timing of PAW pointing or sample acquisition, and ensuring the coordination of science activities while maintaining operational flexibility.

4.3 Expected Results

It is expected that the science operations coordination efforts will lead to the following results:

- A well-defined plan for coordinated science operations, maximizing the science output while minimizing the impact on operations workload and spacecraft resources, embedded in the MSP and the Beagle2 operations planning
- Increased awareness of science activities performed by the partner spacecraft payloads among the instrument teams
- Optimised ground truth data from Lander payloads available to orbiter instrument teams, allowing improved modelling of the Martian environment and higher quality end data products
- A clear understanding of the landing site context with respect to origin and history, macro- and microclimate, dominant mineralogy and petrology
- Increased science communications opportunities as a result of using the orbiter and lander as an integrated science system

The achievement of these goals, and as a consequence the optimisation of the overall science output of the Mars Express mission, is a clear objective of the Mars Express Project Scientist team. It will lead to the utilisation of MEX orbiter and Beagle2 lander not as two separate spacecraft but as an integrated science tool for the exploration of the Martian environment.

5. CONCLUSIONS

The Mars Express orbiter and the Beagle 2 Lander carry a set of instruments, which in many cases measure the same or similar parameters of the Martian environment. The combination of remote sensing and in-situ investigations offers the possibility to achieve both sufficient coverage and ground truth for the space-based observations. The operations of the Mars Express orbiter are planned in advance and defined in the Master Science Plan. The lander operations will, after an initial phase, be planned on a day-by-day basis. A successful coordination of their science operations has to be compatible with both approaches. Planning and implementation of the coordinated operations is done by the MEX Project Scientist Team in cooperation with Principal Investigators and instrument teams.

The coordination of MEX and Beagle2 operations will lead to the utilization of MEX orbiter and Beagle2 lander not as two separate spacecraft but as an integrated science tool for the exploration of the Martian environment. If successful, it will maximise the scientific output of the Mars Express mission.

6. REFERENCES


2. Martin P., Preparation of scientific activities gears up for the ESA Mars Express Mission, presented at ESLAB 36, Noordwijk, NL, 2002
Session 3
Methods for Comparative Planetology
Chairs: H. Wänke & A. Chicarro
SURFACE MINERALOGY OF EARTH-LIKE PLANETS, MOONS, AND SMALL BODIES

P. D. Martin (1), T. B. McCord (2), P. C. Pine (3), M. S. Robinson (4), G. Schwehm (1)

(1) European Space Agency, ESTEC, Research and Scientific Support Department, Postbus 299, NL 2200 AG Noordwijk, The Netherlands, Email: Patrick.Martin@esa.int
(2) Hawaii Institute of Geophysics and Planetology, University of Hawaii, 2525 Correa Road, Honolulu, HI 96822, USA
(3) UMR562-CNRS, Observatoire Midi-Pyrénées, 14 Avenue Edouard Belin, F-31055 Toulouse, France
(4) Department of Geological Sciences, Northwestern University, Evanston, IL 60208, USA

ABSTRACT

Within the past 40 years, advances in ground-based remote sensing and significant return of spectral and compositional data from the orbit of a number of terrestrial and outer planetary bodies have led to a rapidly increasing knowledge of the surface mineralogy of solid objects in our Solar System. These results, combined with progress made in the fields of laboratory measurements and simulations, have allowed for better-constrained interpretation of planetary surface materials, with a particular emphasis on Earth-like planets and moons. The wealth and quality of returned data has permitted numerous comparative mineralogical investigations, hence allowing for comparisons of the evolution processes of the considered planetary objects. It is expected from future missions to bring back a substantial amount of additional compositional data, therefore enlarging the spectrum of already covered solid surface targets of our Solar System. To validate this data, advances shall be made in the area of analytical methods for gathering and analysing planetary samples.

The 0.2-12 µm domain (UV-IR) of the electromagnetic spectrum being essential for adequate identification of mineral and molecular species, the major technique used to identify surface constituents on planetary objects is reflectance spectroscopy. It measures the reflected solar radiation from surface elements and determines the fraction of incident sunlight absorbed as a function of wavelength by the surface materials. The interaction between the solar radiation and the surface materials leads to the presence of absorption bands in the reflectance spectra of the materials, the position and intensity of which is function of the nature of the molecular element and of the location of this element in the crystal. For more detail, the theory of reflectance spectroscopy is summarized in, e.g., [2,3].

Advances in remote sensing technology have led to the development of imaging spectroscopy or spectro-imaging (multispectral – a few spectral channels, or hyperspectral – many channels). It is now being used extensively for the purpose of acquiring simultaneous spatial/spectral information, identifying spectral surface units and analysing their extension, constraining the photometric properties, performing mineralogical mapping, and drawing conclusions on surface mineralogy and geologic evolution. The increase of both spatial and spectral resolutions is making in-situ geology complementary of spectro-imaging observations, with spectral resolutions now comparable to resolutions obtained in the laboratory.

Despite those technological advances that have been used to return quality data from a number of planetary objects, the Earth’s surface has not yet been properly investigated by space-borne instruments appropriate for the detection of mineralogical units at high spatial and spectral resolution. To improve the compositional view of the Earth as seen by the Galileo spacecraft in 1990, new global high-resolution hyperspectral data is expected to be returned in the near future by the first demonstrator satellite in orbit. Beyond Earth, very little is known about the mineralogy of the surface of Venus, optically opaque in the visible. This lack of information is due to poor sampling performed so far by telescopic means, and to the absence of data from


© European Space Agency • Provided by the NASA Astrophysics Data System
hyperspectral infrared spectrometers able to detect variations in the mineralogic properties of the surface within the gaps of the thick venusian atmosphere layer.

The major results obtained from telescopic or orbital spectroscopy and spectro-imaging investigations of solid planetary objects, including the relevant interpretations in terms of surface mineralogy and processes, are reviewed thereafter.

2. MULTISPECTRAL IMAGING DATA OF THE MOON AND MERCURY: POTENTIALITIES AND LIMITATIONS.

2.1 The Moon

Over the last two decades, telescopic Earth-based and Clementine orbital multispectral observations in the ultraviolet, visible, and near-infrared domains, with respective spatial resolutions of 500m-1km and 100-200m, have largely contributed to the spectral characterization of lunar surface units [e.g., in 3]. As a result, it provides with a first global deciphering of the lateral heterogeneity of the lunar surface and detection of the main geological units and boundaries. It greatly improves our knowledge and understanding of the impact cratering process, of the local and regional regolith optical and compositional properties, of the lunar crust structure and nature, of the emplacement of the mare, cryptomare and dark mantle units [e.g., 4, 5]. It also reveals that the importance of the early lunar volcanism [e.g., 6] and of the impact melt contribution may have been widely underestimated in the previous descriptions of the crustal layering and structure [e.g., 7]. As a general consequence, lateral and vertical mixing processes having occurred along the lunar history can be better quantified [e.g., 8].

The current datasets are quite efficient for the purpose of geological mapping and detection at subpixel scale of specific targets of interest. These targets are impact ejecta material, linear units (e.g., dykes), lava flows, impact melt sheets on crater walls and lithological variations with depth (e.g., see the 10-channel UV-VIS multispectral coverage of Copernicus crater [9]). However, these datasets present serious limitations for the purpose of rock petrology identification, detailed stratigraphic chronology and geological interpretation. Indeed, hyperspectral imaging spectroscopy is clearly required for rock identification. Furthermore, the integration of optical spectroscopy data and geochemical data derived from nuclear spectroscopy (gamma-ray, neutron, X-ray) observations is clearly a direction to be emphasized [e.g., 10]. With the expected improvements in spatial resolution, optical heterogeneities already highlighted by Clementine in relation with the photometric and physical properties of the regolithic surface (texture, surface roughness, maturity, degree of crystallinity) at subpixel scale will have to be taken into account.

Among the major outcomes is the deconvolution from the optical local variations between soil maturity and composition effects. Iron, occurring in the form of metallic iron and Fe²⁺ ion in minerals and glasses, dominates the broadband reflectance properties of the lunar surface. Reflectance properties observed in lunar spectra are sensitive to the mineralogical and elemental abundances and to the physical processes, called soil maturation, occurring in the lunar regolith with exposure to solar wind and micrometeorite bombardment. Lunar surface materials darken, redden and lose spectral contrast with increasing maturity. Increasing the Fe²⁺ abundance in the surface materials decreases the reflectance of minerals and increases the spectral contrast. Disentangling these respective contributions due to soil maturity and iron content has been a central issue for long in lunar spectroscopic studies. In the last years, a quantitative reliable method has been proposed [11, 12, 13] which can be applied to the Clementine spectral dataset and led to the production of global compositional maps as shown in Fig. 1. Based on the examination of a suite of Apollo lunar soil samples, one can empirically separate both effects (soil maturity, iron content) in the lunar soil spectral properties in a rather simple way. Integrating the UV-VIS/NIR dataset has resulted in an improved discrimination between maturity and composition [14].

Fig. 1: Global map of iron composition of the Moon, derived from Clementine data [e.g., 12].

Since previous studies suggested that the anomalous surface reflectance variations observed at Reiner Gamma Formation (RGF) might be ascribed to a combination of both optical and compositional effects occurring in the regional mare regolith, a detailed systematic regional study has been made, combining Lucey's approach and spectral mixture analysis [13]. It reveals the existence of 3 basic endmembers relevant for modelling the observed distribution of spectral variations in the RGF vicinity (Fig. 2). The first two components exhibit spectral characteristics consistent with a prevailing contribution of mature mare soils for the surroundings (MB) and of immature mare crater-like soils (RGS) at RGF. The third intermediate-albedo
component (SWS) has general characteristics of a mature mare soil, but with a redder continuum slope. The reported observation can be modelled by a mechanism removing the finest fraction in the soil at RGF and redistributing it in the vicinity with a lateral variable proportion and local accumulations such as at SWS site. According to the available set of in-situ data documenting variations in the chemical composition, in the distribution of particle sizes, and in the degree of maturity with depth in the mare regolith, the characteristics depicted at RGF are those of a subsurface soil layer from a depth on the order of 0.3-0.8 m. This case points out the occurrence at local scale of unusual photometric and spectral behaviors associated with the lunar surface regolith and its scattering properties [e.g., 16].

Fig. 2: Reiner Gamma Formation. From left to right: Fraction images of MB (mare background), SWS (south west swirl), and RGS (Reiner Gamma soil) spectral endmembers. Red is 100% similarity to the considered endmember, blue is 0% similarity.

Their exploration can be addressed with high-resolution stereo-imaging or photolinometry, generating local digital elevation models, and with the production of spectral datasets under various geometry conditions of observation, giving access to the photometric properties of the surface. In turn, geomorphologic studies will benefit of this high-resolution data and geologic interpretation of the spectroscopic information will be significantly improved. This new data is within the reach of the upcoming missions Lunar-A, Smart-1 and Selene. A particular effort is however needed in the field of experimental imaging and spectrophotometric studies to produce extensive databases documenting the various physical and mineralogical/compositional mixing processes involved at mesoscale in an integrated remote-sensing observation made at a macroscopic scale. It goes along with the ongoing work undertaken on the basis of the lunar samples collection and aiming at improving the links among lunar soil mineralogy, chemistry and reflectance spectra for the purpose of predicting reliable composition estimates [17].

2.2 Mercury

Mercury is the least explored of the terrestrial planets, being visited by only one spacecraft flyby nearly 30 years ago [18]. Despite the lack of detailed spacecraft observations of Mercury, new discoveries have been made recently through Earth-based telescopic and radar observations as well as reprocessing of Mariner 10 image data. Mercury is known to have an unusually high uncompressed bulk density demanding an oversize core mostly composed of iron metal (relative to its radius). Thus Mercury’s bulk iron content is high relative to the other terrestrial planets, however it appears to have a FeO-depleted crust and mantle. Analysis of Earth-based spectral measurements (discarded) and Mariner 10 two-color image data [19] indicates that Mercury’s crust has compositional affinities with the lunar anorthositic crust albeit with lesser FeO content (~3 wt%). An important, and related, finding from the Mariner 10 data is that volcanically emplaced smooth plains deposits do not have low albedo (indicating paucity of iron [20]), and do have color properties equivalent to Mercury’s disc-resolved average color (see Fig. 3) [21]. These data indicate that not only is Mercury’s crust deficient in FeO but that its upper mantle and thus probably whole mantle is also low in FeO – a composition consistent with very high temperature equilibrium condensation in the early solar nebula. One might even argue that Mercury’s silicate FeO content is high from the expected temperatures during Mercury’s accretion [22], however the FeO was possibly condensed during the waning stages of accretion when the solar nebula had cooled to the point where FeO could condense as well as through material added from impactors formed at greater AU than Mercury. Understanding the geochemistry and internal structure of Mercury must await detailed orbital measurements from Messenger (launch in 2004) and BepiColombo (launch in 2011).

Fig. 3: Left: Color/albedo variations on Mercury from reprocessed Mariner 10 color data. Right: Color data of smooth plains at Rudaki and Tost. Smooth plains are morphologically similar to lunar mare.

3. MARTIAN MINERALOGY AND SURFACE PROCESSES: TOWARDS A NEW UNDERSTANDING OF THE SURFACE DIVERSITY.

Weathering and alteration processes acting on Mars over geologic timescales cause variations in the
chemical composition, physical properties, and mineralogy of the soils, dust, and rocks (e.g., degree of crystallinity, cation content, oxidation state, grain size, and abundance of water). See detailed reviews in [3,23]. Based on laboratory studies [e.g., 24,25,26], early ground-based and orbital (Viking) observations concluded on basalts with varied degrees of oxidation and grain sizes to be the major constituents of the dark terrains, and on iron-rich alteration products of mafic volcanic glass for the bright materials [e.g., 27,28]. More recent ground-based studies concluded on the presence of hematite-like crystalline ferric mineral phases occurring in minor amounts within the bright terrains [e.g., 29]. From ISM/Phobos-2 data (see Fig. 4), the compositional differences of the bright soils were ascribed to varied histories of chemical alteration instead of simple mixings with darker basaltic materials [30]. Darker terrains exhibit absorption bands characteristic of unweathered volcanic minerals (pyroxene, feldspar, and minor olivine), and present spectral variations which may result from partial coatings by fine materials or from mixing with highly altered dust and soil [31].

Fig. 4: Representative reflectance spectra of martian bright and dark terrains, derived from orbital ISM data. Left: Bright-region spectra compared with laboratory spectra of ferric minerals [30]. Right: Dark-region spectra showing shift towards longer wavelengths [31].

The latest results from the Thermal Emission Spectrometer (TES) onboard Mars Global Surveyor include basaltic toandesitic variations of the volcanic surface materials [32], and aqueous mineralization that occurred in limited regions under ambient or hydrothermal conditions. The detection of gray, crystalline hematite in sedimentary rock formations provides evidence for the long-term stability of liquid water near the surface of Mars [33]. No evidence for carbonates has been found. The lack of evidence for chemical weathering of the martian surface may indicate a geologic history dominated by a cold, dry climate in which mechanical weathering was the dominant form of erosion. As evidenced by results from Mars Pathfinder [e.g., 34,35], much remains to be understood about the precise identity and form of the surface constituents, and more work is required to constrain compositional heterogeneities and physical, weathering/alteration, and climatic processes on Mars. Current and future instruments (Themis, HRSC, Omega, Crism) should, with a better spatial and/or spectral coverage, permit to better constrain the mineralogical inferences, provide additional information about the detection of phyllosilicates, hydroxylated minerals, carbonates, palagonitic-like materials, and maybe decorrelate the respective contributions of the surface and atmosphere.

4. OUTER PLANETS AND THEIR SATELLITES: SURFACE COMPOSITION OF THE MAJOR SATELLITES OF JUPITER AND SATURN.

Knowledge of the surface composition of the large satellites of Jupiter was vastly improved by the Galileo mission. In the past 30 years, electronic detectors and large telescopes have allowed planetary astronomers to obtain infrared reflectance spectra showing that water ice is an abundant surface constituent and that hydrated and/or hydroxylated minerals of some sort are present [36,37,38]. However, the satellites’ small angular size as seen from Earth has restricted Earth-based measurements essentially to hemispherical averages.

Recently, ground-based and Earth-orbital telescopes were used to detect or infer O₂, O₃ and sulfur species [39]. Sulfur was detected in variable amounts on the trailing side of Europa and Ganymede from IUE data. On Callisto, SO₂ was reported from IUE and HST on the leading Jovian quadrant. Ground-based observations detected absorptions ascribed to pairs of O₂ molecules on Ganymede’s trailing hemisphere. Later, HST observations showed the oxygen concentrated at lower latitudes. O₃ was suggested from IUE observations, and HST measurements showed O₃ concentrated on Ganymede’s trailing hemisphere relative to the leading hemisphere. Formation of the several oxygen species probably is due to radiolysis of ice. The sulfur may have originated from Io but could be indigenous as well. All these species are ultimately due to effects of Jupiter’s magnetosphere.

The Voyager spacecraft cameras obtained striking images in several visible wavelength passbands of large areas of the icy satellites with spatial resolution in some areas down to 0.5 km. These images revealed structural and geological provinces suggesting past if not present intense tectonic activity on Europa and Ganymede. The multispectral images revealed correlations of color with geologic units but the composition of the material could not be identified except possibly through geologic inference. A UV-absorbing stain, detected by ground-based and IUE telescopes, was spatially resolved to be centered on the trailing side of Europa with a sinusoidal distribution.
from the center [40]. This is almost surely superficial and exogenic in nature and has been attributed to an S-O interaction due to S" from Io implanted into the icy surface by the Jupiter magnetosphere. On the Galileo mission, the Near Infrared Mapping Spectrometer (NIMS) was the major contributing instrument for the Jupiter satellites, although the Ultraviolet Spectrometer (UVS) also provided important information [41,42,43]. On Cassini/Huygens, the Visual and Infrared Mapping Spectrometer (VIMS) will provide similar information for the Saturn satellites and rings.

On Io, sulfur dioxide is common, and sulfur polymorphs are inferred from visible colors. Silicates are suggested near active volcanoes from color and temperature measurements from Galileo and ground-based IR data. Water ice is the major constituent on Europa (Fig. 5) and Ganymede, and water ice with an OH-bearing material, probably phyllosilicates, covers most of Callisto. Water ice on all icy satellites is present in a variety of grain sizes and crystalline states, due to sputtering and redistribution of water by the particle irradiation and to the surface temperature variations, e.g., finer-grained ice at the poles and more amorphous ice on Europa [44].

![Fig. 5: Representative average NIMS spectra of Europa (Global 1 and Terinc observations) [37]. This example shows the variation from deep water ice absorption bands (icy) to shallow asymmetric bands (non-ice).](image)

Heavily hydrated minerals, most probably MgSO₄ with some Na₂SO₄ from endogenic processes [45,58] and perhaps some H₂SO₄ [46] from radiolysis of these sulfates, are present in the disrupted regions of Europa (e.g., asymmetric bands in non ice spectrum of Fig. 5) and at some dark regions on Ganymede [47,48]. The salt minerals are probably due to materials surfacing from salty liquid layers below by one of several possible geological mechanisms.

A number of important molecules have been detected in small amounts by Galileo on the three outer satellites. These are probably intrinsic to the satellite or due to infall, and some are due to radiolysis of water ice and perhaps carbon. These include CO₂, SO₂, CN, CH₃OH [49,37]. CO₂ has been detected escaping from Callisto [50], implying an active replenishment process. CO₂ is the only molecule detected on all three icy satellites and its extent has been mapped on Callisto and Ganymede [37,51,52]. However, its source and origin have been difficult to determine, although at least some CO₂ seems resident in the darker material and may be due in part to radiolysis of carbon. The SO₂ is present widely and is not uniformly distributed, e.g., concentrated on the trailing side of Callisto although existing everywhere [37,51]. It is probably due to the redistribution of sulfur and SO₃ from Io. Organic molecules CN and CH have been detected on Ganymede and Callisto [49,37], but are present in too small quantities to be mapped from existing data. The origin of these molecules is not determined but they are probably common in outer solar system materials similar to tholins and are present in interstellar ice grains. H₂O₂ was identified by NIMS on Europa [53] and is probably a water ice radiolyis product, as is the several forms of oxygen molecules detected by ground-based telescopes on Ganymede.

So far, only ground-based telescope spectra of Saturn’s satellites exist, indicating that water ice is the major constituent on the surfaces in most cases. Iapetus has a dark, red deposit on one side, perhaps due to simple organics, and Titan of course has an atmosphere with methane and other reduced molecules, probably seriously affecting the surface composition. Evidence of resurfacing on some of Saturn’s icy satellites from Voyager images suggests that Cassini/ Huygens will discover other constituents.

## 5. COMETS AND ASTEROIDS IN THE SOLAR SYSTEM: KEY RESULTS, PERSPECTIVES FOR UPCOMING MISSIONS.

### 5.1 Comets

Results from ground-based observations concluded that cometary surfaces are composed of various water ice mixtures and of water ice mixed with non-ice materials [e.g., 54]. However, the detailed mineralogy and spatial distribution of these materials are still to be constrained. Orbital missions, such as CONTOUR (Comet Nucleus Tour), a NASA discovery mission that will flyby at least two comets, and ESA’s Rosetta mission that will rendez-vous with comet 46 P/Wirtanen, will carry visible and infrared mapping spectrometers. These instruments will allow for better identification of the different ices and ice mixtures, as well as for the characterization of their spatial distribution on the comet nucleus.
A key objective is to identify carbonaceous materials, which probably are mixed with water ice, even in low percentages. The determination of the overall continuum slopes of the reflectance spectrum will provide indications of the presence of organic compounds, as these will "redden" the spectra in a rather diagnostic manner between 0.4 and 1.1 µm. Measuring the spectrophotometric phase curve with high accuracy will determine the physical microstructure and nature of the surface grains. Also, future visible and infrared mapping spectrometry data will provide the means to identify and monitor active areas on the comet nucleus and establish relationships with possible small-scale compositional variations or with the overall mineralogical composition.

5.2 Asteroids

Asteroid science has recently been dramatically improved by the results from the NEAR Shoemaker spacecraft which orbited 433 Eros for a year, returning a wealth of image, spectral, ranging, magnetic, tracking, and compositional data. It revealed that Eros is an undifferentiated body with compositional affinities to LL or L Ordinary Chondrite (OCs) meteorites [e.g.,55,56,57,58]. Specifically, the Near Infrared Spectrometer (NIS) data show that all of Eros falls within the S-Type S-IV to S-II field in band area plots indicating an orthopyroxene to olivine ratio of ∼40:60 [55], a ratio consistent with OCs. The X-Ray Spectrometer (XRS) and Gamma-ray Spectrometer (GRS) compositional measurements are also consistent with an OC composition, in particular the Mg/Si ratio (≈0.8) demands an undifferentiated body [58]. An interesting detail of the XRS compositional measurements is an apparent depletion of S relative to OCs. One interpretation is that Eros is partially differentiated, however a simpler explanation that matches the overall remote sensing picture that Eros consists of an OC composition involves devolatilization of the surface [e.g., 58]. The NIS and Multi Spectral Imager (MSI) subtle color/spectral differences mapped on the asteroid’s surface [55,56,59] support a surface affected by exposure to the space environment – space weathering. The XRS senses only the top few centimeters of the surface, thus measuring those portions most susceptible to high-energy micro-meteorite impacts leading to the hypothesis that the S depletion is an artifact of space weathering. The color and spectral data present no evidence for compositional heterogeneity – further evidence that Eros is an undifferentiated primitive body [55,56,59]. Most surprising is the revelation that Eros has a complex surface geology characterized by a thick regolith, compressional and extensional tectonics, slopes averaging 10°, and never before seen sedimentary processes. The highly mobile global regolith could disguise any subtle compositional heterogeneity that might exist below the surface through horizontal mixing. However high resolution color data (5 m/pixel) reveal no evidence (such as boulders with anomalous color properties) to support such a scenario.

The NEAR mission met all its science goals and returned an order of magnitude more data than was originally planned, giving us our first detailed look at an asteroid confirming long held beliefs and revealing real surprises. NEAR data has shown that future sample return missions will face a simpler task than previously imagined, that many places on the asteroid represent benign landing sites, that there is a complete range of grain sizes easily collected by a scoop or arm (sub-cm to 10s of cm), negating the necessity of complicated drilling or chipping mechanisms.

6. CONCLUSIONS

Significant data has been returned and used for mineralogical interpretations and comparative planetology, with many of the results showing similarities between various bodies of the Solar System, hence allowing for comparisons of their evolution processes. Future planetary missions will bring back a substantial amount of additional compositional data, therefore enlarging the spectrum of already covered solid surface targets of our Solar System. To validate data from upcoming missions, advances shall be made in the area of analytical tools and methods for gathering and analysing samples. A particular effort is needed in the field of experimental imaging and spectrophotometric studies to produce extensive databases documenting the various physical and mineralogical/compositional mixing processes involved at mesoscale in an integrated remote sensing observation made at a macroscopic scale.

7. REFERENCES


26. Morris, R.V., et al., Spectral and other physicochemical properties of submicron powders of hematite (α-Fe₂O₃), maghemite (γ-Fe₂O₃), magnetite (Fe₃O₄), goethite (α-FeOOH), and lepidocrocite (γ-FeOOH), J. Geophys. Res., 90, B4, 3126-3144, 1985.


LABORATORY AND SIMULATION STUDIES

L. Colangeli(1), R. Battaglia(1,2), J.R. Brucato(3), V. Della Corte(1), A. Esposito(1,3), F. Esposito(1), G. Ferrini(1,2), E. Mazzotta Epifani(1), V. Mennella(1), E. Palomba(1), P. Palumbo(1), A. Rotundi(1)

(1) INAF - Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli - Italy, E-mail: colangeli@na.astro.it
(2) Università "Federico II", Piazzale Tecchio, 80100 Napoli - Italy,
(3) Università "Pathenope", Via A. De Gasperi 5, 80133 Napoli - Italy

ABSTRACT

The study of Solar System bodies by remote and in situ measurements provides information about geological, chemical and physical properties of materials present in different space environments. Processes (e.g., thermal annealing, UV irradiation, ion bombardment, gas-solid interactions) contribute with different efficiencies to determine the characteristics of materials, depending on both the local environment properties and on the sensitivity of species to active mechanisms. Identifying the properties and tracing the evolution of compounds may tell us the past history and present status of the Solar System and shed light on the different behaviour of planetary bodies.

A thorough interpretation of the data acquired in situ and remotely is based on the fundamental contribution coming from simulations in laboratory. Experiments are aimed at reproducing environmental conditions, at studying properties of species observed in different environments and at simulating their evolution according to mechanisms active in space. Large scale installations are suitable to provide proper conditions for the test and calibration of systems and experiments used in space missions. Analogues of organic and refractory cosmic compounds are synthesised in laboratory, with controlled chemical and physical status, and studied by different analytical techniques. The application of treatments, similar to those active in space, to well characterised samples provides information both on the reactivity of materials and on the efficiency of processing.

The comparison of laboratory results with data coming from space measurements provides a powerful tool to understand the real nature of cosmic materials and, therefore, to place constraints on their actual evolution in space. This also in the view of a potential bio-genetic evolution of materials.

1. LARGE SIMULATION FACILITIES

Large scale simulation facilities are generally set and used to reproduce at best space and/or planetary environmental conditions. In this way it is possible to develop and calibrate sensors and flight systems used or to be employed in planetary missions. These experiments combine scientific and technological objectives. From one side it is possible to demonstrate the capabilities of new generation instruments to study remotely or in situ targets of the Solar System. On the other hand, data retrieved during tests with bread-boards or copies of flight instruments allow us to trace their behaviour as a function of various boundary conditions and on different test-samples. The data acquired under simulated space conditions offer a precious reference frame for the interpretation of data collected in space and for the optimisation of the scientific return from them.

Only some examples of large installations available in various laboratories are here mentioned with the aim to testify the efforts by different groups to achieve a proper simulation of space conditions.

The Planetary Simulation Chamber available at DLR - Cologne (Fig. 1) offers an internal experimental space of 1.4 m diameter and 1.8 m height, is equipped with a LN2 cooling system (77 K) and reaches vacuum of 10⁻³ mbar. It is suitable to simulate Martian, as well as other planetary environmental conditions.

The Andromeda Planetary Environmental Chamber is available at the Arkansas - Oklahoma Center for Planetary Sciences and is dedicated to study dynamic processes on asteroids, comets, and planetary bodies. It has been already used for experiments on the viability of methanogens in Martian regolith simulant under near-Martian atmospheric conditions. The chamber, shown in Fig. 2, is a cylinder of 0.8 m in diameter and 1.5 m tall. It can be evacuated to < 10⁻² Torr and filled with various atmospheric mixes (e.g., CO₂ dominated), while the internal temperature can range from -180 to +100 °C. It is planned to equip it with a high-powered Xe lamp to produce 1 solar constant, with filters to simulate Martian atmosphere, and with X-ray and β radiation sources. A robotic system shall be implemented for sample manipulation.

demonstrated by results obtained recently thanks to their use. Merrison et al. [1] have studied the capture of magnetic dust in simulated Martian aerosol by the facility in Aarhus, to interpret the results obtained during the Pathfinder mission about the capture on permanent magnets exposed to the Martian dusty environment. Hints about the magnetic properties of Martian dust (e.g., hematite content) have been obtained. Similarly, the group of Naples are presently using their facility to determine the performances and sensitivity of quartz crystal micro-balances to dust and water vapour deposition in Martian atmosphere.

Fig. 1. The Planetary Simulation Chamber at DLR - Cologne.

Fig. 2. The Andromeda Planetary Environmental Chamber at the Arkansas - Oklahoma Center for Planetary Sciences.

While the facilities mentioned above offer the possibility to set various environmental conditions, other simulation chambers are dedicated to specific tasks. Examples of this category are the Wind Tunnel and Atmospheric Simulation Chamber at the Aarhus University - Denmark (Fig. 3) and the Martian Atmospheric Simulation Chamber at the INAF - Osservatorio Astronomico di Capodimonte Napoli - Italy (Fig. 4). The former facility is dedicated to simulate the gas and dust flow under simulated Martian atmospheric conditions, while the latter is aimed at reproducing the detailed atmospheric Martian composition (e.g., water vapour content). The scientific and technical relevance of such chambers is

Fig. 3. The Wind Tunnel and Atmospheric Simulation Chamber at the Aarhus University - Denmark.

2. SMALL SCALE EXPERIMENTS

Laboratory experiments on smaller scales have objectives complementary to those achievable by large scale installations. The main tasks are to reproduce properties of planetary materials and to study / simulate processes active in space environments. Experiments performed in this field are often conducted on terrestrial samples, that are considered analogues of space materials, or on compounds condensed/produced ad hoc by laboratory synthesis methods (e.g., by chemical reactions, vaporisation/condensation, laser ablation).

A typical example of the application of results obtained in laboratory experiments is the interpretation of remote and in situ spectroscopic measurements. The laboratory data can be used for direct comparison with observed spectra or as input in computational codes to synthesise spectra starting from optical constants of selected material. The common feature of this approach is that laboratory data must be obtained on carefully chosen and well characterised materials.
Several materials must be studied to cope with the need of interpreting observations for a wide variety of space environments: carbon-based, silicates, carbonates (especially for Mars), mixed materials. The laboratory program is aimed at tracing the material evolution versus processes simulating transformations expected in space (e.g., thermal annealing, UV and ion irradiation, atomic beam irradiation, laser amorphisation). A thorough characterisation of the samples requires the use of several analytical techniques. Scanning (SEM) and transmission (TEM) electron microscopy are used to study aggregation, morphology, size distribution, internal structure of samples, especially when in dusty form. Electron diffraction is sensitive to crystalline degree, while electron energy loss spectroscopy (EELS) evidences electronic properties. Energy dispersive X-ray (EDX) analysis provides the elemental composition. Spectroscopy in various ranges is sensitive to different features of materials: electronic transitions (far UV - UV), optical gap (visible), molecular resonance and bonds (IR), inner structure (FIR).

Table 1. Selection of natural terrestrial minerals and rocks potentially relevant as Martian analogues.

<table>
<thead>
<tr>
<th>Class of materials</th>
<th>Representative members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonates</td>
<td>Calcite, Dolomite, Siderite,</td>
</tr>
<tr>
<td></td>
<td>Magnesite, Aragonite</td>
</tr>
<tr>
<td>Clays</td>
<td>Kaolinite, Montmorillonite, Sillimanite,</td>
</tr>
<tr>
<td></td>
<td>Nontronite, Schwertmanite, Ferrhydrite</td>
</tr>
<tr>
<td>Feldspars</td>
<td>Albite, Anorthite, Labradorite,</td>
</tr>
<tr>
<td></td>
<td>Orthoclases</td>
</tr>
<tr>
<td>Hydrous Carbonates</td>
<td>Artinite, Gaylussite,</td>
</tr>
<tr>
<td></td>
<td>Hydromagnesite, Dyingsite</td>
</tr>
<tr>
<td>Igneous Rocks</td>
<td>Andesite, Basalt, Palagonite</td>
</tr>
<tr>
<td>Iron Oxides</td>
<td>Hematite, Limonite, Chromite,</td>
</tr>
<tr>
<td></td>
<td>Magnetite, Lepidocrocite, Goethite</td>
</tr>
<tr>
<td>Nitrates</td>
<td>Niter, Nitratite, Nitrocalcite,</td>
</tr>
<tr>
<td></td>
<td>Niromagnesite</td>
</tr>
<tr>
<td>Olivines</td>
<td>Fayalite, Forsterite</td>
</tr>
<tr>
<td>Phosphates</td>
<td>Merillite, Whitlockite, Apatite</td>
</tr>
<tr>
<td>Pyroxenites</td>
<td>Enstatite, Augite, Pigeonite</td>
</tr>
<tr>
<td>Sulfates</td>
<td>Gypsum, Jarosite, Kieserite, Anhydrite</td>
</tr>
<tr>
<td>Others</td>
<td>Quartz, Maghemite, Pyrite,</td>
</tr>
<tr>
<td></td>
<td>amorphous Fe-rich clay</td>
</tr>
</tbody>
</table>

The variety of different materials of potential relevance for planetary applications is so wide that only long term and systematic studies can provide the results required to interpret observational data. As an example, a list of terrestrial analogues of potential relevance for Martian studies is reported in Table 1. Despite the availability of several data banks, the laboratory results obtained so far are often not adequate to be compatible with the characteristics (e.g., spectral resolution) of instruments in use or of future use in planetary exploration missions or do not account in detail of several important physical factors (e.g., granulometry of dusty samples). Therefore, new and more accurate laboratory experiments are required.

![Fig. 4. A scheme (top panel) and a picture (bottom panel) of the Martian Atmospheric Simulation Chamber at the INAF - Osservatorio Astronomico di Capodimonte Napoli - Italy.](image)

In the following, some results of laboratory experiments are reported to illustrate the variety of approaches that can be followed in the context of interpreting planetary measurements.

Very often, to simulate planetary dust, natural minerals are ground and selected in dimension. The SEM image of crystalline forsterite grains is reported in Fig. 5.
Fig. 5. Scanning Electron Microscopy image of ground crystalline forsterite grains.

To characterise grain morphology, SEM images are analysed in order to derive the statistical distribution of sizes (see Fig. 6).

Fig. 6. Size distribution of crystalline forsterite grains ground and selected with sizes smaller than 5 μm.

The elemental composition derived by EDX analysis confirms that the sample falls in the Mg-rich class of olivine silicates (Fig. 7). The infrared diffuse reflectance spectrum of the sample is shown in Fig. 8.

Diffuse reflectance, measured in laboratory for analogues of planetary compounds, can be applied for direct comparison with remote or in situ planetary spectroscopic in the portion of the spectrum where surface reflectance of solar radiation dominates. The same laboratory data can be used to retrieve the emissivity, via the Kirchhoff's law, to be compared with the planetary emission. As demonstrated by various works, the size of the grains drastically influences the profile of the diffuse reflectance spectrum and the contrast of the features (see, e.g., [2]). Specular reflectance data are useful to derive optical constants to become inputs of surface and atmosphere simulation models.

Fig. 7. Elemental composition obtained by EDX analysis of the forsterite sample. The Mg dominance is confirmed.

Fig. 8. Diffuse reflectance spectrum of ground forsterite grains with sizes < 5 μm.

A key subject of analysis on planetary relevant materials is the amorphous-to-crystalline transition mechanism and the definition of environmental boundary conditions suitable for such modification to occur. Since natural materials on Earth rarely occur in amorphous state, laboratory synthesis is required to produce species in this form. The SEM image of an amorphous fayalite (Fe-rich olivine) dust sample produced in laboratory by laser ablation of a crystalline target is reported in Fig. 9. The process is performed in a O₂ atmosphere at 10 mbar.

For amorphous materials it is interesting to study the evolution of spectroscopic properties under thermal

© European Space Agency • Provided by the NASA Astrophysics Data System
processing. An example of the spectral variations for fayalite grains under thermal annealing in vacuum is reported in Fig. 10 [3]. It is evident that the broad bands at 10 and 20 μm, typical of amorphous silicates, become a series of sharp features as long as the amorphous-to-crystalline transition proceeds. The final spectrum is similar to that of the crystalline target.

Fig. 9. SEM image of amorphous fayalite grains obtained by laser ablation (in O₂ atmosphere - 10 mbar) of natural crystalline fayalite mineral used as target.

Fig. 10. Spectral evolution of the IR spectrum of amorphous fayalite (FAYA) grains at 800 °C for different times [3]. The spectrum of crystalline fayalite (FAYC) grains is reported for comparison.

From the analysis of the spectroscopic results, the so-called activation energy for the amorphous-to-crystalline transition can be derived [3,4]. This parameter poses limits on the time and temperature conditions needed for the transition to occur. Therefore, it places constraints on the planetary (and interplanetary) environmental conditions suitable to allow the structural modification of amorphous species towards a more ordered status.

The relevance of the amorphous-to-crystalline transition is of stringent importance, e.g., in the view of the clear identification of crystalline silicate features in IR emission spectra of comets [5,6]. We have to consider that: i) comets are considered to be formed by pristine and rather unprocessed proto-planetary materials and ii) interstellar matter (from which proto-planetary disks are accreted) showed so far no evidence of crystalline silicate features [7,8]. Thus, turbulent radial mixing in the proto-solar nebula [9] and/or annealing of dust by nebular shocks [10] have been invoked to explain the presence of crystalline grains in comets.

Finally, it should be stressed that well characterised analogue samples are relevant not only to interpret remote / in situ planetary measurements, but also to study performances and calibrate sensors / instruments devoted to planetary missions. A typical example is represented by the ESA Rosetta mission, aimed at a rendezvous with comet 46P/Wirtanen. Several instruments included in the scientific payload will study dust properties or will perform spectroscopic characterisation of the nucleus. Many of these experiments have benefited of the use of "cometary dust analogues" as reference samples during the characterisation of performances and calibration phases. This approach is expected to become a common feature of many of the future planetary missions.

3. MICRO-ANALYSES ON EXTRATERRESTRIAL SAMPLES

In the scenario of laboratory studies related to planetary exploration, a key role is played by investigation on extraterrestrial materials. It is beyond the scope of this paper to review the vast field of experimental approaches applicable to samples such as meteorites and interplanetary dust particles.

Recent measurements have been devoted to explore the intimate nature of meteorites. The aim is to identify the different forming components in terms of mineralogy and structure, mainly. This kind of analysis may shed light not only on the composition of parent bodies (e.g., Mars and asteroids), but also about the inter-mixing of minerals, their genesis and past evolution.

The techniques to be used in laboratory for this purpose must be suitable to analyse the properties of materials at micro-scales and must be as much as possible non-destructive. This last requirement derives from the need to guarantee a complete characterisation of small amounts of extraterrestrial samples by a variety of
techniques. Infrared (600-5000 cm\(^{-1}\)) micro-spectroscopy at high (4 cm\(^{-1}\)) spectral and moderate (20-
400 \(\mu\)m) spatial resolution is used in the laboratory of Naples in combination with elemental analysis by EDX
and FESEM (Field Emission Scanning Electron Microscopy) at spatial resolution of about 2 nm. These
techniques allow us to achieve elemental, chemical and morphological characterisation without altering the
sample properties.

The FESEM micro-graphs of the Murchison chondritic meteorite reported in Fig. 11 show the importance of a
detailed morphological characterisation at micro-scale: the chondrules included in a rather uniform matrix are
well evident. Infrared micro-spectroscopy demonstrates that observed inclusions are mainly olivine and enstatite
in composition.

Fig. 11. FESEM images of the Murchison meteorite. The bottom panel displays one chondrule as many
observed in the matrix (top panel).

Micro-analyses have a key role in the characterisation of SNC (Shergotty-Nakhla-Chassigny) meteorites, which
are considered of Martian origin [11]. Their study in laboratory is of great relevance to determine Mars
mineralogy, in combination with results from remote
and in situ planetary measurements. Palomba at al. [12]
have coupled elemental analysis by EDX and infrared
micro-spectroscopy to identify the distribution of main
mineralogic components in a fragment of the Zagami
meteorite. The EDX analysis performed on different
areas of the sample shows the presence of three main
classes of minerals: maskelinite, Fe-rich piroxene and
Mg-rich pyroxene, in agreement with previous results
[13,14]. The infrared micro-spectroscopy on the same
areas shows features typical of the different components
perfectly matching with their attribution.

4. CONCLUSIONS AND FUTURE PERSPECTIVES

The subjects discussed in the present paper demonstrate
that laboratory is a place for both doing science and
testing instruments, in the perspective of studying
planets.

Laboratory installations are suitable to simulate space
environmental conditions, to reproduce material
properties and processes and to study extraterrestrial
samples. The methods and techniques to be applied
differ according to the specific tasks to be achieved and
range from macro- to micro-scales. All experiments
have, however, two main common goals:
- to develop and calibrate sensors and flight systems;
- to produce data to interpret space measurements.

The lines of approach summarised in this paper have a
promising future, as new and more ambitious space
missions will be devoted to explore different planetary
environments by remote and in situ measurements. Only
thanks to a proper preparatory laboratory work these
missions will be really successful.

A remarkable line of development is represented by the
preparation to analyse samples that will be returned to
Earth from different planetary environments. It is not far
the time when "bricks" from comets, asteroids and
planets (Mars, in particular) will be made available to
the scientific community involved in laboratory
experiments. The problems that will have to be tackled
in order to retrieve meaningful information are mainly
related to contamination and preservation of such
precious samples. Efforts for the development of a new
generation of experimental techniques must be started
since now to be ready for the future important
appointments.

5. REFERENCES

1. Merrison J., Gunnlaugsson H., Mossin L., Nielsen J.,
Nørnberg P., Rasmussen K., Uggerhøj E., Capture of
magnetic dust in a simulated Martian aerosol: the


SCIENCE OPERATIONS OF ESA PLANETARY MISSIONS

Detlef Koschny, Andrea Marini, Raymond Hoofs, Miguel Almeida

European Space Agency, ESTEC, Keplerlaan 1, 2201 AZ Noordwijk ZH, The Netherlands, Email: Detlef.Koschny@rssd.esa.int

ABSTRACT

In the year 2003 the European Space Agency (ESA) will launch three spacecraft to targets in our solar system: Rosetta to a comet, Smart-1 to the Moon, Mars Express to Mars. BepiCOLOMBO, to go to Mercury, is in the planning stage. All of these missions have similar requirements to science operations. This paper discusses the peculiarities of these missions and a generic approach to the way the science operations is done and which tools are used at ESA to plan the operations.

1 INTRODUCTION

Three missions to bodies of the solar system will be launched by the European Space Agency (ESA) next year: Rosetta to comet P/Wirtanen in January 2003, Smart-1 to the Moon in February/March 2003, and Mars Express to the planet Mars in June 2003. BepiCOLOMBO will go to Mercury; currently the launch is planned for 2011. Still, the preparation of the science operations already starts now, as the Announcement of Opportunity is planned to come out end of 2002. All of these missions have similar drivers for their operations:

- The instruments are provided by Principle Investigators (PI), delivered to ESA.
- The operations of instruments is the responsibility of PI.
- The instruments share spacecraft resources, like power, data rate, pointing directions.
- The round-trip light time is in the order of minutes to an hour.
- The trajectories and pointing of spacecraft need to be pre-planned.

To optimize the science output of such a mission, the payload operations must be coordinated and pre-planned, sometimes days or weeks in advance. This is the task of the Science Operations Centre (SOC). The methods and some tools used to do this is the topic of this paper.

2 MAIN TASKS OF THE SCIENCE OPERATIONS CENTRES

Science operations is the task of commanding the scientific payload of a spacecraft. Two main entities are involved in this task: The experimenter team, which has built the experiment, and a “Science Operations Centre (SOC)”, which coordinates the operations of the scientific payload. The SOC forwards the operational requests to the Mission Operations Centre, which is responsible to add the spacecraft operations and uplink the final command timeline to the spacecraft. The Mission Operations Centre also provides a Flight Dynamics team which prepares and commands the trajectory and attitude of the spacecraft.

The main tasks of the SOC are defined in the Science Management Plan for the respective mission and are summarized here:

(a) Definition of scientific operations for all mission phases with PI team support.
(b) Mission planning and implementation of experiment operation schedules.
(c) Co-ordination and pre-checking of command sequences generated by the PI teams for the operations of their payload before submission to the MOC.
(d) Together with the PI teams, the creation of a summary of the main scientific results, at regular intervals or for mission highlights.
(e) The preparation of guidelines for science data archiving and - supported by the PI teams - to create the mission data archive.

In the following, we concentrate on items (b) and (c), i.e. how is the scientific mission planned and implemented and which tools are used to co-ordinate and pre-check the operational command sequences.

3 LOGICAL SETUP OF THE CURRENT SCIENCE OPERATIONS

3.1 Rosetta

Fig. 1 shows the logical setup of the science operations for Rosetta. The different boxes depict the experimenter teams including the Lander, the Rosetta Science Operations Centre (RSOC), and the Rosetta Mission Operations Centre (RMOC). The RSOC
consists of staff of the Research and Scientific Support Department (RSSD) of ESA, the RMOC is located at the European Space Operations Centre (ESOC) in Darmstadt.

The first task is to prepare the operational requests, which is done by the experimenter team. Each experimenter team forwards their requests in the form of Operational Request Files (ORF) to the SOC. The ORFs coming from the Orbiter teams are called Orbiter Instrument Operational Requests (OIORs), those coming from the Lander are called Lander Operational Requests (LOR).

The SOC will perform resource checking, i.e. is the sum of the power the experiments needed within the power that the spacecraft has available, is the data volume that is generated lower than what can be stored or downlinked, etc. The SOC will also check whether experiment constraints are not violated. An example: One of Rosetta’s instruments is an atomic force microscope (MIDAS), which will collect dust and scan it with a resolution of a few nanometers. The instrument is sensitive to micro-vibrations. These micro-vibrations may be produced by the mechanical cooler of the infrared spectrometer (VIRTIS). This is just one example for a constraint that needs to be checked at the SOC, where the operations of the different scientific instruments come together.

If constraints are violated, the SOC will propose modifications to the timeline, in order to make the operations consistent with the constraints.

Control Centre (LCC), which adds the subsystem commanding. The consolidated Lander request file will be submitted to the RSOC, just like any other experimenter team’s Operational Request Files.

The consolidated Operational Request Files are called Payload Operations Request (POR) and will be forwarded from the SOC to the MOC. The MOC adds the spacecraft operational requests and will again perform conflict checking. This might result in another iteration back to the experimenter teams. After resolving all conflicts, the final operational timeline is uplinked to the spacecraft, where it will be executed.

The distribution of the downlinked telemetry plus other information (called auxiliary data) like restituted trajectory or pointing data will be performed by the MOC. It will put the information on a computer system called Data Distribution System (DDS), from where it can be retrieved by both the experiment teams and the SOC.

3.2 Smart-1

Other than Rosetta, Smart-1 is a technology mission with both Technology Investigators (TI) and science Principal Investigators (PI) contributing to the payload. However, the logical setup of the science and technology operations is the same as for Rosetta. What is called SOC for Rosetta is called Science and Technology Operations Coordination (STOC) for Smart-1. The STOC is located at RSSD, just as for Rosetta, the RMOC is located at ESOC.

Fig. 2 shows the logical information flow and it can be seen that except for the names of the different parties involved it is identical to Rosetta.

3.3 Mars Express

The setup for Mars Express is slightly different from the two previous ones. The Science Operations Centre is called Payload Operations Support (POS) located at Rutherford Appleton Laboratory (RAL) in England,
whereas the overall management and the long-term planning is done within the Research and Scientific Support Department (RSSD) of ESA.

Just as on Rosetta, there is a Lander Operations Centre (LOC) located also in England (at Leicester University). The current setup is such that the LOC is interfacing directly with the MOC. To ensure coordination between the Lander and the Orbiter science, a working group was set up that is coordinated from RSSD at ESA.

![Diagram of Logical setup of the Mars Express science operations.](image)

**Fig. 3. Logical setup of the Mars Express science operations.**

## 4 OPERATIONAL REQUEST FILES

The format, which the Mission Operations Centre is requiring for the operational requests, is called the POR format (POR = Payload Operational Request). It is not very easy to read for humans, therefore it is not very useful for interactive planning. We therefore introduced a different syntax for providing the Operational Request files, called ITL syntax (ITL = Input Timeline). An example is given in Fig. 4.

![Example timeline. The "#" is a comment character. VIRTIS is an instrument on Rosetta. All telecommand sequences start with "AVRS...". "VVRG..." are parameters of the sequence.](image)

**Fig. 4: Example timeline.**

The basic definition of an ITL consists of a time or an event in the first column, possibly a delta time or a time window in the second column, the name of the experiment, the mode of the experiment, and finally the telecommand sequence with its parameters. For the higher level planning, timelines can be written using modes only, *e.g.* to get a first idea on the power consumption or the generated data volume.

The timeline can also contain pointing requests, that give an indication of where the scientists would want the spacecraft to point to, *e.g.* nadir or pointing to an object.

## 5 PLANNING TOOLS USED AT ESA

### 5.1 Introduction

The science operations for Mars Express is done by the Payload Operations Service (POS), located at RAL in England. This institution is doing the science operations for the CLUSTER mission currently in orbit. They have written a tool called MIRA (Mars Express Instrument Resource Analyser). It will process and plan the science requests and lander contact requests. At ESA/ESTEC, we have developed two tools for performing this task, which are used both on Rosetta and on Smart-1. These will be presented in more detail in the following sections. Also see [1] for details.

### 5.2 The Experiment Planning System

A graphical overview over the Experiment Planning System (EPS) is given in Fig. 5. The EPS is a mission-independent software tool that is used for checking constraints and flagging of the experiment operations. The EPS uses as input the Operational Request Files from the experimenter teams. Optionally, pointing requests can also be handled. All the information needed for checking the syntax and constraints is stored in so-called Experiment Description Files (EDF). There is one EDF per experiment. The EDFs are ASCII files, so to update or add an experiment, there are no changes to the source code of EPS needed. This ensures that the EPS is mission independent.

The EDFs contain the following information:

- All defined telecommands for an experiment.
- All defined telecommand sequences.
- Constraints.
- Possibly a model of the experiment using so-called modules.

Assume we want to model a simple experiment containing a Data Processing Unit (DPU) and a sensor. We could define two modules: DPU and SENS0R. Each one of the modules can have two states, ON and OFF. The ON states will be assigned power values, *e.g.* 5 W for the DPU, 5 W for the sensor. In the definition of the telecommands we would have
commands DPU_ON and DPU_OFF, which switch the DPU on or off. S_ON and S_OFF would switch the sensor on or off. If the EPS encounters these telecommands in the input timeline, it will change the states of the modules accordingly. It will automatically calculate the resources according to the states, i.e., if both modules are on the power will be 10 W, if one is one the power will be 5 W. A constraint could be defined that the sensor is only switched on if the DPU already is on, etc. With this method we can build a model of the experiment to any level of detail. Obviously, the SOC will try to stay at a very top level for this model to keep the planning simple. An experimenter team, on the other hand, may produce a very detailed model of their experiment to get a detailed view of the resource profile during operations.

The Operational Request Files follow the syntax defined in a previous section (ITL syntax), or the POR format as requested by ESOC. EPS can convert ITL syntax to POR syntax and vice versa.

Another input that the EPS can handle is so-called Event files (EVT). These will come during spacecraft operations from the MOC. For planning purposes, they can also be generated with the PTB at the SOC as explained in the following section. The Event files give a correlation between an event (e.g., the spacecraft is more than 10 deg above the horizon of the Perth ground station, acronym is PER_AOS_10 (Perth, Acquisition of Signal, 10 deg)) and time.

If the Operational request contains a telecommand sequence scheduled at a certain event rather than a time, the EPS will look in the Event file and replace the name of the event with the time it occurs. This method allows a very flexible planning. For example, we plan the operations around the asteroid flybys of Rosetta relative to the point of closest flyby without yet exactly knowing what the time of this closest flyby is.

If EDFs are available and the EPS is run with an Input Timeline, it will generate a number of output files. Possible outputs are:

- Conflict file
- Power requirements versus time
- Data rate/volume versus time
- State of the modules versus time

The conflict file will list the time and name of a violated constraint as defined in the EDF. This is the most valuable planning information. The goal of the planning should always be to avoid conflicts. Of course there may be conflicts, which are more severe than others. On Rosetta, e.g., a very hard constraint is to not look closer to the sun than 11 deg, as this might damage one of the detectors on board. This would generate a FATAL conflict. On the other hand, the optical cameras may be able to operate at angles closer than 45 deg, but only with reduced performance due to straylight problems. This may cause a WARNING only.

If the power values and data rates are defined in the EDFs, the EPS can calculate the actually used resources as a function of time and will write these in a file. It will also produce a total value, which can be compared to the available resources from the spacecraft.

It is also possible to model local experiment memories, a spacecraft mass memory, and the downlink to ground. The EPS can therefore also be used to find out whether the data generated by an experiment (e.g., a camera) would be more than can be held in memory or downlinked. The memory contents will be displayed in the output files also.

The EPS is a command-line tool written in ANSI-C. It is working in DOS, UNIX, LINUX, and IRIX. The use of ASCII files to define the experiments makes it completely mission-independent. A graphical user interface is under development.

5.3 The Project Test Bed

The Project Test Bed (PTB) is used to model the environment of the spacecraft. For each mission, there are different setups with different central bodies, e.g., for Rosetta there is a PTB for the escape phase at the beginning of the mission with the Earth at the centre, we have the asteroid flybys, the planetary flybys, and a setup for the comet phase with the comet at the centre.
The PTB reads in the trajectory of the spacecraft as provided by the Flight Dynamics team of the MOC. Alternatively, an internal (but simple) orbit propagator can be used. In each setup, the PTB will simulate, as a function of time, the precise geometry of the spacecraft with respect to the central body, the planets, and the sun.

It is based on a real-time simulator tool called EuroSim written by Fokker B.V. This tool has been used at the Agency before for the simulation of the International Space Station and other missions like Proba. Our system for planetary missions is based on these available systems. We have working PTBs for Smart-1 and Rosetta, which are used in the science planning. For Mars Express, a first version is available, but it was not kept up with the other simulations as the planning tasks was given over to the POS. We are in the process of detailing the available BepiCOLOMBO simulation to a state where it can be used e.g. to size the spacecraft mass memory and to get some first ideas on mapping scenarios.

The PTB will also read so-called Pointing Request files (PTR). These files allow scripting the pointing of the spacecraft. The PTB contains a simple dynamical model of the spacecraft and will perform slews as specified in the PTR. This allows the SOC to get an estimate of slewing times, possible conflicts by illumination constraints, and more.

The PTB can also be set up to produce the Event files that in normal operations are expected from the Flight Dynamics team. This allows for example to study asteroid flybys with different flyby distances and generate events for different distances to the asteroid. These events can then trigger experiment operations.

6 PLANNING THE MISSION

The planning of the mission is done in several steps, starting with a top-level outline and detailing it until individual telecommand sequences are reached. To structure the planning process, it is divided in three different cycles: Long-term planning, medium-term planning, and short-term planning.

6.1 Long-term planning

The goal of the long-term planning is to get a top-level structure on what to do. The input for that is documented by the experimenter teams already in the proposals, later in the Experiment Interface Documents, which contains a section on “intended scientific measurements”. Top-level discussions at the Science Operations Working Group meetings and informal exchange lead to additional input.

The Science Operations Centre or the Project Scientist’s team derives such a top-level plan. In Rosetta, the result is called Science Activity Plan. It lists the sequence of planned Mission Scenarios, which are shorter-term activities like “close encounter”, “landing site selection”, “coma scan”. Each scenario allows identifying priority experiments, e.g. for the landing site selection the camera systems would need to be operating.

For Mars Express, the mission is divided in so-called science sub-phases, which are a combination of possible downlink rate, illumination conditions, etc. Again, a certain sub-phase would be best suited for certain experiments, e.g. sub-phases with an illumination angle between 30 and 60 deg would be the best for imaging, sub-phases which are at night would preferably be used by the radar.

6.2 Medium-term planning

The meaning of medium-term planning differs slightly between the missions. For Smart-1 and Mars Express, this planning cycle involves generating pointing requests and giving the Flight Dynamics team at ESOC the chance to iterate it. The payload operations will be prepared on a preliminary basis. The current plan is to perform this planning during the orbiting of Mars or Moon, parallel to the short-term planning.

In Rosetta, this planning cycle simply means that the two years before the comet is reached, the complete team starts to plan “artificial” operations, both to train the teams and to find ways of solving problems not related to the target object.

6.3 Short-term planning

The short-term planning is where the final operational requests that will be uplinked to the spacecraft are generated. Typically, one week of operations will be
planned. The time to plan one week will be longer than that, so that several short-term planning cycles will have to be run in parallel.

This is depicted in Fig. 7. Only the first planning cycle is shown expanded. The complete iteration is visible. The first step is that RSOC provides requests for the trajectory sufficiently advanced so that Flight Dynamics can prepare the trajectory. They will provide information on the trajectory back to the RSOC. Together with the detailed requests from the experimenter teams, the RSOC prepares a consolidated operations plan. In iteration with the PI teams the RSOC processes the operational files (called OIOR for Rosetta, i.e. Orbiter Instrument Operations Requests). The Pointing Requests will be given to Flight Dynamics, they produce the final pointing and return it to the RSOC. Now the RSOC can finalize the consolidated operational requests and generate the final file, called Payload Operational Request file (POR). The Flight Control Team at the RMOC adds the spacecraft operations and uplinks the commands.

7 CONCLUSION

In the previous paper, we have outlined the concept for science operations for planetary missions, concentrating on Rosetta as an example. We also described briefly the software tools available at ESA to support this planning. Clearly, all the lessons learnt from these missions can be used in preparing the science operations for upcoming planetary missions like BepiCOLOMBO. Solar Orbiter will also have similar requirements on science operations. Using the same concept and the same software tools will increase the efficiency of the Agency and reduce the overall costs of these missions.

8 REFERENCES

Technology and Science from Earth to Moon: SMART-1 Experiments and their operations

A.E. Marini(1), R. Lumb(1), G. D. Racca(1), B. H. Foing(2), M. Dias-Almeida(2)

(1) SMART-1 Project, European Space Agency ESA/ESTEC - Keplerlaan 1, 2200 AG Noordwijk – The Netherlands, E-mail 1st author: andrea.marini@esa.int
(2) Research and Scientific Support Department, European Space Agency ESA/ESTEC - Keplerlaan 1, 2200 AG Noordwijk – The Netherlands.

ABSTRACT
SMART-1, the first European mission to the Moon aimed at demonstrating the Solar Electric propulsion hosts 10 Technology and Science experiments. The monitoring of the spacecraft plasma environment and the thruster contamination produced by thruster is carried out by SPEDE (Spacecraft Potential, Electron and Dust Experiment) and EPDP (Electric Propulsion Diagnostic Package). The miniaturised remote sensing instruments on-board SMART-1 are: AMIE (Advanced Moon micro-Imager Experiment), D-CIXS (Demonstration of a Compact Imaging X-ray Spectrometer), supported in its operation by XSM (X-ray Solar Monitor), and SIR (SMART-1 Infrared Spectrometer) Technology experiments for deep-space communications and navigation are: KATE (Ka-Band TT&C Experiment), based on X/Ka-band transponder which also supports RSIS (Radio-Science Investigations for SMART-1), Laser-link, demonstrating a deep-space laser communication link and OBAN (On-Board Autonomous Navigation experiment).
The Experiments will be performed during two distinct phases of the SMART-1 mission, including 17-month Earth escape phase and a nominal 6-month operational phase in elliptical Moon orbit
The SMART-1 STOC (Science and Technology Operations Co-ordination) carries out the planning and co-ordination of the Technology and science experiments.

1. INTRODUCTION AND BACKGROUND
SMART-1 is the first of the Small Missions for Advanced Research in Technology of the ESA Horizons 2000 Science Plan. These missions have been introduced by ESA as one of the strategic elements into the Horizon 2000 Science Plan. The scientific importance of the SMART-1 mission resides mainly in its preparatory nature for upcoming truly scientific missions and in particular for those missions which will benefit of Solar Electric Primary Propulsion (SEPP) and deep space communications. Among these the Mercury Cornerstone study had clearly identified the SEPP as a key element to enable a low circular orbit. SMART-1 shall demonstrate the use of SEPP on a small mission, but representative of a future deep-space science mission. Therefore the emphasis is placed on the common system aspects and on the peculiar flight dynamics and control techniques needed for implementing the mission profile, rather than on the choice of a particular engine, which is more mission-specific.
The requirement for science output of SMART-1 is secondary to the technology demonstration objective. Nevertheless, in order to demonstrate practical utilisation of SEPP, the spacecraft should travel beyond an Earth orbit and reach some relevant solar system object. The Moon was chosen as a target for its scientific importance and relative ease of access from a commercial Geo-stationary Transfer Orbit (GTO). There is consequently an initial system allocation of 15 kg and 50 W for accommodation of science instruments.
SMART-1 is also a “first-ever” low-budget small mission for Science at ESA and in this sense it explores and tests new ways of implementing cost-effective procurement and efficient management. The budget constraints of SMART-1 also require a cost-effective approach to spacecraft development and verification, not only considering AIT cost, but also associated or induced effort during the design and analysis stages of the development programme.
The very peculiar mission profile – providing a spiral escape from Earth gravitational field in about 17 months and followed by weak capture of the Moon orbit and 6 months of lunar orbit operations – is described in [1].
At the time of writing (August 2002), SMART-1 spacecraft is undergoing the final AIV/AIT phase (see Fig.1) with System and environmental testing. The launch is foreseen after 1st March 2003.


© European Space Agency • Provided by the NASA Astrophysics Data System
2. SMART-1 SCIENCE AND TECHNOLOGY EXPERIMENTS

2.1 Scientific Objectives

Although the main objective of the mission is essentially technological, a great deal of effort was spent to improve its scientific return. Scientific observations can indeed be carried out during both the lunar operational phase and the cruise phase.

The lunar observation phase will be performed from a polar orbit with the peri-lune on the South hemisphere at about 30° from the South Pole and at an altitude varying between 1,900 km and 300 km. The baseline apo-lune will have an altitude varying between 8,400 and 10,000 km.

SMART-1 Lunar scientific studies will concentrate on mineralogical mapping and elemental geochemistry and will include:

- Elemental geochemistry (X-ray imaging spectrometer, with a spatial resolution of 30 km at peri-lune)
- Mineralogy (Near-IR spectrometer combined with camera map-ping)
- Geology, morphology (High resolution camera)
- Exospheric environment (Camera, plasma and dust experiment)

During the long cruise phase the following scientific investigations will be performed:

- Monitoring of X-ray variability of several cosmic sources and the Sun (X-ray spectrometer)
- Cometary detection and auroral X-ray monitoring on both hemisphere of the Earth (X-ray spectrometer)
- Monitoring optical micro-variability of stars
- Space-time variations of the plasma and electron environment in the Earth-Moon space

Finally, as mentioned, the ultimate scientific return of the SMART-1 mission resides in its objective to qualify the use of novel technologies for more ambitious future planetary missions.

2.2 Technology Objectives

As stated previously, the main design drive of the SMART-1 mission is to test in a deep-space representative mission the primary electric propulsion. The mission will qualify the system and its use as primary propulsion. The system aspects such as electrical power supply as well as thrust direction control and mechanical and thermal accommodation are main design drivers. In addition the characterisation of the electromagnetic, plasma and dust environment created by the functioning of the EP is addressed by two instruments: EPDP and SPEDE (see description later).

Other technologically advanced items are addressed by the SMART-1 payload. A new Deep Space X-Ka band transponder will be own as a technology payload. This transponder, essential to BepiColombo will allow also to perform a radio science investigation to monitor the dynamical performances of the electric propulsion system and to measure the rotational state of the Moon, as explained later in this paper. It also aims at assessing capabilities of an advanced X/Ka link for precise Doppler and ranging measurements in preparing future high-precision geodesy and relativity experiments.

Furthermore the possibility of employing laser communication for future Deep Space links will be investigated. The on-board camera (AMIE) will acquire and image the laser beam transmitted by the ESA Optical Ground Station (OGS) in Tenerife (Spain) see fig.2)

![Fig.2: The ESA OGS at the Observatorio del Teide, Tenerife](image)

Two of the science instruments have also been selected due to their technological advances. The D-CIXS X-ray spectrometer has novel features, such as the micro-structure collimator and the Swept Charge Detector. The SIR near-IR spectrometer is of high relevance for planetary research as it is a very compact, miniaturised version derived from a quasi-
monolithic commercial quartz grating spectrometer.

2.3 Science and Technology Payload
description

The payload is composed of technology and scientific experiments and its total mass is about 19 kg. The seven SMART-1 instruments support ten investigations.

EPDP Electric Propulsion Diagnostic Package (2.3 kg, 18 W). A suite of sensors for thruster diagnostics with ion energy up to 400 eV and spacecraft contamination monitoring. The PPA (Plasma Probe Assembly) is composed of a Retarding Potential Analyser and of a Langmuir Probe), see fig. 3

![Fig.3: Flight unit of the EPDP's PPA](image)

The other sensors are meant to detect contamination of neutral Xe ions on the spacecraft surface and are a Quartz Crystal Microbalance (QCM, fig.4) and a Solar cell.

![Fig.4: EPDD's QCM flight unit](image)

SPEDE Spacecraft Potential, Electron and Dust Experiment (0.8 kg 1.8 W). Langmuir probes are made of tiny TiN sensor foils attached on two short (60 cm) booms (see fig.5) for measuring energy range of a few tens of eV, with plasma density from 1/10 to 1000 particles/cm²

![Fig.5: SPEDE booms flight units](image)

KaTE X/Ka-band Telemetry and Telecommand (TT&C) Experiment (6.2 kg, 28 W). A X-up/X-down and Ka-down Deep-Space Transponder running turbo-codes, allowing up to 500 Kbps data rate from lunar orbit.

![Fig.6: KaTE Deep space transponder and the X - and Ka-band horn antennas](image)

D-CiXS/XSM (see Fig. 7) Demonstration of a Compact Imaging X-ray Spectrometer (5.2 kg, 20 W, including XSM). A 12° x 32° FOV spectral imager in 0.5-10 keV range based on Swept Charge Device detectors and micro-collimators, measuring x-ray fluorescence from the lunar surface discriminating the solar background by means of the X-ray solar monitor (XSM, see afterwards)
AMIE Asteroid-Moon Imaging Experiment (2.2 kg, 9 W). A 5.3° FOV miniaturised camera with a 4-band fixed filter (0.75, 0.9 and 0.95 μm wide-band mineralogical filters and a 0.847 μm narrow-band filter for Laser-link). The camera is based on high-density 3-D cube-packed Multi-Chip Module electronics (see fig.8)

SIR SMART-1 Infrared Spectrometer (2.3 kg, 4.2 W). A 1 mrad FOV point-spectrometer with 256 channels operating in the 0.9-2.4 μm wavelength range (NIR) for lunar mineralogy (see fig.9)

LASER-LINK demonstration of a deep-space optical link acquisition (with AMIE), where a laser beam is sent in direction of SMART-1 S/C by the ESA Optical Ground Station situated in Tenerife (see Fig.10). The aim of the experiment is to prepare for deep-space laser communication link, by demonstrating acquisition of the laser-link up to lunar distance and to validate a novel beam arrangement in four sub-apertures for mitigating the effect of atmospheric turbulence on the laser beam. The AMIE camera will be used on-board SMART-1 to image the beam profile and detect the link acquisition.

OBAN (On Board Autonomous Navigation) concept verification (with AMIE), by means of discriminating the motion of a non-stellar target (planet, asteroid) against the starry background in a long exposure image. The camera information will be completed by the on-board star—tracker data and elaborated off-line by the navigation software which generates the navigation data (on-ground simulation)

RSIS (Radio-Science investigations for SMART-1) uses KaTE to perform the characterisation of the X- and Ka-band communication channels and performance, Electric propulsion monitoring and the demonstration of a novel method for measuring the libration properties of a celestial body (the Moon for SMART-1, in preparation of BepiColombo) from the lunar orbit (with KATE and AMIE). The Electric propulsion monitoring is performed by tracking the Ka-band signal while the thruster is on, allowing measurements of the thrust force with a resolution in the order of 10^4 N and of the acceleration of the S/C with a resolution of about 0.1 mm/s^2.

XSM (Fig.11) will monitor the solar X-ray emission in the 1-20 keV range for studying the solar corona activity, both in short and long
time scales. Observing “the Sun as a star” will contribute to test stellar X-ray emission models and study the solar-stellar connections.

For further reading, reference paper [2] describes the science potential and goals of the scientific instruments, while [3] shows how the technology experiments prepare for future ESA Cornerstone missions.

3. SMART-1 EXPERIMENTS OPERATIONS
SMART-1 Experiments are carried out during the whole mission, after spacecraft commissioning.

3.1 Earth Escape Phase
Most of technology experiments are carried out during the cruise phase, together with calibration of science instruments.

In the early cruise phase just after launch, EPDP and SPEDE are commissioned, to be ready monitoring the Electric propulsion Commissioning and first Operations. All the other instruments are then commissioned in the following weeks when passes and orbital environment allow it.

During the first 75 days after Thruster switch-on, the electric propulsion is operating almost continuously, to rapidly raise the perigee to a safe height above the Van Allen radiation belts. During this period only SPEDE and EPDP will be operational, while XSM will start solar monitoring only close to the apogee.

After this phase, the thrusting periods are optimised in a sequence of thrusting and coasting arcs, aimed at raising the orbit apogee and at consuming the least possible fuel.

During thrusting SPEDE is operating almost continuously, supported by EPDP for the electric propulsion monitor or at the same time XSM takes solar spectra.

During the coast arcs, the remote sensing instruments – AMIE, SIR and D-CIXS – will start performing their in-flight calibration and – in visibility of the Ground stations – either KaTE or Laser-link will be operated.

During longer coast arcs (towards the end of the Escape Phase, where also the low-energy proton density is much lower) D-CIXS will perform X-ray observations of celestial objects, the Earth aurora and opportunity targets such as comets. AMIE will take images of the Earth and the Moon (for Public outreach) and target planets for carrying out the OBAN experiments.

During the longest thrust arcs at the end of the escape phase it will be possible to have KaTE pointing at a Ground Station for a long time (~ 1-2 hours) while the thruster operates: in this situation it will be possible to perform the RSIS accurate measurements of the Electric Propulsion performance by tracking in Ka-band.

3.2 Lunar Phase
After the very delicate phase of the weak capture (see [1]), SMART-1 enters a Moon-centred orbit profile and thruster is switched-on to reduce the apo-lune to the nominal observation orbit of 10000 km with a 300 km perilune and an argument of the peri-centre of 270°. During this phase only thruster monitoring will be performed.

The Moon observation phase will mostly run science experiments and KaTE will be also for supporting science data telemetry throughout.

The core science observations for the lunar phase are the following:
- Global X-ray spectral imaging survey (D-CIXS/XSM).
- Near-Infrared spectroscopy and broadband spectral imaging (SIR+AMIE with colour filters).
- High-resolution imaging (AMIE).
- Exospheric & lunar plasma environment monitoring (SPEDE).
- RSIS: Demonstration of measurement method of (lunar) libration from Orbit (KaTE/AMIE).

The orbit will evolve during the 6 months Moon phase with varying height and argument of the perilune [1], offering a variety of illumination and viewing conditions. The spacecraft will be mostly kept in a nadir pointed attitude, with some yaw manoeuvre to be performed for avoiding the spacecraft radiators to be excessively illuminated by the sun. Nevertheless off-pointing will be performed to aim at imaging and spectral target required by AMIE and SIR.

4. THE SMART-1 STOC
The SMART-1 Science and Technology Operation Co-ordination (STOC) is a service within the SMART-1 Project fully integrated within the Operations mission element. The purpose of the STOC is to co-ordinate the Science and Technology Operations and to
support the collection and the exploitation of the mission data. SMART-1, being a low-cost mission, the STOC does not assume the characteristics of an usual Science Operation Centre of a scientific satellite mission, but - making use of available modern information technologies - it rather provides some vital tasks in a flexible and economic way, by being remotely accessed and interacting with the Mission Operations Centre and with the Experiment Operations Facilities. Due to the nature of SMART-1 Mission, science and technology data are collected together and throughout the whole mission, with a prevalence of Science data during the Moon orbiting Phase. The STOC will therefore act as a Science Operations Centre during this part of the mission, without however changing the organisation of the activities or the personnel. The responsibilities of the STOC are in the terms of reference of co-ordination in the preparation and the conduct of SMART-1 Experiment Operations, for the whole duration of the mission till the spacecraft decommissioning. In particular the STOC:

- Supports the Science and Technology Working team in the activities of Science Operation Planning
- Co-ordinates the preparation and conduct of the Experiment operations towards the PI's/TI's
- Takes an active part in the experiment timelining process by generating the Experiment Science Master Plan from the inputs of all Experiments and provides it to ESOC (with no requirement of timeline optimisation).
- Monitors the Data Delivery to the Experiment teams.
- Maintains an archive of Raw data throughout the duration of the mission and supports the post-processing and Science archiving activities at ESA.
- Analyses the results and prepares the "lessons learnt" for use on later projects

The STOC facility is located at ESTEC within the within the joint Science Operations for Planetary Missions (SOPM) facility of the ESA Research and Scientific Support Department (RSSD), which has been established to support all upcoming planetary missions, including Rosetta, Mars Express and BepiColombo. The potential synergy developing the mission operations tools, the re-use of the facilities and the expertise and the advantage of operating in a "specialised" environment, makes this solution the most attractive for the SMART-1 STOC, which is by definition meant to be low-cost and flexible.

A Workspace with the communication means towards the Mission Operation Centre (MOC) and the Experiment facilities and dedicated workstations linked to the generic planning, simulation and archiving facilities of the SOPM completes the SMART-1 STOC infrastructure. The development of the software tools to be used for mission planning and scheduling has been also taken benefit from the synergy with other mission. SMART-1 uses EPS (Experiment planning system) linked to PTB (Project test-bed) as core tools, as thoroughly described in paper [4] in these remaining proceedings.

5. REFERENCES
LUNAR DATA SIMULATION FOR SMART-1

Eduard Vilari(1), Miguel Almeida(2), Detlef Koschny(3), Bernard H. Foing(4)

(1) ESTEC/SCI-SR, P.O. Box 299, 2200 AG Noordwijk (Netherlands), Email: evilar@rssd.esa.int
(2) ESTEC/SCI-SR, P.O. Box 299, 2200 AG Noordwijk (Netherlands), Email: malmeida@rssd.esa.int
(3) ESTEC/SCI-SB, P.O. Box 299, 2200 AG Noordwijk (Netherlands), Email: detlef.koschny@esa.int
(4) ESTEC/SCI-SR, P.O. Box 299, 2200 AG Noordwijk (Netherlands), Email: bernard.foing@esa.int

ABSTRACT

The value of the data collected on a planetary mission does not depend only on the data by itself or its amount, but also on its joint analysis, comparison and integration with previous and/or expected data. The SMART-1 mission has also been prepared to get an opportunity for new science. We have done a search on the past missions that studied our satellite, the Moon. Therefore, previous lunar missions like Luna, Apollo, Hiten, Clementine or Lunar Prospector offer the possibility of comparing the existing lunar databases. This also helps in selecting for SMART-1 a wide range of lunar targets between the hundreds of Mares, Sinus, Catenas and Craters.

On the other hand, simulation is necessary in order to reach this aim, and a simulation program has been developed to plan the SMART-1 payload operations and to pre-evaluate SMART-1 representative data and science return.

Finally, we discuss perspectives how to enhance SMART-1 return in order to get the best results building on the knowledge of the Past with the present ESA technology.

The approved SMART-1 scenario is a lunar mission, including 6 months operations in lunar orbit. The mission is to be launched as Ariane 5 piggyback into GTO (Geostationary Transfer Orbit). The SEPP is used to spiral out from GTO (during some 14 months), then achieve lunar swingby, lunar capture, spiralling in to a near polar lunar orbit with apolune of 10.000 km and perilune of 300-1.000 km.

1.1 Scientific Instruments

- AMIE: Asteroid Moon micro-Imager Experiment (also supporting investigation on On Board Autonomous Navigation (OBAN) and Laser-Link Experiment)
- SIR: SMART-1 Infra-Red Spectrometer
- D-CIXS: Demonstration of a Compact Imaging X-ray Spectrometer
- EPDP/SPEDP: Electric Propulsion Diagnostic Package / Spacecraft Potential Electron and Dust Experiment
- KaTE: X/Ka-band Telemetry and Telecommand Experiment

2. PAST LUNAR MISSIONS

- Apollo - NASA Manned Lunar Program (1963 - 1972)
- Hiten - ISAS Flyby, Circumnavigation and Impact Mission to the Moon (1990)
3. PRIMARY LUNAR DATABASES

Every one of the Past Lunar Missions got useful data, but nowadays we can remark the quality and width of three of them (in order of data volume importance):

3.1 Clementine’s

It took over 1.8 million of pictures of the lunar surface with its different cameras and respective wavelengths with bandwidths:

- Ultraviolet / Visible Camera (UV/Vis) : 415 nm (40 nm), 750 nm (10 nm), 900 nm (30 nm), 950 nm (30 nm), 1000 nm (30 nm)
- High Resolution Camera (HIRS) 415 nm (40 nm), 560 nm (10 nm), 650 nm (10 nm), 750 nm (20 nm)
- Near Infrared Camera (NIR) 1100 nm (60 nm), 1250 nm (60 nm), 1500 nm (60 nm), 2000 nm (60 nm), 2600 nm (60 nm), and 2780 nm (120 nm).
- Long-Wave Infrared Camera (LWIR) 8—9.5 μm

http://nssdc.gsfc.nasa.gov/planetary/clementine.html

![Fig.1. Copernicus. Mosaic of UV/VIS cameras.](image)

3.2 Lunar Orbiter’s

Its two cameras provided quite useful data and pictures, more than 30 years ago:

- High Resolution Camera (HR) and Medium Resolution Camera (MR)

http://nssdc.gsfc.nasa.gov/planetary/lunar/lunarorb.html

![Fig. 2. Simulation by PTB of SMART-1 flying over Hommel’s Crater](image)

3.3 Lunar Prospector’s

Some information can be found at:

http://nssdc.gsfc.nasa.gov/planetary/lunar/lunarporsch.html

4. PTB AND EPS

SMART-1 mission has taken into account these two powerful tools developed by EuroSim and adapted to and by ESA.

4.1 Project Test Bed

Project Test Bed is a simulator whose main task is to support a Science Operations Centre in the planning of the science operations. It allows, for instance:

- Simulating the spacecraft in its space environment
- Recording flight dynamics and science events in a timeline
- Calculating and recording data needed for science planning (illumination conditions, etc.)
4.2 **Experiment Planning System**

Experiment Planning System is a software tools developed for the creation of the experiments operations for a specific time period. It does this by allowing to enter an operation timeline, which is a list of experiment operations and a specific time for when the operation has to be executed, and check for constraints violations on each of the experiments.

![Table](image)

# (c) ESA/Estec

<table>
<thead>
<tr>
<th>#Ref_date</th>
<th>1-July-2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start_time</td>
<td>000_00:00:00</td>
</tr>
<tr>
<td>End_time</td>
<td>000_00:01:00</td>
</tr>
<tr>
<td>Version</td>
<td>1</td>
</tr>
<tr>
<td>Init_mode</td>
<td>AMIE OFF</td>
</tr>
<tr>
<td>Init_MS: AMIE CAMERA</td>
<td>OFF</td>
</tr>
<tr>
<td>Init_MS: AMIE ICU</td>
<td>OFF</td>
</tr>
<tr>
<td>000_00:00:00</td>
<td>AMIE OFF OFF_TO_ON</td>
</tr>
<tr>
<td>000_00:00:01</td>
<td>AMIE ON</td>
</tr>
<tr>
<td>CAMERA ON</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Input Timeline of AMIE experiments (fragment)

5. **AMIE CAMERA**

AMIE is a micro-camera experiment that will contribute to the characterization of surface mineralogy and geology, in combination with other SMART-1 experiments providing the regional context for major geologic features. It is of special interest for imaging the polar permanent shadows and illumination conditions in areas of quasi-eternal light.

The original AMIE concept foresaw a panchromatic, 1024 x 1024 pixel image with a medium field of view of 5.3 degrees by 5.3 degrees. The camera will provide a high spatial resolution, some 50 m/pixel. The scheme of the FOV is on Figure 4.

![Figure 4. AMIE filters](image)

Initially a technology demonstration of a miniaturized micro-imager, AMIE on SMART-1 now has significant science objectives, covering the study of the Moon's morphology, topography and surface texture. The science community realized the potential of such a camera operating in a multi-spectral mode. AMIE has thus been equipped with three filters, in the red (750 nm) and infrared (915 & 950 nm). It will also operate in white light. A test image was obtained on laboratory minerals (see Figure 5).

![Figure 5. Basalt and olivine. Photographed by AMIE camera - all filters -](image)
AMIE will be opening new ground in the field of multi-spectral lunar observation. Whereas the multi-spectral camera aboard the American Clementine mission had constant illumination conditions, SMART-1’s orbit will offer multi-angular imaging. AMIE’s views at different angles correlated with Clementine data of the same lunar areas will allow scientists to establish photometric models, allowing the interpretation of such spectral data.

6. CONCLUSION

Halfway through the year 2004 SMART-1 will reach the Moon. That moment will mean the success of the Solar Electrical Propulsion (SEP) and the beginning of the first geological and mineral studies over the surface done from a European spacecraft.

Hundreds of simulations and studies have been done to prepare the incoming experiments; then it will be time to check if the assumptions taken were correct, and the goals affordable. Past missions - Clementine, Lunar Orbiter - have provided the scientist community a huge database of images over the electromagnetic range. AMIE, SIR and D-CIXS experiments of SMART-1 mission will be able to take records and compare them with the existing ones, as well as enlarging the lunar knowledge on these and other fields.

7. REFERENCES


- RACCA, G., FOING B.H., SMART-1 : The First Time of Europe to the Moon, EarthMoon & Planets 85, 379, 2001

- VILAR, E., Planning and optimisation of payload operations for the SMART-1 & Rosetta missions, RSSD, ESTEC, Stage report, Aug 02.
MEASURING THE ELECTRIC PROPERTIES OF PLANETARY ENVIRONMENTS WITH MUTUAL IMPEDANCE (MI) PROBES

R. Trautner (1), R. Grard (1)

(1) Research and Scientific Support Department, ESA/ESTEC, Noordwijk, The Netherlands
Email: Roland.Trautner@esa.int Tel. +31 71 565 3955 Fax: +31 71 565 4697

ABSTRACT
Mutual Impedance Probes measure the complex permittivity of materials by means of a quadrupolar array of electrodes and associated electronics for generating, recording and processing waveforms [1]. MI instruments have been developed for a number of ongoing space missions. The HASI/PWA MI probe will determine the electric properties of the atmosphere of Titan, Saturn's largest moon, during the descent of the Huygens probe [2]. After landing, the instrument will provide data on the properties of Titan's surface materials. The permittivity probe PP, as part of the SESAME instrument package for the Rosetta Lander, will determine the electrical properties of comet Wirtanen's surface. The main features of MI probes are first recapitulated. Instrument architectures for atmospheric, surface and subsurface investigations are described. Results from recent field test campaigns in harsh environments are presented. A new MI probe prototype employing a linear electrode array for application on mobile platforms or on penetrator devices is described. New application areas for future MI probes and relevant technology requirements are discussed.

frequency $\omega$, is injected between two transmitting electrodes, TX1 and TX2, and induces a voltage $A$ between two receiving electrodes, RX1 and RX2. The frequency is chosen such that the wavelength is much larger than the size of the electrode array. The complex ratio $A/I$ is the mutual impedance of the circuit.

If the amplitude and phase of the measured voltage are $A_0, \varphi_0$ in a vacuum, and $A, \varphi$ in a given environment,

the apparent conductivity $\sigma$ and relative permittivity $\varepsilon_r$ of the medium can be calculated using (1) and (2).

$$\sigma = \frac{A_0 \sin(\varphi - \varphi_0)}{A} \quad (1)$$

$$\varepsilon_r = \frac{A_0 \cos(\varphi - \varphi_0)}{A} \quad (2)$$

where $\varepsilon_0$ is the relative permittivity of vacuum.

The mutual impedance reflects the bulk properties of the medium. Analytical solutions exist also for configurations where the electrode array is placed on the surface of a liquid or on a locally planar ground, and for intermediate configurations.

2. INSTRUMENT ARCHITECTURES FOR ATMOSPHERIC, SURFACE AND SUBSURFACE INVESTIGATIONS

Any Mutual Impedance probe consists of two basic building blocks: The analogue and digital electronics, which are usually accommodated inside a spacecraft or vehicle body, and the transmitter and receiver electrode array, which is arranged such that the material of interest is exposed to the electric field generated by the transmitter electrodes. In the case of an atmospheric probe such as Huygens, the electrodes are situated on deployable booms. For surface operations, the electrode array may be integrated into a landing gear (like on the Rosetta Lander), ejected from a stationary

Figure 1: MI probe principle

1. MI PRINCIPLE

A mutual impedance probe consists essentially of a sensor array, a current generator and a voltmeter as shown in Fig.1. An alternating (sinusoidal) current $I$, of

---


© European Space Agency • Provided by the NASA Astrophysics Data System
Lander (as planned for the Netlanders) or trailed by a mobile platform. For subsurface MI probes based on tools such as drills or moles, body or tether mounted electrodes may be used. For all configurations, the major part of the electronics is accommodated at some distance, often as part of a more complex instrument to save mass and power. Figure 2 shows the basic architecture of a MI probe integrated into an instrument package.

![Fig. 2. MI probe as part of an instrument package](image)

3. RESULTS FROM RECENT FIELD TEST CAMPAIGNS

A laboratory model of the Huygens MI instrument has been tested in various harsh environments. Figure 3 illustrates the measurements performed in snow. The full and dotted lines link the derived conductivity and relative permittivity at frequencies of 360, 1440 and 5670 Hz. The triangle gives the conductivity obtained at 550 Hz, by measuring the resistance of a slab of material using an alternative method. The square is the estimated relative permittivity of snow at 16 kHz, using an empirical formula. An increase of the conductivity and a decrease of the permittivity with frequency are clear signatures of the presence of water.

![Fig. 3. Huygens MI laboratory model test results](image)

The relatively small error bars, the agreement with the results obtained with other techniques and the continuity of the data plots demonstrate further the quality of the MI measurements.

4. MEASUREMENT RANGE AND ACCURACY

MI probes determine the complex permittivity of a material by measuring at a time both magnitude and phase of the receiver electrode potentials. The measurement range and accuracy for both conductivity and relative permittivity are interdependent. The most important factors for measurement range and accuracy of MI probes are:

- Operating frequency range
- Transmitter signal magnitude range
- Receiver circuit signal to noise ratio
- Signal processing and storage capabilities
- Sensitivity to environmental parameters

Other relevant factors are the size and geometry of the electrode array and the instrument design in general. The measurement accuracy has been modeled taking into account key parameters for the measurement such as magnitude and phase measurement precision, precision of component values used in the models for compensation of systematic errors (like parasitic transmitter current losses and receiver input currents) and electrode array geometry factors. Fig. 4 and Fig. 5 illustrate the results of the accuracy analysis for the laboratory instrument for both conductivity and relative permittivity in a conductivity range of 10\(^{-5}\) S/m to 10\(^{11}\) S/m. The measurement range can be extended towards higher conductivities by using higher measurement frequencies. For low material conductivities, the accuracy of the phase measurement at low frequencies is the most important factor limiting the measurement precision.

![Fig. 4. MI conductivity measurement error](image)
5. LINEAR ARRAY MI PROBE

A new MI probe prototype for application on mobile platforms such as rovers or balloons has been built and tested at ESA / ESTEC. It employs a linear electrode, which can be placed on a surface or dragged behind a vehicle. Components such as high impedance preamplifiers and coupling circuits are integrated into the cable shaped electrode. Such an electrode may sense the presence of subsurface structures (buried layers of materials having very different electrical properties compared to the surface materials, such as ground water level or buried icy layers) down to a depth commensurate with the size of the electrode array. First test results confirmed that the prototype performance is nominal.

6. NEW APPLICATIONS FOR FUTURE MI PROBES AND RELEVANT TECHNOLOGY REQUIREMENTS

MI instruments for atmospheric and stationary surface investigations have been developed and built for ongoing space missions such as Huygens and the Rosetta Lander. New research activities and developments are targeted at future applications on mobile platforms for atmospheric and surface operations (such as rovers and balloons) and for applications on subsurface tools (drills and moles). A possible application scenario of surface and subsurface MI probes is shown in Figure 5. In addition to the measurement of the electric properties of materials, mobile devices can acquire information on the size, spatial distribution and nature of subsurface formations. Due to the specific electrical properties of some target materials, MI probes can be used for specialized applications such as the detection of ice and water [3,4]. In order to satisfy the technological requirements of future applications on missions to Mars, Mercury and other destinations, a number of technology areas have to be addressed, such as specialized electrode configurations, high impedance preamplifiers, high temperature electronics and possibilities for further miniaturization of systems and components.

7. CONCLUSIONS

Mutual Impedance Probes allow accurate measurements of the electric properties of materials. They consist of an electronics package and an electrode array for injecting currents into the medium and measuring phase and magnitude of the resulting potentials. MI probes can be easily integrated into facility-type instrument packages. Instruments have been developed for several ongoing space missions. Their performance has been demonstrated by laboratory and field tests. Depending on the application and on the instrument platform, various configurations of electronics and electrode arrays may be employed. The configuration can be optimized for atmospheric, surface and subsurface investigations. New architectures and configurations for applications on future space missions are being developed, and their technology requirements are being investigated.

8. REFERENCES


MARTIAN IONOSPHERIC STUDY WITH THE CNES ORBITER PREMIER


(1) LPCE/CNRS, 3A Avenue de la Recherche, 45071 Orléans cedex 2, France, Email: m.parrot@cnrs-orleans.fr
(2) ACSE, The University of Sheffield, Amy Johnson Building, Mappin Street, Sheffield S1 3JD, UK, Email: H.Alleyne@sheffield.ac.uk
(3) CETP, 4 Avenue de Neptune, St Maur des Fossés, France, Email: jean-claude.cerisier@cetp.ipsl.fr
(4) ESA/ESTEC/SSD, 2201 AZ Noordwijk, The Netherlands, Email: Detlef.Koschny@rss.esa.int
(5) IAP, Boční II 1401, 141 31 Praha 4 Czech Republic, Email: trav@Alenka.ufa.cas.cz

ABSTRACT

After the pioneering investigation of natural waves and plasma in the Martian environment made by the PHOBOS 2 spacecraft, PREMIER offers the opportunity to build upon this experience. The PREMIER orbiter to be launched in 2007 will permit an exploration down to the ionospheric level. In the long term, it should be possible to study latitudinal, diurnal and seasonal variations. We propose an integrated set of instruments for studying fluctuating electric and magnetic wave fields, and thermal plasma parameters. The main scientific objectives of this consortium named PLASMA PACKAGE are related to the following DYNAMO scientific objectives: Surface-atmosphere interaction, and interaction of atmosphere with interplanetary medium. DYNAMO is a payload for studying atmospheric escape and mapping magnetic and gravity fields.

1. INTRODUCTION

As part of DYNAMO [9] onboard the PREMIER orbiter of Mars we propose a set of instruments, the Plasma Package, for the survey of the upper atmosphere and its connections with the interplanetary medium. The Plasma Package shall thus contribute to an extended study of the couplings between the lower neutral atmosphere, the ionosphere, and the shocked solar wind. It shall also give information on the erosion of the Mars’ atmosphere by the solar wind. The Martian Ionospheric electron Density, Speed, and Temperature investigations (MIDST) experiment shall reliably and accurately measure the density, and temperature of the thermal electron population over wide ranges with a suitable time resolution. To achieve the MIDST objectives, two complementary techniques are actually required, the mutual impedance (MI) and the Langmuir probe (LP). These techniques shall also allow the radio-frequency electric field spectrum (MI), one component of the flow velocity (MI), the spacecraft potential (LP), and possibly the integrated solar EUV flux (LP) to be monitored all along the PREMIER orbits around Mars. The Magnetic Search Coil (MSC), experiment shall study fluctuating magnetic field waves. Finally, the Wide Angle Imager (WAI), shall detect dust storms and meteors.

2. SCIENTIFIC OBJECTIVES

The main scientific objectives of the DYNAMO package that will be aboard the PREMIER orbiter of Mars is to study the atmosphere of the red planet in a wide range of altitudes, latitudes, longitudes, and solar zenith angles. Seasonal effects which are known to be important at Mars will also be paid attention to. Here, the term “atmosphere” must be interpreted in its broadest sense, that is including the ionosphere, and the exospheric neutral and charged particles of planetary origin, which extend deeply into the interplanetary medium and strongly interact with it. This is particularly true at Mars due to its low gravitational field compared, for example, with those of the Earth and Venus.

The lowest altitude reached by PREMIER (150-170 km) will not allow the inner part of the Martian atmosphere to be monitored in situ, nevertheless its extended exploration of the ionosphere should compensate that. The study of the ionosphere of a planet is indeed an essential part of the understanding of the atmosphere of this planet, the ionosphere being actually a tracker of the neutral atmosphere. As a matter of fact, changes in the neutral atmosphere such as variations of density or neutral wind has usually an immediate effect on ionospheres. Since the induced magnetic field of solar wind origin deeply penetrates thanks to charged particles, down to about 150 km (model prediction), the ionosphere should also act as a catalyst in the coupling between the solar wind and the upper atmosphere.

We know from previous missions to Mars that the interaction between the solar wind and the atmosphere is, at Mars, intermediate between that of Venus and those of comets, in particular during high solar activity [4, 5, 17, 18, 24]. The Phobos 2 and, mostly, the Mars
Global Surveyor (MGS) missions have proved that no Martian magnetic field could be invoked to significantly stand off the solar wind [1]. Therefore, the Mars-solar wind connections are more like those of Venus than of those of the Earth. At Venus, dissipationless currents flow in a thin layer, below the ionopause and produce magnetic fields that lead to stand off and divert (drape) the solar wind around the ionopause. When the ionospheric thermal pressure is less than the incident solar wind ram pressure, the ionosphere of Venus is also permeated by the draped interplanetary magnetic field. In this case, no sharply defined ionopause is currently observed at Venus. At Mars the pressure imbalance seems to prevail, therefore by analogy with Venus this could explain why no distinct Martian ionopause drop-off is actually seen. Magnetic structures called flux ropes have been identified in the ionosphere of Venus, they should also be observed at Mars.

Nevertheless, the Venus-type ionospheric interaction could not strictly apply to the Mars-solar wind interaction. Multiple magnetic anomalies have indeed been discovered by MGS. These complex and strong anomalies, which are thought to be the signature of a past intrinsic field fossilized in the crust, could play a significant role at Mars [1]. They could, for example, bear a part in the unshielding of the surface of Mars from the induced ionospheric magnetic field. Neutral particles from Martian origin are known to escape above the exobase (~200-km altitude), whenever their upward velocity, is larger than the velocity for escape from Mars (5 km s\(^{-1}\)). They participate in the solar wind erosion of the Mars’ atmosphere, through photoionization and pick-up processes, like the ones that are responsible for the interaction of comets with the solar wind [20, 22]. Finally, a plasma transition region, called planetopause, magnetopause, ion-composition boundary, protonopause, and lately Magnetic Pile-up Boundary (MPB), has been brought to the fore by Phobos 2 and MGS. This new plasma boundary could be the actual obstacle to the solar wind flow rather than the ionopause. Observing the location and variability of such plasma boundaries shall therefore give insight into the large scale interaction of the Mars’ space environment with the external medium.

Although the broad lines of the Martian space environment and its links with the outside world seem to be understood, actually little is known about the near-Mars plasma environment, in particular, the nightside ionosphere of Mars remains unexplored. This is due to the limited number of missions devoted to this subject, the very few parameters that have been measured, the insufficient ranges of measurements, and the very poor space and time coverage of observations [19, 20, 27, 28].

With Mars Express and Nozomi, PREMIER offers the opportunity to study in great detail the structure, dynamics, and energetics of the upper layers of the Mars’ atmosphere and possibly the variations that occur as a function of solar wind, solar flux, and season variations [8, 18]. It is worth saying again that upper atmospheric regions are of critical importance to an overall investigation of the evolution of planets, and their ability to sustain water and may be life because ionospheres and atmospheres modulate the exposure of planetary surfaces to charged particles and electromagnetic radiations.

### 2.1 MIDST scientific objectives

MIDST is designed to monitor the thermal component of the Martian ionospheric electron population, to survey the radio-frequency electric field component of natural waves, and to measure the spacecraft potential. It shall also appraise one component of the plasma flow speed and possibly the integrated solar extreme ultraviolet (EUV) flux. MIDST shall operate all along the PREMIER orbits around the red planet, and especially the DYNAMO dedicated elliptical orbit. It shall therefore fulfill parts of the main scientific objectives assigned to the DYNAMO package. More specifically, MIDST will address the following objectives.

*To study the couplings between electrons, ions, and neutrals and to scrutinize the processes that control the evolution of the upper atmosphere (convection, transport, ionization, recombination).* Atmosphere ionized constituents, in particular electrons, are known to couple with ionized and neutral gases and to influence processes that command the evolution of the upper atmosphere of Mars. MIDST shall then be useful to look into these couplings. It shall indeed provide reliable and accurate in situ measurements of the thermal electron density and temperature in the upper part of the atmosphere of Mars, namely the ionosphere and above the regions. MIDST shall then, for example, allow the physicochemical processes that lead to the formation, expand, and evolution of the ionosphere to be analysed. The electron density and mostly the electron temperature are critical parameters that govern loss processes of planetary ions. Furthermore, the temperature measurements shall address the problem of electron energetics. A large data set together with the help of numerical modelling tools shall allow assessing mechanisms that direct electron temperature heights. Measuring the electron temperature with accuracy is therefore essential to evaluate the efficiency of the plasma-atmosphere coupling processes. Radio-frequency waves have never been measured below 830 km, the lowest altitude reached by the Phobos 2 spacecraft. Characterizing the spectral
distribution of natural waves in a frequency range from 2 kHz up to 4.5 MHz, which cover the electron plasma frequency range expected above 130-km altitude, shall lead to correlative studies with particles detector observations. High frequency waves have been observed together with anisotropy of particle distribution functions, non-Maxwellian populations, beams, or particle flux enhancements. They are often involved in heating, thermalization, and dynamics processes. Consequently, they shall be crucial for the plasma regimes and encountered regions to be identified and characterized.

To contribute to the understanding of the interaction between the upper atmosphere and the interplanetary medium. The horizontal magnetic field that is created below the ionopause and the magnetic field that penetrates into the ionosphere whenever the solar wind ram pressure exceeds the ionospheric pressure (including structures like flux ropes and nightside holes) shall be overseen in sizing the plasma scale height. Measuring in situ the electron density profiles and their evolutions shall therefore be crucial. In turn, the induced magnetic fields strongly influence the thermal equilibrium of the ionospheric electrons, which shall be characterized by the MIDST electron temperature profiles determination. The ability of the MIDST experiment to estimate the plasma flow velocity shall allow us to survey the anti-sunward transport of the ionosphere due to day-night pressure gradients or to look into the flow from the upper to the deep ionospheric layers.

Plasma boundaries, for example the ionopause and magnetic pile-up boundary (see next section) are the large scale signatures of the Mars-interplanetary medium interaction. Their locations and shapes indeed depend on the solar EUV flux, solar wind dynamic pressure, and possibly interplanetary magnetic field orientation. Studying these boundaries is consequently one of the key direction to learn about this interaction.

To determine the topology of the planetary ionized environment and its response to the solar wind variations. Radio-occultation and landers observations have shown that the altitude of the electron density peak of the Martian ionosphere was 125 km in the subsolar direction and reached 170 km at the terminator. The subsolar and terminator topside terminations of the ionosphere (ionopause) have not been observed in situ, but they should be at respectively 300 km and 1000 km above the planet surface. The MIDST measurements shall thus determine the ionospheric extent and its variations. It is worth noting that for example the terminator ionopause height is assumed to control the nightside ionospheric supply. Although it should be marginal, dust storms are known to heat the lower atmosphere, and thus to increase the altitude of the ionospheric density peak by 20-30 km.

The magnetic pile-up boundary, which is thought to be the Martian obstacle to the solar wind flow is typically located at 660-km altitude in the subsolar direction and 1250 km in the terminator plane. MIDST shall therefore determine the locations of these plasma boundaries and scan their variations as a function of the solar EUV flux and solar wind changes. The Phobos 2 mission have shown that radio-frequency electric field and spacecraft potential measurements are essential to study plasma boundaries and to identify the encountered regions. The additional plasma density, electron temperature, and flow velocity determined by MIDST shall provide new information about heating, compression, and motion. MIDST shall also characterize the encountered plasma regimes and detect ionospheric irregularities and structures, plasma clouds, Venus-like holes, and streamers.

To provide constraints to the study of the erosion of the Martian atmosphere and the escape of its constituents into space. Erosion processes produced by the solar wind have long been recognized as important in the quantification of the Martian atmospheric mass budget. Thermal escape of hydrogen, nitrogen, carbon, and oxygen atoms and mostly transfer of momentum of the solar wind to the planetary upper ionosphere indeed represent a viable way for the atmosphere to loss mass. As the electron density and temperature are key parameters to evaluate the efficiency of these loss mechanisms, the MIDST shall participate in the atmospheric erosion scrutiny. For example, electrons are thought to govern the formation of energetic oxygen atoms, through dissociative recombination of O3+, which are the majority ions in the ionosphere of Mars. These energetic oxygen atoms, whose velocities are very large, even larger than the Martian escape velocity, may then be ionized and picked up by the solar wind. Frictions are exerted on the Martian wake plasma by the solar wind, they also lead to ionospheric erosions, through the formation of thermal plasma clouds and streamers that should be detected by the MIDST.

2.2 MSC scientific objectives

- Study waves in the upper atmosphere of the planet and application of the wave measurements as a tool for studying properties of the Mars magnetic field.
- Study of the atmospheric electricity by detecting electrostatic discharges in the dust storms: Major dust storms which can last up to several weeks occur on Mars. Electrification in dust clouds during storms is a known phenomena in the low terrestrial atmosphere. Data show that dust storms are electrically active and produce remarkable electrical perturbations in the
fair-weather electrical state of the atmosphere. An electrostatic discharge in gases can occur when the potential difference is equal to a given breakdown voltage, but the characteristic parameters of the Martian atmosphere are quite different from those on the Earth. The Martian atmosphere is mainly composed of carbon dioxide CO₂ (> 95 %), surface pressure is lower than on the Earth - of the order of 5.6 mbar (560 Pa), and temperature is between 140 and 300 K. The potential difference required for breakdown in gases decreases with gas pressure, - Paschen's law. Reference [7] measured the breakdown potential for different CO₂ densities, and their results show that for an atmospheric density ~ 2 \( 10^{15} \, \text{cm}^{-3} \), potential gradients of only 20-25 kV m⁻¹ are required for breakdown in the Martian atmosphere (on the Earth at sea level in a dry atmosphere, this value is 3 \( 10^7 \, \text{kV m}^{-1} \)).

Previous work on dust electrification in the Martian atmosphere only concerns laboratory experiments. Reference [10] performed an experiment to investigate whether dust becomes electrified when agitated in a CO₂ atmosphere at low pressure (10 mbar), and whether any breakdown could be observed as a consequence. They showed visible breakdown and measured potential gradients of about 5 kV m⁻¹. This may be applicable to the low Martian atmosphere and shows that electrification of agitated dust can cause visible breakdown in a CO₂ atmosphere at low pressure.

Under wind action, particles of radius close to 50 \( \mu \text{m} \) take part in frictional electrification in the low Martian atmosphere (up to 10 km). The charge of each sand grain could be \( \sim 1 \, \text{fC per } \mu \text{m of radius} \). On the basis of the known characteristics of dust in the Martian atmosphere, and of past laboratory experiments, the charge that dust particles can acquire through collisions under wind action have been studied [21]. The simulation results indicate that it is possible to reach breakdown in the Martian atmosphere. Electrical discharge is able to generate electromagnetic waves propagating in the atmosphere and then in the ionosphere, and it is possible to compute the accurate waveform to reconstruct the propagational circumstances [11].

- Elucidate some of the mechanisms governing the atmosphere/solar wind interface which are still poorly understood: role of wave-particle interactions and turbulent transport; transfers through the various boundaries; effect of pick-up ions and associated instabilities on mass-loading, etc... The weak gravity and the large ion Larmor radii make these questions much different compared to the known magnetized planets.

2.3 WAI scientific objectives

Meteoroid particles entering the Martian atmosphere will contribute to the chemical composition of its upper regions [29, 30]. A layer centered around 80 km and composed of metals (Fe and Mg) and metallic ions (Fe⁺ and Mg⁺) is predicted. It has however never been detected. This ion component originates from the meteoric deposition itself, charge exchange processes between metals and O₂ and UV photoionization of metals. The observations of the proposed imager will give some important constraints to the modeling of the meteor entry, and also to the modeling to this specific part of the ionosphere of Mars. Therefore the Imager will address one of the scientific objectives of the DYNAMO package, which is the study of the plasma around Mars. The imager will also be used at daytime to continuously provide a very wide-angle view of Mars. This will allow to study the development of dust storms. They provide a tracer for the circulation of the atmosphere. In particular, WAI will address the following scientific questions:

- How many meteoroids and of which size enter the atmosphere of Mars?
- Are there any major meteor streams on Mars?
- What's the difference of the interaction between meteoroids and the atmosphere of Mars and Earth?
- How big is the chance of meteoroids penetrating the atmosphere and potentially impacting the surface?
- What is the extend and movement of dust storms?

Thus, in addition of constraining the ion and metal deposition into the atmosphere, WAI will contribute to dust physics and to the impact risk on the surface. The large-scale study of dust storms in the atmosphere (at daytime) will allow to look at the atmospheric circulation and thus directly address one of the main science goals of the mission. To answer these questions, WAI will detect meteors in the Martian atmosphere. Additionally, it will be possible to detect lightning activities (see section 2.2) and aurorae. This is achieved by obtaining images with typical exposure times of 1 s and analysing them on-board via software for increases in the light level. The images where a possible event is detected will be down linked. From these images, the light curves of the meteors can be measured. Using available particle entry models, the size of the meteoroid that caused the event can be estimated. We modelled the meteoroid environment at Mars and estimate about 0.1 to 10 meteors per hour observing time to be visible, also depending on the activity of any meteor streams. A good average number would be 1 per hour. Assuming 100 % false detections,
we can expect a typical number of 2 detections per hour, maximum would be 20. It will be possibly to highly compress the image data, since the event will only cover a small part of the image.

3. SCIENTIFIC PAYLOAD

The payload of the Plasma Package is composed of 4 sensors:
- MIDST, a Mutual Impedance Probe and a segmented Langmuir Probe,
- MSC, a Magnetic Search Coil,
- WAI, a Wide Angle Imager.

4. MEASURED PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Frequency range</th>
<th>Amplitude range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time resolution</td>
<td></td>
</tr>
<tr>
<td>Electron Density</td>
<td>0.1 s - 5 s</td>
<td>0.1 - 5 (10^3) cm(^2)</td>
</tr>
<tr>
<td>Density fluct.</td>
<td>0 - 2 kHz</td>
<td>±(0.2 - 25)%</td>
</tr>
<tr>
<td>Electron Temp.</td>
<td>1 s - 5 min</td>
<td>30 - 10(^6) K</td>
</tr>
<tr>
<td>Electric field</td>
<td>2 kHz - 4.5 MHz</td>
<td>60 dB dyn. range</td>
</tr>
<tr>
<td>SC potential</td>
<td>1 min</td>
<td>- 10 V, + 10 V</td>
</tr>
<tr>
<td>Drift velocity</td>
<td>1 s - 5 s</td>
<td>100 - 5,000 m s(^{-1})</td>
</tr>
<tr>
<td>EUV flux</td>
<td>1 s</td>
<td>photons cm(^2) s(^{-1})</td>
</tr>
<tr>
<td>Mag. Waveform</td>
<td>0.1 Hz - 2 kHz</td>
<td>80 dB</td>
</tr>
<tr>
<td>Mag. Spectrum</td>
<td>0.1 Hz - 10 kHz</td>
<td>80 dB</td>
</tr>
</tbody>
</table>

5. THE MISSION

For the first time the PREMIER orbiter shall cover altitudes from 170 km (and may be less) to about 1000 km, a wide range of latitudes, longitudes, and solar zenith angles of the Martian space environment. Three PREMIER scientific phases have been identified:
- During the PREMIER scientific “Phase 1” (Polar and circular orbit, 500-km altitude, period of 2 h and 3 min,
- During the PREMIER scientific “Phase 2A” (Polar and circular orbit, 300-km altitude, period of 1 h and 53 min),
- During the PREMIER scientific “Phase 2B” mainly dedicated to the DYNAMO investigations (elliptical orbit, 170-km altitude perigee, 1000-km altitude apogee, period of 2 h and 7 min).

6. THE DATA PROCESSING

The Plasma Package experiments will use the service of the DYNAMO Mission Center which will be located in PARIS. The tasks of this Mission Center are:
* to survey the health of the experiments,
* to decommutate the data,
* to transform the data in physical values,
* to plot Quick-Looks of the data,

* to prepare the TC plan including the memory filling optimisation,
* to put data products at experimenters’ disposal through a web server.

7. SUMMARY

The Martian ionosphere formation, expanse, and evolution are poorly known. For example, the nightside ionosphere has never been explored. The mechanisms by which atmospheric constituents are created, escape to space, and the rates of escape will be investigated. MIDST will make accurate in situ measurements of the electron density, the plasma drift velocity, and mostly the electron temperature that is known to play a fundamental role in these processes. Correlations with the NetLander NEIGE (NetLander Ionosphere and Geodesy Experiment) experiment will be done [2].

MSC will study the fluctuating magnetic field from a few Hz up to 20 kHz. Correlation with the NetLander ARES (Atmospheric Relaxation and Electric field Sensor) experiment will be done [3, 23].

WAI will address the meteors and dust storms monitoring.

The Plasma Package as part of the PREMIER scientific payload shall thus definitely contribute to improve our understanding of the Martian space environment behaviour in space and time.

8. REFERENCES


IO-JUPITER SYSTEM: A UNIQUE CASE OF MOON-PLANET INTERACTION

Anil Bhardwaj(1) and Marykutty Michael (2)

(1) Space Physics Laboratory, Vikram Sarabhai Space Centre, Trivandrum 695022, India, spl_vssc@vssc.org
(2) Engineering Physics, University of Virginia, Charlottesville, VA 22904-4238, USA, mm2eq@virginia.edu

ABSTRACT

Io and Jupiter constitute a moon-planet system that is unique in our solar system. Io is the most volcanically active planetary body, while Jupiter is the first among the planets in terms of size, mass, magnetic field strength, spin rate, and volume of the magnetosphere. That Io is electrodynamically linked to Jupiter is known for nearly four decades from the radio emissions. Io influences Jupiter by supplying heavy ions to its magnetosphere, which dominates its energetic and dynamics. Jupiter influences Io by tidally heating its interior, which in turn drives the volcanic activity on Io. The role of Io and Jupiter in their mutual interaction and the nature of their coupling were first elaborated in greater detail by the two Voyagers flybys in 1979. Subsequent exploration of this system by ground-based and Earth-satellite-borne observatories and by the Galileo orbiter mission has improved our understanding of the highly complex electrodynamical interaction between Io and Jupiter many fold. A distinct feature of this interaction has been discovered in Jupiter’s atmosphere as a auroral-like bright emission spot along with a comet-like tail in infrared, ultraviolet (UV), and visible wavelengths at the foot of Io flux tube (IFT). The HST and Galileo and Cassini imagining experiments have observed emissions from Io’s atmosphere at UV and visible wavelengths, which could be produced by energetic electrons in the IFT. In this paper an overview on these aspects of the Io-Jupiter system is presented, which by virtue of the nature of its electrodynamical coupling, has implications for the extrasolar planetary systems and binary stars.

INTRODUCTION

Io-Jupiter system bears a unique distinction in our solar system. While Io is the most volcanically active planetary body, the Jupiter is the largest planet, has strongest magnetic field, fastest spin, biggest and most powerful magnetosphere, and the densest of the planetary atmospheres. The study of this distinct moon-planet system is important because it is very different from the Lunar-Earth relationship, it helps advance our understanding of basic plasma-neutral-surface interactions, and it has implications to understand similar processes occurring at other places in our solar system and in extrasolar planetary systems.

Jupiter is 5.2 times farther from the Sun compared to Earth, and has an equatorial radius of 71492 km, which is 11 times (1400 times in volume) that of the Earth. Jovian atmosphere is mostly made of the simple molecules of hydrogen and helium with sulfur, oxygen and nitrogen in small amount. Jupiter’s magnetic moment is about 4.3 Gauss-R\textsuperscript{3}J, which is 20,000 times greater than that of Earth, with magnetic field direction opposite to that on Earth and inclination of 9.6\textdegree, which is close to 11\textdegree tilt on the Earth. The general form of Jupiter’s magnetosphere resembles that of Earth with dimensions about 1200 times greater, as the solar wind pressure at 5.2 AU is only 4% of its value at 1 AU. Note that even though Jupiter is only slightly bigger than Saturn, its magnetic field is 4 times as big in each dimension. While the magnetic field of Earth is generated by the iron core, the Jovian magnetosphere is generated by the motion of magnetic material inside the liquid metallic shell. At about 1000 km below the cloud top the hydrogen atmosphere becomes thicker and finally changes phase to become liquid hydrogen. Under the liquid hydrogen layer a layer of metallic Hydrogen exists which causes the Jovian magnetic field. Unlike a dipole at Earth’s core, a quadrupole and octupole also contribute to produce the Jovian magnetic field, which explains the shape of its magnetosphere. Jupiter, with a 10-hour day, is the fastest rotator among the planets. The power for populating and maintaining the magnetosphere of Jupiter comes principally from the rotational energy of the planet and the orbital energy of Io, whereas the power source for Earth’s magnetosphere is principally the solar wind.

Jupiter and its satellites constitute a miniature solar system. Galilean satellites are the important among them, which were discovered by Galileo Galilei in 1610. Io, the innermost among them (while Europa, Ganymede and Callisto are the others in order of their distance from Jupiter), has a radius of 1820 km, which is only 2% larger than that of the Moon. Io orbits Jupiter at a distance of 5.9 R\textsubscript{J} away from Jupiter while Moon is at a distance of 60 R\textsubscript{E} from Earth. Io orbits within Jupiter’s intense magnetic field while the radius of the Earth’s distant magnetic tail...
occurs at about 30 $R_E$. That means while Io constantly couples with Jupiter’s magnetosphere, Moon does not. Io is a volcanically active body due to the tidal heating produced by Jupiter and its orbital resonance with the other Galilean satellites, while the volcanoes on Moon have occurred between 3-4 billion years ago. The volcanoes on Io emit SO$_2$ rich material that is one of the major sources of SO$_2$ atmosphere on Io. Moon has a very thin atmosphere with He, Ar, Na, K, Ne etc as the constituents. A detailed review on Io is presented in [1].

The nature of an atmosphere on Io is of great scientific interest ever since the discovery of the Ionian ionosphere by the Pioneer spacecraft in 1973. SO$_2$ has been discovered in Io’s atmosphere in infrared by Voyager, in ultraviolet by HST and Galileo, and in millimeter wave by ground-based observations [2,3,4,62,63]. The presence of SO has been observed by [5] in the millimeter wave rotation lines and S$_2$ in ultraviolet by [6]. Recently, the presence of SO$_2$, SO and H$_2$S in the exosphere of Io has been inferred from Galileo magnetometer data [7]. Na, K, and Cl in trace amounts are also present in Io’s atmosphere [1,8,9].

The sources of SO$_2$ in the atmosphere of Io are volcanic activity, sunlit sublimation of surface SO$_2$ frost, and sputtering. The major sources are the former two, but the relative importance between them is poorly understood. Sublimation produces a uniform atmosphere, whereas volcanic activity is responsible for the spatial variation or patchiness of the atmosphere of Io. A number of models have been developed to study the sublimation driven, volcanic, and sputtered atmosphere of Io [cf. Ref. 2 for review]. The important photochemical models that describe the vertical structure of the atmosphere of Io are those given by [10,11,12].

As mentioned earlier, Io orbits inside the magnetosphere of Jupiter, and therefore the atmosphere of Io is in constant interaction with the magnetospheric plasma. In the present paper the interaction of Jovian magnetospheric plasma with the atmospheres of Io and Jupiter is discussed in context of Io’s atmospheric emissions and emissions from the footprint of Io flux tube in Jupiter’s atmosphere. The application of the Io-Jupiter system study to the astrophysical and cosmic objects, like binary stars and extra-solar systems, is also discussed.

**IO-JOVIAN MAGNETOSPHERE-IO PLASMA TORUS CONNECTION**

Sporadic radio waves of 22.2 MHz (decametric wavelengths) and constant radiation of frequency from 300 to 3000 MHz (decametric wavelengths) are observed from Jupiter in the late 1950s. The electrodynamic relation between Jupiter and Io was discovered when [13] noticed that these emissions are correlated with the Io’s orbital position. Steady state models were developed initially by [14] and [15]. The orbital velocity of Io is less than the rotational velocity of the Jovian magnetospheric plasma. Therefore the Jovian field lines and the associated plasma stream past Io at a relative speed of 57 km s$^{-1}$, which sets up a potential difference of ~400 kV across Io resulting in a current of ~$10^6$ A that closes through the Jovian and Ionian ionospheres. The existence of such a current system was verified with the observations made by Voyager 1 [16]. The interaction of magnetospheric particles with the Io’s atmosphere and surface causes sputtering and the sputtered ions are picked up by the magnetic field lines and get trapped. Hence, Io’s orbital path is populated with S$^+$ and O$^+$ ions (photochemical products of SO$_2$), which is known as the Io plasma torus. The Io plasma torus is a manifestation of the relation between Io and Jupiter’s magnetosphere. Na, K and Cl are also found in the torus [1,8,9]. The coupling between the plasma torus and the Io’s atmosphere is studied by many groups, the recent one is by [17]. The bombardment of Jovian magnetospheric particles onto the Io’s surface can cause generation of soft x-rays, as recently discovered by Chandra X-ray Observatory [18].

![Fig. 1. A schematic showing the Jupiter, Io, Jovian magnetic field lines, Io plasma torus, and its other three Galilean satellites (not drawn to scale).](image-url)

Recent Galileo observations have improved our knowledge about the Io-Jupiter relation considerably. The particles and field instruments detected strongly perturbed fields, beams of energetic electrons and ions, and a dense, cold decelerated plasma flow in Io’s wake [19,20,21,22]. After having several observations by the Galileo magnetometer, [23] argued that the uncertainty regarding the Io’s magnetic moment cannot be eliminated. From the asymmetries observed by the Galileo plasma wave instrument during the various flybys [24] suggested that the Io’s ionospheric plasma density is being strongly influenced by the magnetospheric plasma flow around Io,
similar to the radio occultation experiment observations reported by [25]. Detection of emissions at the Io footprint in Jupiter’s auroral atmosphere in the infrared, ultraviolet, and visible wavelengths [cf. review by Ref. 26] revealed that the particles associated with Io reach Jupiter. The interactions of magnetospheric plasma with Io’s and Jupiter’s atmospheres produce emissions, which are discussed in the next sections.

MULTI-WAVELENGTH FOOTPRINTS OF IO FLUX TUBE IN JUPITER’S ATMOSPHERE

As stated above, the Io is involved in a complicated electrodynamical coupling with its plasma torus and Jupiter's magnetosphere. The electrodynamical interaction between Io and the Jovian magnetosphere results in an electric circuit that runs from Io along Jupiter’s magnetic field lines and closes through the Jovian ionosphere (near 65° north and south latitude) at each foot of the Io Flux Tube (IFT). Where the particles carrying this current impact the atmosphere of Jupiter, an auroral-like spot of emission results.

The first direct evidence of the IFT footprint was obtained in a near-infrared image of Jupiter’s H$_2^*$ emissions at 3.4 μm (cf. Fig. 2) in 1992 [27]. Subsequently, the IFT signature has been observed at far ultraviolet, by HST (cf. Fig. 3), and visible (by Galileo SSI) wavelengths [28,29, 30,31,32,33; cf. Ref. 34, 26 for reviews]. Part of the Jovian decameter radio emissions, and especially the “S-bursts”, are believed to be directly related to the IFT. The emitted power in IFT footprint is ~0.5×10$^{11}$ W in IR (H$_2^*$ emissions), <~10$^{13}$ W in FUV, ~5×10$^8$ W in visible, and ~10$^{10}$ W in radio wavelength ranges [34]. The radio footprints are low-frequency (~2–40 MHz) radio emissions with specific time-frequency characteristics, both L and S bursts, that originate a few thousand kilometers above the optical spots [35,38].

![Fig. 2. Image of Jupiter at 3.4 μm obtained at NASA-IRTF.](image)

The distinct faint emission feature at the foot of L=6 R$_J$, seen in both N and S poles, is the IFT footprint. The footprint is more clearly visible in the southern hemisphere along with a trail of weaker emission extending tens of degrees in longitude downstream of the Io footprint.

![Fig. 3. HST-STS false-colour image of Jupiter in FUV. The spot emission features, both in the northern and southern polar regions, equatorward of the main auroral oval are the IFT footprints. A weaker tail-like feature extending several tens of degrees along the L=5.9 shell associated with both N and S spots are also clearly seen. (from STScI web site http://oposite.stsci.edu/pubinfo/1998.html).](image)

The total power of multi-wavelength radiation emitted from the IFT footprint is ~10$^{11}$ W, which is <1% of the total auroral emission output of Jupiter [26]. The two main features of the “optical” IFT footprints are:

1. A spot of emission having an elliptical or roughly circular shape. The dimension of the IFT footprint is typically larger than the size of the Io projected on to the planet (since the magnetic flux tubes become narrower with decreasing distance to Jupiter, the 3640 km diameter of Io projects an ellipse of ~120×200 km at Jupiter’s atmosphere).
indicates that the region of Io-Jupiter interaction is somewhat larger than the size of Io.

2. A faint comet tail-like feature associated with the main footprint has also been observed at UV, visible, and IR wavelengths (cf. Figs. 2, 3). This trail of emission follows the Io’s magnetic latitude in the downstream direction along the L=5.9 footprint line of magnetic latitude, and its brightness decreases slowly with increasing distance from the main footprint. The trail is observed to extend to 100° in longitude. This implies that Io’s magnetic footprint extends well beyond the instantaneous magnetic mapping of Io and persists for a few hours after Jupiter’s magnetic field has swept past Io.

These faint auroral trails are seen when Jupiter’s dipole is tilted towards Io. In this orientation, Io is at its highest Jovianentric latitude and is temporarily outside of its plasma torus. There is no perfect explanation for the occurrence of the faint trails associated with the main footprint. At present they seem to be more consistent with the multiple reflecting Alfven wave model [36,37,38], but instead of discrete spots continuous emissions are seen. Within the observational sensitivity of the HST-STIS, no emission extending in the upstream direction from Io is detected in FUV [30]. Recent study [39] indicates that the FUV IFT footprint emissions are excited at higher altitudes compared to the main polar aurora and also suggests that drop in the brightness of emissions along the Io trail is due to decrease in the electron number flux rather than softening of the electron mean energy. The average energy of the electron producing the main FUV footprint is estimated to be ~50 keV.

The IFT footprints are less visible in the northern hemisphere over longitudes were the surface magnetic field is stronger (compared to the conjugate point in the southern hemisphere). This is true because the precipitation is easier in lower magnetic field regions (because of lower mirror point altitudes) compared to higher B-field region. Conversely, radio emissions are more intense over longitudes where the surface magnetic field is stronger, because radio emission generation requires substantial (unstable) reflected electron population. The emission from the IFT footprints provide a very useful “fiducial mark” for evaluating models of jovian magnetic field, particularly close to the planet, since although the longitude of the emission may vary according to the dynamics of the IFT, the latitude of the “optical” spot maps out the L-shell of Io at 5.9 R_J. This wonderful constraint has helped the development of a new model (VIP4) of Jupiter’s magnetic field [40] which is not yet perfect, but much improved for interpreting IFT radio emission morphology [38].

EMISSIONS FROM IO’S ATMOSPHERE

Emissions of atomic oxygen and sulfur from Io in the ultraviolet region were observed initially by [41] using the IUE. The FOS and GHRS instruments on HST have also observed ultraviolet emissions from Io [42]. The HST-STIS has sufficient resolution to study the spatial structure of the UV emissions in detail [43,44,45]. Enhanced UV OI and SI emissions above Io’s limb near its equator, called equatorial spots, are observed at sub- and anti-Jovian positions that move with the changing orientation of Jupiter’s magnetic field and appear uncorrelated with the locations of the volcanic vents: suggesting the relation between Io’s atmosphere and the Jovian magnetosphere.

The visible emissions of Io are observed by Voyager, ground-based telescopes, Galileo, HST, and Cassini. Visible emissions were observed for the first time by Voyager 1 spacecraft [46]. From 1990 onwards the stellar echelle spectrograph at Kitt peak has been observing OI 630 nm emissions from Io [47,48,49]. The OI 630 nm emissions are also observed by HST WFPC2 and STIS [cf. Ref. 44,49]. The visible emissions of Io from Galileo observations are reported by [50,51,52]. These visible emissions consist of blue, red, and green band emissions. Imaging sub-system on Cassini has observed visible emissions from Io in late 2000 and early 2001 on its way to the Saturn [53]. The Io’s visible emissions are also found to be correlated with the orientation of the Jovian magnetic field, thereby reinforcing the relation between Io and the Jupiter’s magnetic field.

The probable sources of UV and visible emissions from Io are the excitation of atmospheric neutrals by photoelectrons, torus plasma electrons, and jovian magnetospheric electrons. Ref [42] have ruled out resonant scattering and recombination as the possible sources since the contributions from them are too low. Ref. [54] developed a Monte Carlo model to calculate the photoelectron-excited intensities of neutral and ionized sulfur and oxygen UV emissions, and SO and SO2 band emissions in the wavelength region 240-430 nm, from the atmosphere of Io and found that the calculated brightness are too low to account for the HST-observed brightness.

Ref. [55] described the interaction of torus plasma electron flux with the atmosphere of Io using a three-dimensional two-fluid plasma model. This work assumes a SO2 atmosphere with a scale height of 100 km and column density $3 \times 10^{16}$ cm$^{-2}$ and an atomic oxygen mixing ratio of 0.1. Though this model could explain the radiation patterns observed by HST-STIS, the observed brightness could be explained only if the mixing ratio of oxygen relative to SO2 was increased to 0.2 in the upper atmosphere of Io.
Intense beams of energetic (>15 keV) electrons propagating parallel and anti-parallel to the magnetic field have been observed by the Galileo Energetic Particle Detector (EPD) [20,21,22]. Ref. [56] have reported the detection of large fluxes of bi-directional magnetically field aligned electrons at lower energies (0.1-5 keV) by the Galileo Plasma Science (PLS) instrument. The energy flux of low energy electrons is ~2 erg cm⁻² s⁻¹, which is about 2 orders of magnitude larger than the energy flux of high energy electrons observed by the EPD. The penetration of these electrons into the atmosphere of Io and subsequently the production of emissions are investigated by [57] using a Monte Carlo Model [58]. The brightness is calculated for two model atmospheres of Io developed by [10,11,12]. The model results showed that the electron spectrum observed by the Galileo-PLS in the energy region 0.1-5 keV are capable of producing the HST- STIS observed emissions if a fraction of these electrons precipitate into the atmosphere of Io. Table 1 presents the intensities of emissions calculated using the electron flux observed by PLS instrument and HST-observed intensities of certain important emissions. Ref [59] used the Monte Carlo model [58] to calculate the brightness of emissions in the visible region observed by the Galileo-SSI. The visible emissions observed by Galileo-SSI consist of blue (380-440 nm), red (615-710 nm), and green (510-605 nm) bands. The probable sources of these emissions are SO₂ band in the blue region, OI (630 and 634 nm) in the red region, and OI (557.7 nm) and Na (589 and 589.6 nm) in the green region. The atmospheric models developed by [10,11] are used in the study and the preliminary results from this study are presented in Table 1. These intensities are calculated using the electron energy spectrum observed by the PLS instrument onboard Galileo. It is seen from the table that the red and blue emissions could be produced by the interaction of the PLS-observed field-aligned electrons with the atmosphere of Io, but this process could produce only 25% of the observed green emissions. Further results will be published soon with an upgraded model and improved input data.

<table>
<thead>
<tr>
<th>Features</th>
<th>Calculated (kR)</th>
<th>Observed (kR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OI (130.4)</td>
<td>5.9</td>
<td>20.5</td>
</tr>
<tr>
<td>OI (135.6)</td>
<td>1.84</td>
<td>7.3</td>
</tr>
<tr>
<td>SI (147.9)</td>
<td>1.44</td>
<td>1.0</td>
</tr>
<tr>
<td>Blue</td>
<td>11.8</td>
<td>6.2</td>
</tr>
<tr>
<td>Red</td>
<td>2.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Green</td>
<td>1.7</td>
<td>2.0</td>
</tr>
</tbody>
</table>

The wavelengths in parentheses of the features are in nm.
*1Observed by HST-STIS.
*2Observed by Galileo-SSI.

### IMPLICATIONS FOR BINARY STARS AND EXTRA-SOLAR SYSTEM

An analogy of the Io-Jupiter system is now being applied to the astrophysical and cosmic objects. A conducting body traversing a magnetic field produces an induced electric field. When the circuit is closed, a current will set up, resulting in resistive dissipation. The Jupiter-Io system therefore operates as a unipolar inductor. Another potential cosmic unipolar inductor could be a planet orbiting around a magnetic white dwarf [60]. These systems have a similar configuration, with the differences being their orbital period and separation, the masses and radii of the two components, and the magnetic moment of the magnetic body.

![Fig. 4. A schematic illustration of a unipolar inductor consisting of a magnetic and non-magnetic white dwarf pair in a close binary orbit. [from Ref. 61].](image)

In fact the generation of electric current between the magnetic object and non-magnetic body result in heating of the atmosphere/surface of the magnetic object where the current touches it and would result in production of emission in the polar region of the magnetic object. This could be one way of revealing a planet around a white dwarf that is difficult to detect otherwise [60].

Ref [61] proposed that binary stars consisting of a magnetic and a non-magnetic white dwarf can also be a cosmic unipolar inductor (cf. Fig. 4). In this model the luminosity is caused by resistive heating of the stellar atmospheres due to induced currents driven within the binary. This source of heating is found to be sufficient to account for the observed x-ray luminosity of the RX J1914+24, and provides an explanation for its puzzling characteristics [61].

Close binaries of this type can have short period and secondaries larger than the planet-size bodies. Provided that the spin of the magnetic white dwarf and the orbital rotation are not synchronized (so that there is a relative
motion between the secondary and the magnetic field lines of the primary) and that the density of the plasma between the white dwarf is high enough, the unipolar inductor will operate.

REFERENCES


NEW CORRELATIONS OF PLATE AND PLUME TECTONICS

1Yu.V.Barkin, 2J.M. Ferrandiz, 2M.G. Ferrandez

1Sternberg Astronomical Institute, Moscow, Email: barkin@sai.msu.ru
2Alicante University, Spain; Email: jm.ferrandez@ua.es

ABSTRACT

New computer methods of the study of the ordered structures in position of Earth formation is developed [1]. In papers [2]-[4] the ordered character of the lithosphere plate positions and their motion was described. Geometric, kinematic and dynamics regularities of the plate motion were discovered. Grid phenomenon of the ordered distribution of planet and satellite formations was described [5]-[8]. In given report some from mentioned regularities have been obtained confirmation by using special method of mathematical-statistical analysis of the non-random distribution of the centres of the Earth’s formations (CF): hot spots, triple junctions of the plates, centres of mass of the plates and others. By axography method a computer search of the Earth surface was realised to determine poles for which the EFC have tendency are located along selected parallels and meridians. As criteria’s of the latitudinal and longitudinal ordering of EFC we use statistical parameters (Qn) and (Qm) which are the sums of the numbers of the Earth CF situated in thin (order 1-2°) meridian or latitudinal strips of the given pole. These parameters are defined with some small step for all surface of the Earth and their values are illustrated in colour figures on the plane «latitude-longitude».

1. ORDERED STRUCTURES OF THE PLANETS AND SATELLITES

In Barkin’s paper [5]-[8] the new phenomenon of the ordered positions of the centres of the different formations (CF) of the planets and satellites was discussed. A tendency of alignment of different CF along parallels and meridians of the special inclined reference systems of the celestial bodies is observed. In another words a concentration of the CF along definite meridians and parallels (we will call their as active parallels and meridians) is observed. This regularity in the planet (satellite) surface structure was discovered in result of elementary analysis of the coordinates of the CF positions.

Above-mentioned regularity can be called as grid phenomenon (or net-phenomenon, or phenomenon of the cellular structure in positions of the geological formations [6]). This regularity has universal character and is observed in formation positions of Earth-like planets and moons. In accordance with this law the centres of definite part of formations of celestial body (volcanoes, hot spots, eruptive centres, triple junctions of the plates, some craters, mountains and others formations) are located closely to selected parallels and meridians of the special planetary reference system (their can be a few). In general case these reference systems are inclined with respect to the planet (satellite) equator. Regularity in the relative positions of the selected meridians (and parallels) also is observed. The angular distances between them are divisible to 2°. Grid-phenomenon of the geological, geographical, geophysical and others structures and formations in first was discovered empirically for the Earth and then was generalised to the others Earth-like planets and satellites: Mercury, Venus, Mars, Moon, Io and some others.

The purpose of the paper is to develop computer methods of the searching of the special inclined reference system for which the mentioned regularity is observed more clearly. Practically we must point out only orientation of the polar axis of this reference system relatively a body reference system (or its pole position). In this sense discussed method can be called as «axography» method. This title was suggested in first by M.G. Ferrandez, which has been carried the main contribution in construction of computer methods of studying of the ordered structures of CF positions [1].

2. AXOGRAPHY METHOD

In accordance with axography method the important parameters of ordering as measure of «quality» for given pole is introduced. Different criteria’s of alignments were suggested: 1. Alignments along meridians of given pole (parameter Qn); 2. Alignments along parallels of given pole (parameter Qp); 3. Alignments along equator (for given pole Qf) and some others. In the paper we will discuss only parallel and meridian alignments.

2.1 Algorithm of axography method

The basic idea is to explore the space of polar axes of reference looking for ones that give the highest number of alignments (which we call as «quality»). Alignments (tried up to date) exhibit spin symmetry, the nodes (2D
particular reference system in the XY plane) of a tentative reference system are less important than the pole. This short cut makes the search algorithm more efficient also. The distribution of the searching poles we can illustrate over the search space. But more easy to turn into a plane picture (Longitude/Latitude grid) with colours to code the goodness of fit. The actual procedure is:

1. Cut the Longitude/Latitude plane into latitude strips of uniform spacing (with a step Delta Lat). And devide each strip in a number of rectangular tiles proportional to the cosine of the latitude. Approximating in this way a uniform angular distribution over the sphere.

2. Now we have a finite set of poles to try (the centers of set of tiles, constricted as explained). For each of these poles we can compute several quality indexes, to turn into a color for the corresponding tile. The distribution of poles cover the sphere with a maximum angular separation related in a simple way to the Delta Lat parameter. In fact, less than twice the value of Delta Lat, with an exact supreme to be computed when circumstances will ask for the small effort to do so.

3. Prior to any subsequent computation, all the spots (triple junctions, Venus montes, etc.) to be analysed, are converted to a (new) Reference System with that pole as Polar axis.

4. The following criteria’s («quality index» for the pole) can be up to date:
   \( Q_r \): For any pair of spots (centers of formations) add 1 point to the quality index if the spots differ than a tolerance (Delta Sep) in Longitude;
   \( Q_l \): For any pair of spots, add 1 point to the quality index if the spots differ less than Delta Sep in Latitude;
   \( Q_s \): For every spot, add 1 point if the absolute value of the (new) Latitude of the spot is less then Delta Sep (i.e. the spot is near the new equator).

   Of course another quality indexes (parameters of alignment) can be suggested for analysis of the ordered structures of the planet and satellites. For example can be used a mixed index by adding \( Q_r \) and \( Q_s \) for each pole. But for first studies it seems preferable to display each index in separate picture.

2.2 Method of analysis of non-random distribution of the formation centres

First confirmation of the ordered positions of the centres of planet (satellite) formations presents directly the axography method. In reality applications of these method let us to determine only a few poles with high index of quality. These pole correspond to definite directions in the body reference system. In general case they are inclined. It means that discussed regularity of the latitudinal and longitudinal alignments of centres is observed in special inclined reference systems. Method of axography let us to determine these important reference systems. First studies of the active pole positions confirm our hypothesis about dynamical origin of the observed regularity [7], [9].

For more effective analysis of the ordered structures of the centres also was realised computer analysis of their non-random distribution. In accordance with method we fulfilled two realisations of the axography method. In first this method was applied to the real distribution of the point-centres and then to random distribution over all sphere same number of the centres. Comparison of obtained results has shown that for random distribution of centres the poles with sufficiently high quality index practically are absent. But for many tested of axography method pictures for the real centres positions a few poles with high value quality indexes are observed. Intensity of quality indexes for active poles surpass similar indexes for random centres positions in 1.5 - 2 times. These facts we consider as confirmation of the global regularity of the ordered latitudinal and longitudinal positions of the formation centres.

2.3 Applications of axography method

Axography method in first were used for searching of the high quality poles of the latitudinal and longitudinal distribution of the centres of the different Earth formations: hot spots, triple junctions, volcanoes, centres of mass of the plates and others. More wide application this method has obtained for analysis of centres positions of the formations for Mercury, Venus, Mars, Moon, Io, Europa, Callisto, Ganymede, Triton and others (see second our paper in this proceedings).

3. HOT SPOTS ORDERED POSITIONS

3.1 Hudson Bay Pole and LRS

On the Fig.1 and Fig.3 the results of the application of axography method to the positions of the hot spots (80 dark points on the plane «latitude-longitudes») are presented. From Fig.1 it follows that the more high value of index \( Q_r \) has the special inclined axis with the Northern pole 63° N, 265° E at the western cost of Hudson Bay. This pole we will call as Hudson Bay Pole (HBP). The polar axis with HBP defines the special inclined reference system of the Earth. But this reference system was discovered earlier as reference system of the ordered position and motion of the lithosphere plates [2]-[4]. It was called as Lithosphere Reference System (LRS).
Another results of application of axography method to hot spots distribution on the Earth surface \( Q_p \) index are presented on Fig.3. This criteria let us to determine pole locations of the latitudinal ordering of the hot spots. In result were found new poles located near far from lithosphere equator (Brazil and Taiwan regions). One pole is situated about Santa Elena island. The axes corresponding to the mentioned on the Fig.3 poles (in first Brazil-Taiwan axis) have important geodynamical meaning. In accordance with developed geodynamical conception \([7], \ [9]\) these poles reflect the relative displacements of the Earth shells, for example the liquid core displacements in geological period of time. More detailed picture of \( Q_p \) pole position shows other correlation's between hot spots activity and tectonic processes \([9], \ [1]\). The pole Santa Elena 15°S, 18°W corresponds to the trend direction of the center mass of the Earth due to subduction of the oceanic plates: 14°6' S, 12°4'W. 2). The pole Gram-Chaco 24°S, 61°W corresponds to direction of displacement of the main Earth shells obtained for their non-spherical models and from gravitational field analysis: 19°7' S, 61°5'W. 3). Mata-Grosu pole 22°S, 56°W corresponds to displacements of the liquid core found from dynamical model of the Earth shell behaviour: 15°S, 30°W. These results confirm important dynamical role of the shell dynamics and corresponding mass redistribution of the Earth for global tectonic process.

3.2 Brasil-Taiwan dynamical axis and genetic correlation's of the plume and plate tectonics

Results of the p. 3.1 and p.3.2 discovery a very depth genetic correlation plate tectonic motion and plum and volcano activity of the Earth. Ordered positions of the hot spots and centres of the plates are observed in one inclined reference system (LRS). Practically here we have obtained evidence that one mechanism controls and dictates plate motion and processes of magmatism \([9]\). The active poles of formation ordering carry very important geodynamical information about relative displacements of the Earth shells in geological timescale.

4. Triple Junction Ordered Positions

On the Fig.4,5 and 6 the results of application of axography method to triple junction positions. Ordered
positions of the triple junctions of the plates were predicted for the Principal Geodynamical Reference System (PGRS) with the northern pole situated near lithosphere equator: 25.5°N, 71°E (North of India), and with southern pole: 25.5°S, 108.5°W (Easter Island) [6].

This empirical result was confirmed by axography method. In reality a mentioned pole has high value of parameter $Q_a$ (of the meridian ordering). But define new poles with similar high values of $Q_a$ were found by computer method (Fig.4). These poles are concentrated near from intersections of the big circles:

1. 37°, 78°W;
2. 27°, 43°W;
3. 60°, 86°E;
4. 60°, 48°E.

Fig.4. Triple junction positions and axography poles ($Q_a$ index).

Observed poles of the ordered positions of the hot spots and triple junctions do not coincide. But independently from relative motion of the plates they also have some correlation's. So the more active pole $Q_a$ of the triple junction ordered positions are situated near meridian 90°W of the LRS. From another hand Taiwan poles of the hot spots ordering (criteria $Q_a$) are situated near the poles of triple junctions ordering (East-Chinese, criteria $Q_a$) The poles of the triple junctions ordering ($Q_a$ criteria) looks as displaced on 55° along corresponding Greenwich parallel with respect to HBP (pole of the meridian ordering of the hot spots) (see. Fig 4, 5).

Fig.5. Triple junction positions and axography poles ($Q_b$ index).

Axography method was applied also to the random distributions of the same number of points (58) on the sphere. In result we have obtained new evidences of non-random positions of the triple junctions (Fig.6). So bigger value of parameter $Q_a$ for triple junction of the Earth is equal 58 and for random positions of the same number of points is about 38. Similarly for $Q_b$ criteria (ordering along parallels) we have corresponding values 123 and 79.

Fig.6. Random triple junction positions and axography poles ($Q_b$ index).

So triple junction positions are characterised by the ordered component. It is an indirect confirmation that plate motion also has ordered character.

Barkin's work was accepted by RFBR grant N 02-05-64176 and Spain Sabatico grant SAB-2000-0235.

5. REFERENCES

1. Ferrandez M.G. and Barkin Yu.V., Sagas about axography method and its applications to the planets and satellites. Email discussion in 2001-2002.


5. Barkin Yu.V., Antipodes of the planets and satellites and mechanism of their formation, Abstracts of papers of 32-th Microsymposium on Comparative Planetology, Moscow, pp. 5-6, 2000.


7. Barkin Yu.V., Mechanism of cyclicity of the natural processes and formation of the ordered structures of the planets and satellites, Abstracts of papers of 32-th
Microsymposium on Comparative Planetology, Moscow, pp. 9-10, 2000.


ORDERED POSITIONS OF FORMATION CENTERS OF THE EARTH-LIKE PLANETS AND MOONS

M.G. Ferrandez(1), Yu.V. Barkin(2), J.M. Ferrandez(1)

(1) Alicante University, Spain, Email: jm.ferrandez@ua.es
(2) Sternberg Astronomical Institute, Universitetskii pr-1, 13, Moscow, 119899, Russia, e-mail: barkin@sa1.msu.ru

ABSTRACT

Elementary empirical analysis of the distribution of the centers of different formations of the Earth-like planets and moons (CPF) leads us to a conclusion about ordered character of their positions. The definite inclined belts of the concentration of the CF, grid phenomenon in their latitudinal and longitudinal distribution, antipodal CF are observed on the surface of the Mercury, Venus, Moon, Mars, Io and others planets and satellites [1]. Here we present some first results of the mathematical-statistical analysis by axography method of the ordered center positions of different formations (montes, dorsa, pateras, albedo features and others) for above-mentioned bodies of the Solar system.

1. AXOGRAPHY METHOD

In last years the new phenomenon of the ordered planetary distribution of the geological structures of the Earth-like planets was discovered [1]. It can be called as grid-phenomenon (or net-phenomenon), or phenomenon of the cellular structure in positions of the geological formations. In accordance with this law the centers of definite part of formations of celestial body (volcanoes, hot spots, eruptive centers, triple junctions of the plates, some craters, paterae, mountains and others formations) are located closely to selected parallels and meridians of the special planetary reference system (their can be a few). In general case these reference systems are inclined with respect to the planet (satellite) equator. Regularity in the relative positions of the selected meridians (and parallels) also is observed. The angular distances between them are divisible to 2\(^a\). Grid-phenomenon of the geological, geophysical, geophysical and others structures and formations in first was discovered empirically for the Earth and then was generalized to the others Earth-like planets and satellites: Mercury, Venus, Mars, Moon, Io and some others.

The authors suggest new methods of the studies and evaluations of the ordered positions of the centers of the planet and satellite formations (CF). The problem is reduced to statistical analysis of the point positions (CF) on the spherical surface of the corresponding celestial body. The central method (it was called as axography method) let us to realize computer search of the special active poles on the sphere for which the CF are characterized the more elegant latitudinal-longitudinal order in their positions. When definite part of the CF located and concentrated along selected parallels and meridians for given pole. As criteria's of the latitudinal and longitudinal ordering of CF we use statistical parameters (Q\(_L\)) and (Q\(_\phi\)) which equal to sum of the numbers of CF pares situated in thin (order 1-2\(^c\)) meridian or parallel strips of the given pole. These parameters are defined with some small step for all surface of the celestial body and their values are illustrated on the plane "latitude-longitude".

The method of computer analysis let us to study the alignment of the CF along meridians and parallels of the special inclined reference systems. In first we define coordinates of the more important poles for which the parameters of the alignment CF along meridians (Q\(_L\)) and along parallels (Q\(_\phi\)) have the more high values and then we compare calculated parameters with their similar values but obtained for random distribution of the same number of the points on the sphere. For more intensive pole we define position of the equatorial planes of the inclined reference systems (or big circles on the Mercury surface with inclination I\(_L\) and with node longitude L\(_\phi\)). For analysis in our reports we have used the Internet Data of USA Geological Surveys about planet and satellite formations.

The method axography for analysis of longitudinal ordering (Q\(_\phi\) criteria) let us to analyze some big circles on the planet surface with intensive concentration of its FC. Earlier a similar empirical method of analysis of the active meridians (in special inclined reference system) were applied for analysis of the order of FC (Mercury, Venus, Mars, Moon, Io and others [1]). The comparative analysis of active meridians and parallels by both methods indicate a existence of the ordered structures of planet and satellite formations. The axography method also confirm a set predictions of the ordered structures (active meridians) for the Earth, Moon, Mercury, Mars and for some Earth-like moons.

2. MERCURY

2.1. Latitudinal and longitudinal ordering of Albedo features

In result of application of this method to the centers of Mercury Albedo Features (CMAF) and to random distributions of the same number of points on the sphere we have obtained some evidences of their
ordering. The latitudinal ordering of CMAF has place with respect to the some inclined axes. We will call their as active axes of the latitudinal or longitudinal ordering of the centers. Poles of the more active axis on the Mercury surface were determined by axography method. They are presented on the coordinate plane “latitude-longitude” as red spots (Fig.1,2). On all pictures of paper (starting from Fig. 1 and Fig.2) along abscissa axis the east longitude are calculated from 0° to 360°.

The coordinates of the northern pole of these IRS (established by criteria $Q_2$) are:

1). 89.5°N, 202.5°W;
2). 38°N, 224°W;
3). 43°N, 3°W;
4). 67°N, 28°W.

The planes orthogonal to the pointed polar axes of IRS define the big equatorial circles (inclination, longitude of node):

1). 0.5°, 112.5°W;
2). 52°, 134.5°W;
3). 47°, 183°W;
4). 23°, 208°W.

The pole 1). 89.5°N, 202.5°W characterizes latitudinal ordering of the CMAF with respect North and South of the Mercury. The fact of the latitudinal ordering of the different formations of Mercury (mons, planitia, dorsum, rupes, vallis, albedo features) with respect to the Mercury polar axis (or pole 1) was predicted earlier in the paper [1].

Position of the pole 2) corresponds to the main giant planetary formation of the Mercury Caloris Planitia (30°5 N, 189°8 W) with diameter 1300 km and Caloris Montes (39°4 N, 187°2 W). Formations: Chechov crater (36°2 S, 61°5 W) with diameter 199 km, Mirini Rupes (37°3 S, 39°5 W) are situated on the opposite side of Mercury (at Southern pole of axis 2).

To the Southern pole of the axis 3). 47° N, 183°W can be corresponded Dostoevscij crater with diameter 411 km (45°1 S, 176°4W).

The fine system of the big circles with high concentration of FC are presented on the Fig.2. They were found with help of the $Q_n$ criteria ($I, L_o$):

I. 46°, 167°W;
II. 81.5°, 36.5°W;
III. 52°, 90°W;
IV. 75°, 338°W;
V. 49°, 285°W;
VI. 84°, 155°W;
VII. 90°, 45.5°W.

2.2. Predicted ordered structures of the Mercury

The active big circles (of the concentrated positions of the centers of different Mercury formations) close to III and V also have been discovered earlier. III. 55°, 85°W; V. 35°, 265°W [1]. So the second from them was called as inclined dynamical equator of the Mercury. It is formed by centers of the following Mercury formations:


Of course we used the restricted list of formations and described result has approximate character.

The first studies of the ordered positions of the Mercury formations were fulfilled in the paper [1]. In accordance with grid phenomenon the centers of the definite formations of the celestial body (Mercury) are located closely to the nodes of the original grid formed by the selected parallels and meridians of the special planetary reference system. In general case these reference system are inclined with respect to the axis of rotation of the corresponding celestial body. Very effective illustrations of the grid phenomenon are given on the coordinate plane “latitude-longitude”. On this plane the grid nodes and consequently the formations centers are located along definite straight lines. These lines intersect the nodes of the grid. In particular these lines can present parallels and meridians of the planet or satellite reference systems (see for example our reports.
about ordered positions of the Mercury, Moon, Io and others formations. For example for Mercury formations in the paper [1] were pointed two similar straight lines of formation centers.

2. Goethe, Haystack Vallis, Solitudo Hermae Trismegisti, Liguria, Solitudo Phoenicis, Solitudo Persephones (meridian 45°E).

Mercury formations also form a few another meridians. Two from them are orthogonal to the above-mentioned big circle and determine the Mercury inclined reference system. Similar belt can be pointed also for the belt with inclination 55° and with longitude of ascending node 85°W. And the belt orthogonal to the IDE of Mercury has inclination 90° and a longitude of the ascending node 175°W (355°W). By the way last big circle is close to the big circle VI.

Full grid picture let us predict the existence of the new big geological formations with centre positions along meridians 200° E, 211° E, 219° E, 245° E, 251° E, 305° E, 330° E, 336° E, 344° E, 352° E.

3. VENUS

3.1. Grid of Venus formations

For Venus a grid phenomenon is illustrated the more clearly. For example the many from mountains of this planet (their centers) presented on the coordinate plane “longitude-latitude” are situated along the definite straight parallel lines [1]. Centers of many others formations of the Venus are located closely to the nodes of the original grid formed by the selected parallels and meridians of the special planetary reference system. These ordered structures of Venus in first were studied empirically with pure geometrical positions and in given work we have confirmed some from them and determined new ordered structure of the planet. As simple example we can point here two groves of formations the centers of which are situated closely to two straight inclined mutually-orthogonal lines:

1. Atanua, Rhea, Samodiva, Rhpisunt, Ithlya, Innini;

In previous papers [1] also were established some inclined belts (meridians) of the concentration of the Venus formations. To compare with results of the our mathematical-statistical study of the Venus formation positions we shortly describe this empirical results. Venus formations are characterized by the more fine and elegant order. Its belts comparatively easy are established in results of analysis of major structures and formation grid. One from predicted big circles had an inclination 35° and a longitude of the ascending node 40° E. Here we present the full list of formations located along of this belt:

Alpha Regio, Utii Hiaita Mons, Nokomis Mons, Yolkai-Estam Mons, Ganis Chasma, Nahas-tsaa Mons, Tkashi-mapo Chasma, Mem Loimis Mons, Zewana Chasma, Spandarmat Mons, Parga Chasma, Tsan Nu Mons, Nephys Mons, Inmini Mons, Dione Regio, Cipactli Mons.

Venus formations also form and very clearly present a few another meridians. Two from them are orthogonal to the above-mentioned big circle and determine the Venus inclined reference system. We omit here the full lists of formations for the belts corresponding to above mentioned meridians. Some from these empirical results were confirmed in given study of the ordered Venus structure.

3.2. Application of axography method

Application of this method to Venus formations and to random distributions of the same number of points on the sphere let us to confirm a existence of the order in positions of Venus mountains. The main phenomena in the latitudinal and longitudinal ordered positions of the centers of Venus mountains are illustrated by the Fig. 3 and 4.

The latitudinal ordering of the mountains has place in special inclined reference system (IRS) in full analogy with phenomena of the ordering of the hot spots positions for the Earth. The northern pole of this IRS has coordinates: 37°N, 285°E (Fig.4). This pole is located in the northern part of the famous Beta Regio not far from Rhea Mons (32°4 N, 282°2 E, 217 km) and opposite pole is located on the South of Aphrodites Tierra near by Juno Chasma (30°5 S, 111°1 E). And our poles can be called as Rhea pole and Juno pole.

Fig.3

Also another pole with coordinates 48°N, 279°E can be pointed but with smaller value of parameter Q. Only 1/7 part of the full number of the Venus mountains is situated in the southern hemisphere of the established IRS (Fig.3). Also a concentration of the mountains

© European Space Agency • Provided by the NASA Astrophysics Data System
along inclined equator of the Venus is observed. This big circle is orthogonal to the polar axis of IRS and has inclination 33° and longitude of ascending node 8°E.

Earlier on the base of analysis of chasmata and another formations position was pointed another big circle with inclination 35° and node longitude 40°E [1].

Another global structures of Venus were observed by analysis of the Qa parameter distribution (Fig.4). From them we point the next more active big circles:

1. 48°, 8°E;
2. 49°, 12°E;
3. 53°, 15°E;
4. 76°, 215°E;
5. 51°, 123°E;
6. 84°, 170°E
and others.

Origin of the ordered structures of the Venus (and another's celestial bodies) we connect with small relative displacements of the Venus shells in geological periods of time [1].

4. MARS

In result of statistical analysis of the positions of the conditional centers of mountains (CF) on the spherical surface of the Mars the new facts of their ordered positions were established. By axography method a computer search of the Mars surface was realized to determine poles for which CF are characterized by latitudinal-longitudinal order with tendency are located along selected parallels and meridians. Calculated values of criteria’s of the latitudinal and longitudinal ordering of CF (statistical parameters (Qa) and (Qa)) with respect to the all possible poles on the Moon sphere are illustrated in standard form on the plane “latitude-longitude” (Fig.7,8). In accordance with axography method these parameters are defined with some small step for all surface of the celestial body and their values are illustrated in color figures on the plane “latitude-longitude”.

4.1. Ordered positions of the centers of Mars montes.

In result of application of axography method to the Mars montes and to random distributions of the same number of points on the sphere we have obtained some evidences of non-random positions of the mountains centres. The latitudinal ordering of the mountains has observed more clearly in some special inclined reference systems (Fig.5).

The coordinates of the northern pole of these reference system obtained by criteria Qa are:


The planes orthogonal to the pointed polar axes of IRS define the big equatorial circles (inclination, longitude of node):

1). 73°, 259°W; 2). 82°, 149°W; 3). 32°, 321°W; 4). 53°, 117°W.

Another inclined big circles with intensive concentration of CF in their vicinities were found by criteria Qa (Fig.6):

I). 52°, 114°W; II). 39°, 272°W; III). 64°, 114°W; IV). 72°, 5°W; V). 73°, 258°W.

Important property of the distribution of the Mars CF analogous to established for the Venus montes was discovered. Approximately 1/7 part of the full number of the Mars mountains is situated in the southern hemisphere of IRS with equator II).

4.2. Predicted ordered structures of Mars

It is worth to remark that the big circles 1) and 2) earlier were discovered in the papers [1], [5] in result of empirical analysis of Mars formations positions.
montes, tholi, paterae and others). The corresponding characteristics of these circles are:

I. 54.0°, 110.0W [5]; 55.0°, 110.0W [1];
II. 35.0°, 290.0W [5]; 35.0°, 270.0W [1].

In the paper [5] also the two other active circles were pointed: 47.0°, 50.0W and 43.5°, 206.0W.

The grid phenomenon also can be illustrated for the different Mars formations. So two gropes of Mars formations are located along two straight lines on the coordinate plane “latitude-longitude” [1]:

1). Labeatis Mons, Olympus Mons, Avernus Colles, Tyrrhena Patera, Tyrrhena Mons, Anseris Mons, Peneus Mons, Peraea Mons, Alpheus Colles, Hellespontus Montes;

2). Chasma Boreale, Erebus Mons, Elysium Mons, Elysium Chasma, Hyblaeus Chasma, Libya Montes.

Main from inclined belts of the Mars formations has an inclination 35° and longitude of the ascending node 290.0W. This belt combine the following Mars formations:

Cydonia Mensa, Nilokeras Mensae, Lunae Mensa, Sacra Mensa, Nilus Mensae, Pavonis Mons, Pavonis Chasma, Tharsis Mons, Arsia Chasma, Arsia Mons, Chryse Montes, Tyrrhena Patera, Arabia Terra.

Similar belt can be pointed also for the belt with inclination 55° and with longitude of ascending node 110.0W. And the belt orthogonal to the IDE of Mars has inclination 90° and a longitude of the ascending node 20.0W (200.0W).

4.3. Antipodal gropes of the montes

We point out also on the antipodal position of the two par groups of Mars montes.

Group A: Gonnus Mons (41.6N, 90.8W), Labeatis Mons (37.8°N, 75.9°W), Pindus Mons (39.7°N, 88.9°W), Tanaia Montes (39.7°N, 90.8°W).

Group B: Anseris Mons (30.1°S, 273.2°W), Ausonia Montes (28.5°S, 260.0°W), Centauri Montes (38.5°S, 263.1°W), Coronae Mons (34.5°S, 271.5°W), Hellas Montes (37.9°S, 260.9°W), Peneus Mons (31.2°S, 273.9°W), Peraea Mons (31.3°S, 273.9°W), Tyrrhena Mons (24.5°S, 258.7°W).

The average values of latitudes and longitudes of the CF from gropes A and B are distinguished on 7.6° and 180°=0.3°. The next group of mountains C: Ascræus Mons (11.3°N, 104.5°W), Geryon Montes (8.0°S, 80.7°W), Pavonis Mons (0.3°N, 112.8°W), Tharsis Montes (2.8°N, 113.3°W), Echus Montes (6.8°N, 78.2°W) is antipodal to the Lybia Montes (2.7°N, 271.2°W). The average values of latitudes and longitudes of the CF from gropes C and D are distinguished on 10.2° and 180°+6.7°. A possible mechanism of origin of the observed ordered Mars structures is discussed.

5. MOON

The confirmations of the ordered structures in position of the centers of the lunar formations (CF of mares, montes, rima, dorsa and some others) were obtained by axography method and by method of the mathematical-statistical analysis of their coordinates in the Moon reference system. Some results of the computer calculation of criteria’s of the alignment along meridians (Qₐ) and along parallels (Qₐ) are presented in standard form (Fig.7, Fig.8).

5.1. Ordering of the Moon formations

The main poles of the latitudinal ordering of the Moon CF are: 57°S, 16°E; 81°N, 129°E (montes), 11°N, 19°E; 56°N, 33°W (mares, Fig.7), 73°N, 141°E; 56°N, 150°E (dorsum, Fig.8), 29°N, 10°W (rimae) and others.
In particular, the developed methods have been let us to construct the big circles of the intensive location of the CF (\( \Omega_c \) and \( L_c \) are their inclination and longitude) were determined:

1. 32°, 125°W;
2. 57°, 73°E;
3. 33°, 74°W and others.

These PC big circles are in good accordance with more earlier empirical constructions of the big circles of the location of mares and others formations of the Moon:

1. 30°40', 110°12°W (Franz, 1912, [3]);
2. 54°, 110°E (Lipskij, Rodionova, 1972, [4]),
3. 55°, 113°E (Barkin, 2000, [1]);
4. 36°, 82°W (Lipskij, Rodionova, 1972, [4]),
   35°, 73°W (Barkin, 2000, [1]).

Comparative analysis of the observed locations of the Moon CF with random experimental distribution of the same number of the CF have let us to shown a existence of the ordered (regular) component in observed distribution of the Moon CF concretely in terms of numerical characteristics of parameters \( \Omega_c \) , \( \Phi_c \) and others.

6. IO

In result of statistical analysis of the positions of the paterae centers (PC) on the spherical surface of the Io the new facts of their ordered positions were established. By axography method a computer search of the Io surface was realized to determine poles for which the PC are characterized by latitudinal-longitudinal order (it means that definite part of the PC have tendency to locate along selected parallels and meridians for given pole). As criteria’s of the latitudinal and longitudinal ordering of PC we use statistical parameters \( \Omega_c \) and \( \Phi_c \) which equal to sum of the numbers of PC situated in thin (order 1-2°) meridian or parallel strips of the given pole. These parameters are defined with some small step for all surface of the celestial body and their values are illustrated on the plane “latitude-longitude”.

6.1. Io ordered structures

In result of application of this method to the Io paterae and to random distributions of the same number of points on the sphere we have obtained some evidences of non-random positions of their centers (Fig.9).

The latitudinal ordering of the mountains has place in some special inclined reference systems (IRS). The coordinates of the northern poles of these IRS were obtained by criteria \( \Omega_c \): 1). 67°N, 353°W; 2). 38°N, 237°W; 3). 57°N, 225°W; 4). 88°S, 120°W and others. The planes orthogonal to the pointed polar axes of IRS define the big equatorial circles (inclination, longitude of node):

1). 23°, 263°W ; 2). 52°, 327°W
3). 33°, 315°W ; 4). 5°, 227°W.

Another inclined big circles with intensive concentration of the paterae centers in their vicinities were found by criteria \( \Omega_c \) and some other modified criteria:

I). 55°, 291°W; II). 38°, 291°W;
III). 73°, 35°W; IV). 88°, 110°W;
V). 73°, 258°W.

Obtained results point out on highly ordered PC positions with respect to the Io northern and southern poles. This phenomenon earlier was predicted and illustrated in the paper [1] as grid phenomenon for Io paterae positions.

In accordance with grid phenomenon the centres of the definite Io formations have tendency to located along some parallel straight lines on the plane “latitude-longitude”. We point out here on the two chains of formations:

1. Tvashtar Catena, Volund, Donar Fluctus, Ot Mons, Pele, Danube Planum, Echo Mensa;
2. Chalibis regio, Amirani, Maui, Prometheus, Dorian Montes, Epaphus Mensa, Nemea Planum.

In previous papers [1] also were established some inclined belts (meridians) of the concentration of the Io formations. The plane of dynamical equator for Io has an inclination 35° and longitude of the ascending node 35° W. Two another meridians are orthogonal to the above-mentioned big circle and determine the Io inclined reference system. Similar belts can be pointed also for the belt with inclination ± 55° and with longitude of ascending node 215°W. And the belt orthogonal to the IDE of Io has inclination 90° and a longitude of the ascending node 125°W (305°W). We omit here the full lists of formations for the belts.
corresponding to above mentioned meridians. Some from these empirical results were confirmed in given study of the ordered Io structures.

7. Conclusion

Results obtained by axography method, applied to the observed center positions of the different formations and to their model random distribution, confirm the phenomenon of the latitudinal-longitudinal ordering with respect special inclined reference systems and confirm set of earlier predicted results: grid phenomenon, inclined belts of the active concentration of the planet (satellite) formations and others.

Observed correlation’s in positions of the active pole of ordering with fundamental planetary formations (for the Earth, Venus, Mars and others) illustrate dynamical origin of discussed phenomena. For dynamical explanation of the observed ordering of the celestial body formations we suggest the mechanism of the relative displacements of its shells due to gravitational action of the external celestial bodies [1].

Barkin’s work was accepted by grants of RFBR 02-05-64176 and Spain Sabatico grant N SAB2000-0235.

References

1. Barkin Yu.V., Abstracts of papers of 32-th Microsymposium on Comparative Planetology, pp. 5-12, 2000. a). Antipodes of the planets and satellites and mechanism of their formation, pp.5-6; b). Formations grids of the planets and satellites, pp. 7-8; c). Mechanism of cyclicity of the natural processes and formation of the ordered structures of the planets and satellites, pp. 9-10; d). Inclined belts of the planet and satellite formations, pp. 11-12.
3. Franz J., Der Mond. 2 Auflage, Leipzig. 75 p., 1912.
CORE FORMATION: A NEW MODELLING APPROACH

Ruth Ziethe¹, T. Spohn¹, and S. Turek²

¹Institut für Planetologie, University of Münster, 48149 Münster, Germany
²Institute of Applied Mathematics, University of Dortmund, 44227 Dortmund, Germany

ABSTRACT

Core formation in terrestrial planets is still not well understood although this process is of importance for our understanding of the thermal evolution of a planet and the history of its magnetic field. Because core formation is among the earliest processes in planet formation and evolution, the initial conditions for thermal evolution models are, to a significant extent, determined by this process. The initial temperature of the core and its state are determined by the amount of energy dissipated during core formation. One possible scenario for the formation of a planetary core is the settling of liquid iron from a solid matrix (Stevenson, 1990). Assuming that a planet in the late state of accretion has a magma ocean, there soon will form a layer of molten iron at the bottom of the magma ocean. Since the iron has a higher density than the underlying planetary mantle, it will probably sink due to Rayleigh–Taylor instability. According to Woidt (1978) the sinking iron will attain the shape of spheres because the viscosity of the liquid iron should be much smaller than that of silicates. We model the Stokes falling of an iron sphere through a silicate mantle with temperature dependent viscosity of the mantle material by using a finite element code (FEATFLOW) written by Turek (1998). We solve the incompressible Navier–Stokes equation coupled with the energy and mass equation. With these models the effect of the temperature dependence of the silicate rock viscosity on the differentiation rate and the temperature of the core after core formation can be estimated.

Key words: Core formation, planetary interiors, Computational Fluid Dynamics.

1. INTRODUCTION

Core formation is an important albeit still little understood process of differentiation for terrestrial planets and satellites. While it is still not entirely clear whether or not the Moon has a small iron core of perhaps 400 km radius (Konopliv et al., 1998) it is widely held that the terrestrial planets all have iron rich cores. For Mars this is implied by the moment of inertia factor of 0.365 that requires a core (Sohl & Spohn, 1997; Spohn et al., 1998). For Mercury, a core is necessary to explain the high density of the planet of 5340 kg m⁻³ (Vila et al., 1988). While the Earth’s core is well established, there is no direct evidence for a core in Venus. But it is commonly assumed that Venus has a core in analogy to its sister planet Earth. Even the inner three of the Galilean satellites of Jupiter Io, Europa, and Ganymede are likely to have cores as the recent gravity data from the Galileo mission suggest (see for a recent review Sohl et al., 2002).

It is commonly assumed that core formation is among the earliest processes in the interior of a planet. W-Hf isotope studies show that for Mars and Earth core formation occurred soon after accretion and took only a few tens of million years to complete (Lee & Halliday, 1997; Halliday & Lee, 1999). However, there is some indirect evidence that core formation in the Galilean satellites may have taken much longer, perhaps gigayears (Spohn & Breuer, 1998).

Core formation may have been strongly affected or have even been triggered by collisions not only between planets and small bodies like asteroids or comets but also between planet sized objects. Giant collisions may have melted large parts of a planet or may have even removed its outer shell by vaporization, a possibility cited for the early Earth and Mercury. For Earth, the vaporized material may have helped to form the Moon (Cameron, 1997) and for Mercury vaporization of part of its early outer silicate shell may explain its unusually high density (Cameron et al., 1988). Large amounts of energy were probably dissipated as heat in the very first part of planetary evolution in its outer layers which helped not only to melt the iron but also may have facilitated its path to the deep interior by weakening the solid material of the protoplanet.

1.1. Scenarios for core formation

It is, of course, possible that the core forms during heterogeneous accretion in which case a particular

Figure 1. Simple core formation sketch. a) homogeneously accreted protoplanet, b) during heavy bombardment, the outer shell melts because of the conversion from kinetic energy of impactors to heat and a magma ocean is formed, c) light and heavy (mainly iron) components separate in the magma ocean, the magma ocean freezes and is layered afterwards, d) because of the unstable state of a heavier medium overlying a lighter one, the iron sinks to the planet's centre in a Rayleigh–Taylor instability. On the way to the centre of the planet the migrating iron collects iron from the deeper mantle (possibly migrated through pores).

Core formation-model is not required. Although heterogeneous planet formation cannot be disproved by pure observation it is nevertheless improbable due to many reasons as discussed in Boss et al. (1989), Wasson (1988), Ringwood (1979, 1984), Jacobs (1987). Therefore, we assume in this paper that the planet formed homogeneously and that the core formed later through planet-wide differentiation. According to Stevenson (1990) the following scenarios for the separation of iron from silicate material are conceivable:

- Percolative core formation. The planetary mantle is assumed to be a porous medium – porous on the scale of crystal size – containing finely distributed liquid iron. For a sufficiently large permeability the iron melt is able to migrate through the silicate matrix to form larger melt bodies. The permeability depends on a variety of parameters such as the surface energy between the melt and the solid phases and on the degree of melting or melt concentration. The difference in surface energy between the phases largely determines the dihedral angle of the melt pocket. Only for dihedral angles smaller than 60 degrees can an interconnected melt film between the solid grains form and allow an effective transport of the melt by percolation. For larger angles the droplets will stay isolated and will be trapped by the solid. Unfortunately, the dihedral angle between iron melt and low pressure silicate phases (< ≈ 3 GPa) of is commonly larger than 60 degrees (van Borgen & Waff, 1986) and a continuous melt film is not likely to be possible. For perovskite, the dihedral angle seems to be smaller than 60 degrees (van Borgen & Waff, 1988), meaning that as argued by Stevenson (1990) the existence of a perovskite layer is a condition for core formation with the percolation model. However, if perovskite is necessary, then this model of core formation can only work for the big terrestrial planets Earth and Venus in which the pressure increases rapidly enough for a thick lower perovskite proto-core to form. In Mars, the depth to the perovskite proto-core will only be approximately equal to the depth of the present core-mantle boundary (Sohl & Spohn, 1997). In any case, the percolation model allows for a hot initial core (after core formation) because the surface to volume ratio for the melt is large and effective heating due to viscous dissipation is possible.

- Core formation by rainfall. This model assumes that the planet is completely or almost completely molten after accretion. The iron can easily form drops, which will sink and form a core in the centre of the planet. Although this core formation mechanism is easily understood, it is perhaps rarely applicable since it is difficult to see how terrestrial planets and smaller satellites could ever have been completely molten. An exception is the Moon for which isotope data suggest that at least half the volume was molten (Zitaat, Palme?) probably as a consequence of its formation from a hot vapour cloud.

- Core formation by diapirism. In this model kilometer sized iron melt blobs sink through the solid silicate mantle due to their higher density. The sinking is possible because the solid mantle on long enough time scales (millions of years) undergoes solid state creep and behaves like a very viscous fluid. The sinking starts through a Rayleigh–Taylor instability (compare Figures 1 and 2). Chemical equilibrium between molten iron and silicate rock is not expected if the iron bodies are big enough. Stevenson (1990) argues that the formation of big iron particles is difficult. Moreover, this author argues that convective flow in the planet will disrupt big blobs and even frustrate their formation. However, the diapir model is attractive because it allows the formation of a planetary core on the short time scales suggested by the isotope data without requiring a completely molten planet (Stevenson, 2000). Diapirs can not only form by collection of distributed iron – by using, for instance, the percolation model – but also by formation of an iron layer at the bottom of a magma ocean after impact event during heavy bombardment and following a Rayleigh–Taylor instability.
2. MODEL DESCRIPTION

2.1. Basic equations

We model the flow around a cylinder in a fluid with temperature dependent viscosity and solve the incompressible Navier–Stokes equations coupled with the energy and mass conservation equations. The equations are:

\[ \nabla \cdot \mathbf{u} = 0 \quad (1) \]

\[ \rho \frac{Du}{Dt} = -\nabla p - [\nabla \cdot \tau] + \rho g \quad (2) \]

\[ \rho \frac{DT}{Dt} = -(\nabla \cdot \mathbf{q}) - p(\nabla \cdot \mathbf{u}) - (\tau : \nabla \mathbf{u}) \quad (3) \]

where equation (1) is the continuity, (2) the momentum balance equation and (3) the energy equation. In these equations \( \mathbf{u} \) denotes the velocity, \( p \) the pressure, \( T \) the temperature, \( g \) the gravity acceleration and \( q \) the heat flow. The stress tensor \( \tau \) is defined as:

\[ \tau_{ij} = -2\nu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad (4) \]

\[ \tau_{ij} = -\nu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad (5) \]

For the viscosity \( \nu \) we assume the following weakly temperature dependent viscosity law (e.g., Turcotte & Schubert (2002)):

\[ \nu(T) = \nu_0 \cdot e^{a(T - T_0)} \quad (6) \]

where \( \nu_0 \) is some reference viscosity, \( T_0 \) is the maximum temperature.

2.2. Numerics

To solve the equations (1), (2), and (3) we use the finite element package FEATFLOW written by Turek.
2002ESASP.514D...7F

Figure 3. The setup for the model (left) and the coarse discretization (right).
Left: A single cylinder is placed in an area large enough to avoid influences from the walls. At the walls we use a free slip boundary condition; there is no friction at these boundaries. The cylinder is modeled to have a no slip boundary condition at the surface. Furthermore there is a fixed temperature at the surface of the cylinder (see section 'boundary conditions' for details).
Right: The model is discretized due to a finite element discretization with quadrilateral elements. This grid is refined during the simulation.

(1998). It is a powerful tool to solve incompressible flow problems in nonstationary flows. After discretization and refinement the equations are solved in every element due to a multigrid solver. The setup is shown in figure 3. The coarse grid is shown on the right hand side of figure 3. For the calculation and postprocessing the grid is refined up to four times; 2944 elements and 3048 nodes were calculated. Because of the immense numerical effort we use only a two dimensional model. To study the physical effects this should be sufficient. For the future we plan to extend the model to three dimensions.

2.3. Boundary conditions

At the beginning the initial conditions \( u_0 = u_{00}(x, y, t = 0) \) and \( u_1 = u_{00}(x, y, t = 0) \) are given. In addition conditions at the boundaries are needed for all times. The velocity component perpendicular to the boundary is called \( u_n \) and the component tangential to the boundary \( u_t \). We assume a 'no-slip' condition at the interface between the cylinder and the fluid and we call the outer boundary of the area \( \Gamma_1 \) and cylinder's surface \( \Gamma_2 \).

\[
\frac{\partial u_n(x, y)}{\partial n} \bigg|_{\Gamma_1} = 0, \quad \frac{\partial u_t(x, y)}{\partial n} \bigg|_{\Gamma_1} = 0 \quad (7)
\]

At the outer boundaries of the area we use a 'free-slip' condition, there are no frictional losses along the wall:

\[
\frac{\partial u_n(x, y)}{\partial n} \bigg|_{\Gamma_2} = 0, \quad \frac{\partial u_t(x, y)}{\partial n} \bigg|_{\Gamma_2} = 0 \quad (8)
\]

The inflow condition is given by both velocity components are given:

\[
u_n(x, y) = u_0, \quad \nu_t(x, y) = u_0 \quad (9)
\]

for known values for \( u_0 \). For the outflow condition the velocity components should not change in the direction perpendicular to the wall:

\[
\frac{\partial u_n(x, y = y_{max})}{\partial n} = 0, \quad \frac{\partial u_t(x, y = y_{max})}{\partial n} = 0 \quad (10)
\]

For the temperature we use a Dirichlet' boundary condition:

\[
T \bigg|_{\Gamma_2} = T_0 \quad (11)
\]

meaning that the temperature of the cylinder is given. The model allows to implement a time dependent function for \( T_0 \) if the temperature is to evolve in time. In this work \( T_0 \) is considered to be constant.

3. RESULTS

The flow around the cylinder shows the expected properties with respect to the streamfunction and velocity field. Figure 4(a) illustrates the paths of fluid elements in a flow with low Reynolds number \((Re \leq 1)\). The flow is almost symmetrical upstream and downstream, the right–hand half of figure 4(a) is the mirror image of the left–hand half. The presence of the cylinder has an effect over large distances. Even many diameters away from the cylinder, the velocity is clearly different from \( u_0 \) \((u_0 = 1)\) (4b). It can be shown that the highest velocity occurs at 90 degrees to the accumulation point in front of the sphere and its value is \( u_{max} = 2|u_0| \).

Because of viscosity there is a pressure gradient along the surface of the cylinder. A pressure gradient is

Figure 4. Flow around the cylinder; detailed view close to the cylinder. a) Streamlines for the flow around the cylinder at low Reynolds number. The lines indicate the paths of fluid elements. b) Velocity field around the cylinder. The values are normalized to the absolute value of the inflow velocity.
Figure 5. Temperature (left) and viscosity (right). The heat from the cylinder is transported mostly to the trailing side. As a consequence, a low-viscosity channel forms in the wake of the cylinder.

needed to move the fluid adjacent to the surface against the shear forces.

The heat of the cylinder is transported to its trailing side by the flow. A channel of higher temperature than the surrounding forms in the wake of the cylinder. Because of the temperature dependence of the viscosity a low viscosity channel forms in the wake of the cylinder. (Fig. 5) This causes several effects: the speed of the cylinder will increase, as will the speed of possible further cylinders that may follow the first one.

Because of the temperature dependent viscosity the material close to the cylinder is less viscous compared to the material far from the cylinder, because the highest temperatures in the whole area are situated on the cylinder’s surface. This causes a reduction of the shear forces and the drag force decreases. Figure 6 shows the results for the drag force for various inflow velocities. As expected the drag force is lower for a material where the viscosity decreases with rising temperature (fig. 6 solid line) than the drag force for a material having the constant viscosity of the cold material (fig. 6 dash-dotted line). Since viscous forces operate over large distances the high viscosity of the colder part of the area influences the behaviour of the material close to the cylinder.

That makes clear why the drag force for the temperature dependent case is larger than the calculated drag force for a cylinder surrounded by a material with the (constant) viscosity of the maximum temperature (fig. 6 dashed line).

In figure 7 the drag forces for different viscosity contrasts are shown. A viscosity contrast of 10 is depicted in black, 100 in blue and 1000 in red. The solid lines denote the temperature dependent case and the dashed lines denote the case where the cylinder is surrounded by material with the maximum viscosity. Again the influence of viscous forces on the drag forces are higher if the viscosity contrast is larger.

4. DISCUSSION

In this paper we have shown that the drag force on a cylinder in a fluid flowing around the cylinder is substantially reduced if the viscosity is temperature dependent and if the cylinder is hotter than the ambient fluid. The factor by which the drag is reduced depends on the degree of temperature dependence of the viscosity.

These results can be applied to the problem of a cylinder and, with some reservations, to a sphere sinking in a viscous fluid. In this case, equating the drag force with the body force will allow the terminal velocity of the body under consideration to be calculated. The difference between a sphere and a cylinder is likely to be a numerical factor of order unity. Applying the results of Figure 7 to a diapir then suggests that a diapir of iron of a given size in a protoplanetary mantle will sink two to three times faster if the viscosity of the protoplanetary mantle is mildly temperature dependent with a viscosity contrast of ten. If the latter is by a factor of one hundred, the increase in sinking velocity is by a factor of four to five, and, finally, by a factor of six to seven if the viscosity contrast is by a factor of one thousand. The restricted range of values available today from the numerical calculations in terms of viscosity contrast suggests that a further increase in the temperature dependence of the viscosity will not result in much more than perhaps an order or two of magnitude increase in the sinking velocity even for very large viscosity contrasts. This result is not
unexpected since the momentum diffusion length is many times the radius of the sphere and many times larger than the thermal diffusion length. The consequence of this is that the viscosity far away from the diapir still has a substantial effect on its sinking velocity. For instance, if a 10 km radius iron diapir sinks in a $10^2$ Pa s viscosity mantle by roughly 30 km than the same diapir will be expected to sink 300 km if the viscosity is strongly temperature dependent. This will allow the formation of the Earth’s core in roughly 10 Ma but it is still questionable whether or not the assumed value of the viscosity is applicable and whether or not the diapirs can grow that large. After all the deep interior of the Earth or, even more likely, a small terrestrial planet may be relatively cool and more viscous assumed above. Certainly, assuming in addition a stress dependence of the rheology will help.

In the future we plan to increase the range of viscosity variations by extending the viscosity contrast due to the temperature dependence and by studying the effects of a stress dependent rheology.

Acknowledgements
We would like to thank Dominik Gödeke and Christian Becker for DeViSor, and Rainer Schmachtel for the new CC2D solver. This project is supported by the DFG.

References


MARS GEODESY WITH NEIGE: SIMULATION OF THE MARTIAN ORIENTATION PARAMETERS ESTIMATION

M. Yseboodt\textsuperscript{1,2}, P. Rosenblatt\textsuperscript{1}, V. Dehant\textsuperscript{1}, J.P. Barriot\textsuperscript{3}, and T. Van Hoolst\textsuperscript{1}

\textsuperscript{1}Royal Observatory of Belgium, 3 avenue Circulaire, 1180 Bruxelles, Belgium.
\textsuperscript{2}Email: m.yseboodt@oma.be
\textsuperscript{3}Observatoire Midi-Pyrénées, UMR 5562 / GRGS, 14 av. E. Belin, 31400 Toulouse, France.

ABSTRACT

By analysing the radio Doppler shifts between an orbiter around Mars and a set of landers on Mars surface, and between this orbiter and the Earth, the NETlander Ionosphere and Geodesy Experiment (NEIGE) will provide Mars orientation parameters, particularly polar motion, nutation and length-of-the-day (rotation rate) variations. Information about the martian core will be deduced from nutation estimations.

Due to the small amount of data that will be obtained (a few observations each week), the experiment objectives will only be achieved if an adapted strategy for the parameter adjustment is established. We have simulated the NEIGE data and the modeling of the rotational parameters. After 50 weeks, we obtain a precision in polar motion and rotation speed amplitude estimation at the level of a few mas.

Key words: NEIGE, nutation, rotation, Mars.

1. THE NETLANDER MISSION TO MARS

In 2008, four identical landers dedicated to network science will land on the red planet to study its internal structure and atmosphere. Nine experiments will be on board such as a seismometer, a magnetometer, meteorological sensors and a geodesy experiment. An orbiter will relay data from the station network to the Earth, and will also be used for the geodesy experiment. Several European countries and the USA participate in the NetLander mission. See Harri et al (1999) for more details.

2. NEIGE

The NETlander Ionosphere and Geodesy Experiment, NEIGE, part of the NetLander mission, aims at mea-

Figure 1. Principle of the NEIGE Experiment.

uring the Doppler shifts of two radio-links (1) between the NetLanders and the orbiter and (2) between the orbiter and the Earth (see Fig.1). Mars orientation parameters will be estimated from the analysis of the Doppler shifts, particularly the precession rate (long-term motion of the rotation axis around an axis perpendicular to the ecliptic), the nutations (periodic motions of the rotation axis in space, see Fig.2), the polar motion (motion of the rotation axis in a frame tied to the planet) and the rotation rate variations (length-of-day).

As nutations are influenced by the interior structure of Mars, information about the martian core, such as the core dimension, the density or the physical state, will be deduced from nutation estimations (Dehant et al, 2000, Van Hoolst et al, 2000a and 2000b, Van Hoolst et al, this issue). Seasonal exchange of CO\textsubscript{2} between the atmosphere and the ice caps will also be investigated from the annual and semi-annual variations of the rotation rate (Defraigne et al, 2000).

In addition to these geodesy objectives, ionosphere properties like the total electronic content will be deduced from the perturbation on the radio-signals (Morel et al, 2002). See also Barriot et al (2001) or


© European Space Agency • Provided by the NASA Astrophysics Data System
2002ESASP.514D...7F

Figure 2. Definition of the nutation angles $\Delta\psi$ and $\Delta\epsilon$.

the web site cited at the end of references for more information about NEIGE.

3. MARTIAN ORIENTATION PARAMETERS

In this study, we want to test the quality of the adjustment of the Martian rotation parameters. Due to the small number of measurements that will be available in 2008-09 (1 pass/lander/week), it is appropriate to use only a limited number of parameters to adjust.

Therefore, we evaluate polar motion (PM) not as a temporal series, but as a sum of annual, semi-annual and Chandler frequencies for each component (X and Y polar motion). This means that only 12 parameters (two parameters for each frequency) must be determined instead of about 100 coefficients (one each week during one terrestrial year). The Martian Chandler period is taken equal to 200 days (Van Hoolst et al., 2000a). For the rotation rate variations, the 50 parameters of the temporal series have been reduced to four parameters for the annual and semi-annual frequencies.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>#</th>
<th>Frequency</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar Motion (X)</td>
<td>6</td>
<td>Annual</td>
<td>PX1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semi-annual</td>
<td>PX2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chandler Wobble</td>
<td>PXC</td>
</tr>
<tr>
<td>Polar Motion (Y)</td>
<td>6</td>
<td>Annual</td>
<td>PY1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semi-annual</td>
<td>PY2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chandler Wobble</td>
<td>PYC</td>
</tr>
<tr>
<td>Rotation rate</td>
<td>4</td>
<td>Annual</td>
<td>UT1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semi-annual</td>
<td>UT2</td>
</tr>
<tr>
<td>Nutations</td>
<td>2</td>
<td>$F$ and $\sigma_0$</td>
<td>NUT</td>
</tr>
</tbody>
</table>

Table 1: The 18 rotation parameters that will be adjusted

We didn’t investigate precession modeling in this study, as the precession period is about 175 000 years, which is large compared to the one year period of data acquisition.

Nutation can also be modeled by trigonometric series modeling instead of temporal series (100 parameters if $\Delta\psi$ and $\Delta\epsilon$ are estimated each week during one terrestrial year). The main frequencies to be investigated are the 6 nutations induced by the Sun. Nutations induced by Phobos and Deimos are very small. The amplitude and phases of the rigid nutations can be computed from astronomical theory with a high precision, see Roobbeek (2000) for example. To each rigid nutation, a non-rigid part must be added. In particular, the response of Mars to the forcing of the Sun could be influenced by the resonance of a possible fluid core. This can be mathematically accounted for by convolving (a product in the frequency domain) a transfer function with the rigid nutation series (Folkner et al., 1997, Sasao et al., 1980):

$$r'_m = r_m(1 + F - \frac{\sigma_m}{\sigma_m + \sigma_0})$$
$$p'_m = p_m(1 + F - \frac{\sigma_m}{\sigma_m - \sigma_0})$$

where $p_m$ and $r_m$ are the rigid prograde and retrograde amplitudes of the nutations for the angular frequency $\sigma_m$ (annual, semi-annual,...), and $p'_m$ and $r'_m$ are the prograde and retrograde amplitudes of the nutations after applying the transfer function, $F$ and $\sigma_0$ are the two unknowns, they are a function of the core physical state, size and flattening. $F$ is zero for a solid core and about 0.02 for a liquid core; it mainly depends on the polar moment of inertia of the fluid core. $\sigma_0$ is the frequency of the Free Core Nutation (FCN) which is a rotational normal mode connected with the fluid core. Differences between rigid and non-rigid nutations are shown in Fig.3.

Figure 3. Nutation in obliquity ($\Delta\epsilon$) as a function of time. Nutation with a solid core (Blue) versus nutation with a liquid core (Red). The right graph zooms on the first minimum of the nutation in obliquity (left graph).

<table>
<thead>
<tr>
<th>$F$</th>
<th>No liquid core</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Liquid core with a 1468 km radius</td>
</tr>
<tr>
<td>0.02</td>
<td>Large liquid core with a 1700 km radius</td>
</tr>
<tr>
<td>0.2</td>
<td>Estimated FCN rate with a 1468 km core radius (rad/day)</td>
</tr>
</tbody>
</table>

$\sigma_0 = -2\pi/240$

Table 2: Numerical values for nutation convolution parameters $F$ and $\sigma_0$. 

© European Space Agency • Provided by the NASA Astrophysics Data System
Together with the 12 parameters for PM and the four parameters for the rotation rate, a total of 18 parameters must then be determined.

4. SIMULATIONS

Only a small number of measurements will be made because of power limitations on the landers. In order to evaluate the number of weeks of observation needed to correctly model the interior structure of Mars, we determined the evolution of the rotational parameters precision and of their associated variances with observation time, by using both analytical and numerical simulations.

4.1. Numerical simulations

Synthetic Doppler shifts between the landers and the orbiter, and between the Earth and the orbiter have been simulated over 100 arcs of one week each. The orbiter's initial position and velocity have been adjusted for each arc, while the 18 orientation parameters defined in § 3 have been globally adjusted. The orbitography software used was GINS ("Géodésie par Intégrations Numériques Simultanées" developed in the GRGS Toulouse), which makes estimation of geophysical and orbit parameters through a least squares process. Each frequency is modeled by a cosine and sine amplitude.

![Figure 4](image)

**Figure 4.** Test of the numerical adjustment of four parameters among 18: true errors (in blue) and a posteriori standard deviation (in red), as a function of the observation time.

In Fig 4, true errors and a posteriori standard deviations are plotted as a function of the observational time. The "true error" is defined as the difference between the value of the parameter used for the data simulation and the value coming out with the least square process. "True errors" usually decrease with the observation time. The behaviour during the first observation weeks is chaotic, probably because the amount of available data is too small. For $F$ and $\sigma_0$, we need to iterate the least squares process in order to obtain smaller errors. This means that the adjusted value coming out with the process is used as an a priori model for the next iteration of the adjustment to the simulated data. After a second iteration, the true error is very close to the uncertainty. The configuration used is a network of four landers on Mars (three landers around Tharsis, one in Hellas impact basin), and a 500 km circular heliosynchronous orbit. Observations are performed during one pass/lander/week.

![Figure 5](image)

**Figure 5.** Numerical correlation between the 18 model parameters after 42 weeks for a network of four landers, and a 500 km circular heliosynchronous orbit (93° inclination). See Table 1 for label axis explanation.

![Figure 6](image)

**Figure 6.** Numerical correlation between the 18 model parameters after 100 weeks for a network of four landers, and a 500 km circular heliosynchronous orbit.

Figs.5 and 6 show the evolution of the correlations between the 18 model parameters w.r.t. time. The correlations between parameters decrease with time,
they are still large after 30 weeks, but largely decrease after 100 weeks (around one martian year). From Fig.6, we see that rotation rate and PM components have larger correlations if they correspond to the same frequency.

4.2. Analytical simulations

In this section, the Doppler observable is expressed analytically as a function of configuration parameters (the lander and orbiter positions and velocities) and the rotation parameters between the inertial and the planetary frames. The Doppler expression has been derived with respect to the 18 rotation parameters. A variance-covariance study has then been performed for a particular orbit by evaluating the derivatives once per week per lander. Each frequency is modeled by a phase and an amplitude.

![Figure 7. Standard deviation of four parameters among 18 as a function of the observational time (obtained by a variance-covariance analysis). The red dots are the a priori standard deviations.](image)

We plotted the a posteriori standard deviations as a function of the observational time in Fig.7. One can see that they decrease very fast to small values w.r.t. time. Their behaviour is very similar to the true error after the second iteration and to the uncertainties found in the numerical simulations. We used the same configuration as for the numerical study.

The time evolution of the correlation between 18 parameters has also been investigated (Figs.8 and 9). One can see that they also decrease with the observation time.

For comparison, correlations between 18 parameters after 100 weeks by considering only one lander have been computed by considering the same orbit, see Fig.10. As expected, the correlations between the...
parameters are large because of the missing lander-orbiter geometries.

![Graphs showing F (Nutation) and σ (Nutation)]

Figure 11. Standard deviation of four rotation parameters as a function of the observational time (obtained by a variance-covariance analysis), for different numbers of landers.

In Fig.11, we compare the parameter variances as a function of the number of landers included in the network. We see that some parameters like polar motion components can not be determined with only one or two nearby landers.

5. CONCLUSIONS

Our simulations show that the best strategy in order to have small uncertainties on the rotation angles for the NEIGE experiment is to minimize the number of parameters to solve, given the small amount of data. The different geophysical signals have been adjusted through a sum of frequencies for polar motion and length-of-day. Two parameters (F and σ₀) have been used to evaluate the effect of the core on the nutations.

After 50 weeks, the correlations between the parameters are small, and we obtain a precision in polar motion and rotation rate amplitude estimations of a few mas. This will permit in particular to model accurately the CO₂ sublimation-condensation processes. Furthermore, F and σ₀ will be known with sufficient accuracy to be used for core properties determination if four landers are used.

With only one lander, the correlations between parameters are too large and the precision obtained is not sufficient to meet the NEIGE objectives. Three landers on Mars’ surface give approximately the same results as four landers. If only two geometrically well spread landers (not on the same latitude or longitude, distance larger than 5 000 km) are used, the same precision can be obtained by monitoring the rotation parameters over a longer period of time (some weeks more).

ACKNOWLEDGEMENTS

We would like to thank S. Willems for help with the computations and J. Vienne for his comments. This study was funded by an ESA/Proex 6 contract.

REFERENCES


Dehant V., Van Hoolst T. and Defraigne P., Comparison between the nutations of the planet Mars and the nutations of the Earth, Survey Geophys., 21, 1, pp. 89-110, 2000.


Van Hoolst T., Dehant V. and Barriot J.-P., Interior of Mars from non-rigid nutations, this issue, 2002.

http://ganimede.ipgp.jussieu.fr/GB/projets/netlander/
http://smsc.cnrs.fr/NETLANDER/

© European Space Agency • Provided by the NASA Astrophysics Data System
MASS TRANSFER AND TRITON’S CRATER ASYMMETRY

Marchi S., Barbieri C., Lazzarin M.

Astronomy Department, Vicolo dell’Osservatorio 2 Padova, Italy
Email: marchi@pd.astro.it
Email: barbieri@pd.astro.it
Email: lazzarin@pd.astro.it

1. WHAT IS THE MASS TRANSFER?

By mass transfer we mean the exchange of mass between two satellites belonging to the same satellitary system.
It takes place in the following steps (see Fig. 1):
1) Impact of a satellite (parent body, PB) with an heliocentric body (e.g. comet).
2) After the impact, fragments are released on planetocentric orbits.
3) Part of the fragments collides with another satellites (target, i.e. Triton in the present work).

Fig.1. Sketch of the mass transfer process.

2. ORBITS

Due to the complex history of Triton’s orbit (for this reason we refer to the past of Triton as proto-Triton, PT), we considered many orbits related to each other by the conservation of the angular momentum, all with fixed inclinations. Strictly speaking this corresponds to well sufficient to lead proto-Triton to complete melting. Of course any trace of cratering at this stage would be lost. On the other side, in case of gas drag evolution of proto-Triton, its melting could be avoided.

In our simulation we refer to this case.
The PT’s semi-axes considered vary between 14 and 100 Neptune’s radii.
For the PB, we consider many circular orbits with zero inclinations. These orbits correspond to the positions of the present inner satellites. Their semi-axes vary between 3.5 and 6.5 Neptune’s radii.

3. THE SIMULATIONS

In our simulation the impact has been modeled as a cratering event, so the fragments are released inside a cone with semi-aperture of 40deg.
The ejecta’s velocities module vary uniformly between 0 and 1.5 km/s.
Both the Triton and the fragments evolution happens with fixed (a,e,i), while both Ω and ω circulate uniformly.

4. RESULTS

1) The distribution of impacts is focused on the leading side of proto-Triton (see Fig. 2);
2) the transfer’s duration is always less than 3000y;
3) the mean impact velocities are in the range 10-15 km/s (see Fig. 3);
4) the shape of the impacts’ distribution is only weakly dependent on the orbital configuration of both the PB and proto-Triton;
5) the cumulative crater distribution we infer in our model agrees with the observed one (see Fig. 4).


© European Space Agency • Provided by the NASA Astrophysics Data System
Fig. 2. Density of impacts. The center of the image corresponds to the antapex of motion. The poles are at $\theta_s = 0.180$ deg. The impact density is much greater on the apex. The contours are slightly depending on the PT and PB orbital configurations.

Fig. 3. Impact velocity distributions for some values of PT's semi-axis in units of Neptune's radii. As shown, the impact velocities are depending on the orbital configuration.

Fig. 4. Simulated cumulative craters counts rescaled on the observed distribution (black squares). $D_0$ is the arbitrary dimension of the largest crater that our model produces. Notice that the model goes to zero at about 90 deg.

5. REFERENCES
Session 4
Interiors, Surfaces, Exospheres and Impact Processes
Chair: R. Grard
GEOCHEMISTRY OF MARS BASED ON LABORATORY ANALYSES OF MARS METEORITES AND IN SITU ANALYSES FROM PATHFINDER

Heinrich Wänke

Max-Planck-Institut für Chemie, P.O. Box 3060, D-55020 Mainz, Germany, Email: waenke@mpch-mainz.mpg.de

ABSTRACT

Using laboratory data on the composition of Mars meteorites, it was possible to make reliable estimates of the bulk composition of the Martian mantle and core. According to these estimates the Martian mantle contains, compared to the Earth’s mantle, about twice as much FeO and also higher concentrations of moderately volatile and volatile elements. Also - contrary to the case of the Earth’s mantle - Mn, Cr, and especially P are not depleted in the Martian mantle. All chalcophile elements were found to be highly depleted in the mantle of Mars and a homogeneous accretion of Mars was inferred. The core mass of Mars was estimated to 21%, consisting of Fe and Ni and about 14% S.

The APX-spectrometer on the rover Sojourner returned the first analyses of Martian rocks, which turned out to be rich in silica and potassium, but low in Mg resembling terrestrial andesites. The Martian soil which was found to have almost identical composition on all three sites where analyses were carried out, i.e. Viking 1, Viking 2, and Pathfinder. The soil can be explained as mechanical mixtures of diminuated basalts, compositionally similar to the Martian meteorites, their weathering products, as well as the andesitic component similar to the Pathfinder rocks. The APX-spectrometer was not able to detect carbon above the detection limit of 0.5 wt%, corresponding to about 5% carbonate.

1. GEOCHEMISTRY OF MARS BULK COMPOSITION BASED ON ANALYSES OF METEORITES.

A small group of meteorites represent surface rocks of Mars expelled by large impacts. They can be and have been studied in all details in laboratories on Earth. In that way the situation can be compared what in the future can be expected to be done with samples returned automatically from Mars. The Martian meteorites, previously called SNC-meteorites, comprise a small group of differentiated meteorites (shergottites, nakhlitites and Chassigny). Their total number (at present 26) increases steadily due to new discoveries. According to their oxygen isotopes and trace element ratios all come from one parent planet. The crystallization ages of these meteorites are with one exception strikingly low, between 0.16 and 1.3 Ga. They contain a trapped gas component very different to any observed in other meteorites, but identical to Mars atmosphere both in element and isotope ratios, which are known from the Viking landers.

All SNC-meteorites are igneous rocks of quite variable composition. The high FeO contents of the SNCs reflect the high FeO content of the Martian mantle compared to the terrestrial mantle, while their MnO and Cr₂O₃ concentrations indicate that, contrary to the Earth’s mantle, the Martian mantle is not depleted in MnO and Cr₂O₃. The depletion of Mn and Cr in the terrestrial mantle, although not yet fully understood, must be a special feature of the Earth. Similarly, the much lower abundance of phosphorus than that of other moderately volatile elements in the Earth’s mantle has not yet been explained satisfactorily. However, in the mantle of Mars the abundance of P is in line with that of other elements of similar volatility.

Using element correlations observed in SNC-meteorites and general cosmochemical constraints, Dreibus and Wänke [1] and Wänke and Dreibus [2] have estimated the bulk composition of Mars as shown in Table 1 and Fig. 1. The mean abundance value for

![Fig.1. Estimated abundances of moderately volatile and moderately siderophile and chalcophile elements in Mars’ mantle together with the respective data for the Earth's mantle. Note for the mantle of Mars the higher abundance of volatile and moderately volatile elements Na, Ga, P, K, F, Rb, Cl, and Br and the low abundance of chalcophile elements Cu, Ni, Co, Mo, In, Zn, and Tl.](image-url)
the elements Ga, Fe, Na, P, K, F, and Rb in the Martian mantle were found to be 0.35 and thus exceed the terrestrial value by about a factor of two.

Table 1. Bulk composition of Mars as derived from SNC-meteorites

<table>
<thead>
<tr>
<th>Element</th>
<th>Mantle + Crust</th>
<th>Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgO</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>3.02</td>
<td></td>
</tr>
<tr>
<td>SiO₂</td>
<td>44.4</td>
<td></td>
</tr>
<tr>
<td>CaO</td>
<td>2.45</td>
<td></td>
</tr>
<tr>
<td>TiO₂</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>FeO</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td>Na₂O</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>P₂O₅</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Cr₂O₃</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>MnO</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>305</td>
<td></td>
</tr>
<tr>
<td>Rb</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Cs</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Core mass % 21.7</td>
</tr>
</tbody>
</table>

The most recent addition in Fig. 1 is the abundance of lead in the Martian mantle [3]. From Fig. 1, it is evident that Pb does not behave as a chalcophile element on Mars as it is not more depleted as other elements of similar volatility. Useful abundance values for Pb can only be obtained by precise isotope analysis to distinguish between radiogenic and primordial lead. This is a good example of the importance of sample analysis in the laboratory.

The abundances of elements in the Martian mantle - especially the similar abundances of several geochemically very different elements (W to Rb in Fig. 1) - indicate that the two-component model put forward to explain the composition of the Earth’s mantle is also well suited for Mars. According to the two-component model as formulated by Ringwood [4] and Wänke [5] both Earth and Mars, and probably also Venus and Mercury, were formed by two chemically distinct components but with different mixing ratios for the individual planet. These components are:

(i) Component A: highly reduced and free of all elements with equal or higher volatility than Na, but containing all other elements in CI abundance ratios. Iron and all siderophile elements in metallic form, even Si may be partly present as metal.

(ii) Component B: oxidized and containing all elements - including volatiles - in CI abundances. Iron and all siderophile and, of course, all lithophile elements present as oxides.

Looking at Fig. 1, the low abundances of all elements with chalcophile character in the Martian mantle is obvious. It has been taken as evidence for the extraction of these elements from the Martian mantle by FeS segregation to the core during the accretion process of the planet, which allowed a nearly perfect equilibration of component A and component B. The mixing ratio (component A/component B) was estimated to be 60:40. In the equilibration process two reactions were thought to be of special importance:

(a) Sulphur supplied in the form of FeS and sulphates by component B and metallic FeNi from component A formed a sulphur-rich FeNi alloy:

\[ \text{MgSO}_4 + 4\text{Fe} \rightarrow \text{FeS} + 3\text{FeO} + \text{MgO} \]  (1)

During core formation (segregation of the (Fe,Ni)S alloy) extraction of elements from the mantle took place according to their sulphide-silicate partition coefficients rather than their metal-silicate partition coefficients.

(b) Water from component B reacted with the metallic iron of component A:

\[ \text{Fe} + \text{H}_2\text{O} \rightarrow \text{FeO} + \text{H}_2 \]  (2)

The large amounts of H₂ (corresponding to about 0.4% of the planet's mass) led to hydrodynamic escape of heavier species like rare gases, etc. Only trace amounts of water remained in the Martian mantle. Carbon from component B at least partly dissolved in the remaining metallic iron and was extracted into the core [6].

The high contribution of component B led via reaction (1) to a sulphur-rich core (about 14% S), while the total mass of the core became considerably smaller. Reaction (2) was responsible for the Martian mantle becoming FeO rich but H₂O poor.

The sulphide-silicate equilibrium in the Martian mantle indicates its saturation with FeS. The FeO content of the Martian mantle is about a factor of 2 higher than that of the terrestrial mantle. As a consequence, the sulphur abundance in the Martian mantle is expected to be substantially above the S abundance in the Earth's
mantle as the solubility of FeS in silicates increases strongly with the FeO content. Hence, the observed high concentrations of sulphur in mantle-derived magmas as represented by the shergottites (sulphur content between 600 and 2800 ppm) are not surprising.

The size and the composition of the Martian core, with its over 14% of S as given in Table 1, fits not only well with the geophysical data, i.e. the Martian moment of inertia factor and the planet's density, but also falls within the error to a value of 16% S, obtained by Schubert and Spohn [7] for a completely fluid core that is only weakly convecting, which in turn could explain the very weak magnetic field of Mars.

To estimate the water content of the Martian mantle, Dreibus and Wänke [8] used data on the water content of Shergott 180 ppm measured by Yang and Epstein [9]. Shergott is enriched in La by a factor of 5 relative to the Martian mantle. Assuming a similar enrichment for H_2O, a mantle concentration of 180:5 = 36 ppm was found. This is exactly the value obtained earlier by Dreibus and Wänke [10], comparing the solubility of H_2O and HCl in silicate melts and using the abundance of chlorine in the Martian mantle as deduced from SNC meteorites. The exact match is, of course, purely fortuitous considering the uncertainties.

Over years, the dry Martian mantle as proposed by Dreibus and Wänke [8] has been questioned in light of water-rich inclusions observed in SNC meteorites [Johnson et al., 11; McSween and Harvey, 12]. However, it was not known if the host phases of these inclusions have crystallized from mantle-derived magmas or represent material from a water-rich Martian crust taken up by intrusions and over-plating of mantle-derived magmas. The contradictory evidence of a dry Martian mantle as indicated by the low water content of SNC meteorites and the erosional Martian surface features, which seem to require large amounts of water, has been discussed by Carr and Wänke [13].

The value of 180 ppm H_2O for Shergott used by Dreibus and Wänke [8] for their estimate of the water in the Martian mantle could be considered to be too low as the sample was preheated to 350°C to remove water of terrestrial origin. Karlsson et al. [14] have shown that indeed SNC meteorites give off considerable amounts of water below 350°C, but both H/D analyses [Watson et al., 15] as well as oxygen isotopes [Karlsson et al., 16] indicate a large portion of terrestrial water contribution below 350°C.

In total, Karlsson et al. [14] extracted 640 ppm H_2O from Shergott without preheating and 260 ppm above 350°C. However, the oxygen isotopes indicate that for all the SNC meteorites, apart from the presence of terrestrial contamination, a large fraction of the water, although Martian, is not derived from the Martian mantle, but obviously represents Martian surface water with an oxygen isotope composition up to three times further away from the terrestrial isotope fractionation line than the oxygen in the silicates of SNC meteorites. At the high temperatures during magma generation in the Martian mantle isotopic equilibration between oxygen of the silicates and of water would certainly have been established. Hence, only a fraction of the water found in SNC meteorites can be mantle derived and the other non-terrestrial part must come from the Martian surface. The oxygen isotopes of the surface component might have been created by nonlinear isotope fractionation by non-thermal escape of oxygen space [Jakosky, 17].

A similar scenario as described for Mars might in fact also have happened on Earth. For further details see Carr and Wänke [13]. If the Earth has received most of its present water late in its accretion history, this water might have had oxygen isotopes very different from those of the oxides in the mantle. Later on, subduction and recycling of the oceanic crust might have continuously brought water from the surface into the originally dry mantle and isotopic equilibration with the oxygen of the silicates took place there.

2. GEOCHEMISTRY OF MARS SURFACE BASED ON IN SITU ANALYSES.

The APX-spectrometer (Alpha-Proton-X-ray-Spectrometer) mounted on the rover Sojourner, returned chemical analyses of Martian rocks for the first time. Viking Landers 1 and 2 only returned in-situ X-ray fluorescence (XRF) analyses of soil samples, as neither rocks were within reach of the arms nor devices were on board to acquire and prepare solid rocks.

The APX-spectrometer of Pathfinder was originally designed for the Russian Mars 1996-Mission, which failed during launch. The APXS technique enables us to chemically analyze soils and rocks without any sample preparation. The spectrometer’s sensor head (Fig. 2) simply had to be put against the sample so that it was irradiated by alpha particles emitted by curium-244 sources. The instrument had three modes of operation:

1.) Alpha-back-scattering, known as Rutherford scattering. It is measured by solid-state detectors.
2.) In a few cases the nuclei of the target undergo alpha-proton reactions.

3.) The alpha particles also interact with the electron shell of the target nuclei and give rise to characteristic x-rays, which are analyzed by a solid state x-ray detector. The alpha-back-scatter mode is especially useful for the analysis of light elements C, N and O, while the x-ray mode is advantageous for all elements heavier than Na.

The data for the most reliable because least disturbed soil samples are given in Table 2.

The chemical composition of all soil samples was almost identical, independent of the color and appearance of the ground. Altogether the soil composition at Ares Vallis was very similar to the landing sites of Viking 1 and 2 at Chryse and Utopia. This might well mean that the Martian soil is homogeneous on a global scale, having been distributed and mixed by impacts and storms.

The agreement with Viking data is good in most cases (Table 2). From the measurements of the XRF-spectrometer on board Viking 1 and 2, Clark et al. [18] reported only upper limits of 0.15% K2O, whereas from the APXS of Pathfinder 0.6 ± 0.1% K2O was found in soil samples. It seems likely that Clark et al. [18] might have overestimated their sensitivity considerably; otherwise one has to assume that the much higher K concentrations in the Pathfinder soil are derived from local K-rich rocks.

Because of the huge background due to the CO2 of the thin Martian atmosphere, the sensitivity of the APXS for C was drastically reduced. So that only upper limits of 0.8 wt% C (corresponding to about 5% MgCO3) could be given for both soil and rock samples at the Pathfinder landing site.

In fact, carbonates should not be expected in the Martian soil because of the dominance of SO2 [Wänke and Dreibus, 21; Clark, 22]. Shergottites, the most abundant group of Martian meteorites, contain mantle derived concentrations of ~ 200 ppm H2O, ~100 ppm CO2, and between 1200 and 5600 ppm SO2. Terrestrial MORB contain ~2000 ppm H2O and similar concentrations of SO2 and CO2. On Mars, which is much poorer in H2O and CO2, but similar or richer in SO2, it is expected that SO2 dominates the volcanic gases. At least part of SO2 will be quickly transformed to SO3, which together with water vapor will produce sulfuric acid, which in turn will decompose carbonates and return CO2 to the atmosphere.
Table 2. Composition of soils and rocks at the Mars Pathfinder landing site. Corrected data (August 2000) in weight-\%, normalized to 100\%, based on APXS Mainz re-calibration. The average error given in the last line includes the error due to counting statistics as well as the error from the calibration curves.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Na&lt;sub&gt;2&lt;/sub&gt;O</th>
<th>MgO</th>
<th>Al&lt;sub&gt;2&lt;/sub&gt;O</th>
<th>SiO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>P&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;</th>
<th>SO&lt;sub&gt;3&lt;/sub&gt;</th>
<th>Cl</th>
<th>K&lt;sub&gt;2&lt;/sub&gt;O</th>
<th>CaO</th>
<th>TiO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Cr&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;3&lt;/sub&gt;</th>
<th>MnO</th>
<th>Fe&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;3&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soils</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-4, soil</td>
<td>1.00</td>
<td>9.95</td>
<td>8.22</td>
<td>42.5</td>
<td>1.89</td>
<td>7.58</td>
<td>0.57</td>
<td>0.60</td>
<td>6.09</td>
<td>1.08</td>
<td>0.2</td>
<td>0.76</td>
<td>19.6</td>
</tr>
<tr>
<td>A-5, soil</td>
<td>1.05</td>
<td>9.20</td>
<td>8.71</td>
<td>41.0</td>
<td>1.55</td>
<td>6.38</td>
<td>0.55</td>
<td>0.51</td>
<td>6.63</td>
<td>0.75</td>
<td>0.4</td>
<td>0.34</td>
<td>23.0</td>
</tr>
<tr>
<td>A-10, soil</td>
<td>1.32</td>
<td>8.16</td>
<td>7.41</td>
<td>41.8</td>
<td>0.95</td>
<td>7.09</td>
<td>0.53</td>
<td>0.45</td>
<td>6.86</td>
<td>1.02</td>
<td>0.3</td>
<td>0.51</td>
<td>23.6</td>
</tr>
<tr>
<td>A-15, soil</td>
<td>0.97</td>
<td>7.46</td>
<td>7.59</td>
<td>44.0</td>
<td>1.01</td>
<td>6.09</td>
<td>0.54</td>
<td>0.87</td>
<td>6.56</td>
<td>1.20</td>
<td>0.3</td>
<td>0.46</td>
<td>23.0</td>
</tr>
<tr>
<td>Mean Soil</td>
<td>1.09</td>
<td>8.69</td>
<td>7.98</td>
<td>42.3</td>
<td>0.98</td>
<td>6.79</td>
<td>0.55</td>
<td>0.61</td>
<td>6.53</td>
<td>1.01</td>
<td>0.3</td>
<td>0.52</td>
<td>22.3</td>
</tr>
<tr>
<td>Mean Viking soil*</td>
<td>-</td>
<td>6.4</td>
<td>8.0</td>
<td>47.0</td>
<td>-</td>
<td>7.9</td>
<td>0.5</td>
<td>&lt;0.15</td>
<td>6.4</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>19.7</td>
</tr>
<tr>
<td>Cemented Soil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-8, Scooby Doo</td>
<td>1.56</td>
<td>7.24</td>
<td>9.09</td>
<td>45.6</td>
<td>0.61</td>
<td>6.18</td>
<td>0.55</td>
<td>0.78</td>
<td>8.07</td>
<td>1.09</td>
<td>--</td>
<td>0.52</td>
<td>18.7</td>
</tr>
<tr>
<td>Rocks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-3, Barnacle Bill</td>
<td>1.69</td>
<td>3.20</td>
<td>11.02</td>
<td>53.8</td>
<td>1.42</td>
<td>2.77</td>
<td>0.41</td>
<td>1.29</td>
<td>6.03</td>
<td>0.92</td>
<td>0.1</td>
<td>--</td>
<td>16.2</td>
</tr>
<tr>
<td>A-7, Yogi</td>
<td>1.19</td>
<td>6.71</td>
<td>9.68</td>
<td>49.7</td>
<td>0.99</td>
<td>4.89</td>
<td>0.50</td>
<td>0.87</td>
<td>7.35</td>
<td>0.91</td>
<td>--</td>
<td>0.47</td>
<td>16.7</td>
</tr>
<tr>
<td>A-16, Wedge</td>
<td>2.30</td>
<td>4.58</td>
<td>10.24</td>
<td>48.6</td>
<td>1.00</td>
<td>3.29</td>
<td>0.41</td>
<td>0.96</td>
<td>8.14</td>
<td>0.95</td>
<td>--</td>
<td>0.65</td>
<td>18.9</td>
</tr>
<tr>
<td>A-17, Shark</td>
<td>2.03</td>
<td>3.50</td>
<td>10.03</td>
<td>55.2</td>
<td>0.98</td>
<td>1.88</td>
<td>0.38</td>
<td>1.14</td>
<td>8.80</td>
<td>0.65</td>
<td>0.05</td>
<td>0.49</td>
<td>14.8</td>
</tr>
<tr>
<td>A-18, Half Dome</td>
<td>1.78</td>
<td>3.91</td>
<td>10.94</td>
<td>51.8</td>
<td>0.97</td>
<td>3.11</td>
<td>0.37</td>
<td>1.10</td>
<td>6.62</td>
<td>0.82</td>
<td>--</td>
<td>0.52</td>
<td>18.1</td>
</tr>
<tr>
<td>Shergotty**</td>
<td>1.29</td>
<td>9.28</td>
<td>7.07</td>
<td>51.4</td>
<td>0.80</td>
<td>0.33</td>
<td>0.01</td>
<td>0.16</td>
<td>10.0</td>
<td>0.87</td>
<td>0.20</td>
<td>0.53</td>
<td>19.4</td>
</tr>
<tr>
<td>CI *** meteorite</td>
<td>0.68</td>
<td>15.5</td>
<td>1.55</td>
<td>22.8</td>
<td>0.23</td>
<td>13.5</td>
<td>0.067</td>
<td>0.062</td>
<td>1.26</td>
<td>0.073</td>
<td>0.39</td>
<td>0.23</td>
<td>23.5</td>
</tr>
<tr>
<td>Calc. soil free rock</td>
<td>2.46</td>
<td>1.51</td>
<td>11.0</td>
<td>57.0</td>
<td>0.95</td>
<td>0.30</td>
<td>0.32</td>
<td>1.36</td>
<td>8.09</td>
<td>0.69</td>
<td>--</td>
<td>0.55</td>
<td>15.7</td>
</tr>
<tr>
<td>Av. error (rel %)</td>
<td>40</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>25</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*Data from Clark et al. [18] normalized to 95.6 % to account for Na, P, Cr, and Mn.
**Data from Banin et al. [19]. ***Data from Palme et al. [20], CI meteorites contain in addition 3.5 % C and 17 % H<sub>2</sub>O.

One minor element, which could be analyzed with meaningful accuracy by APXS at the Pathfinder landing site on Mars, is Mn. Manganese strongly correlates with Fe on the Earth and the Moon. Hence, the Fe/Mn ratio is indicative of the samples' parent body. The mean Fe/Mn ratio of Pathfinder rock and soil samples of 39 ± 11 agrees well with the Fe/Mn ratio of shergottites of 39 ± 2. This qualifies as additional proof of the parent body of the Martian meteorites (Fig. 3).

A similar argument is derived from the Cr concentration in Martian meteorites and the Pathfinder rocks and soil, except that the accuracy of the measured Cr concentration is lower due to the lower absolute concentrations especially in the rocks. In the case of Mn, we have to admit that rocks from any parent body not depleted in Mn in its silicate phase like eucrites will plot close to the Mars data points, too.

Another minor element of importance is P. However, the P peak sits on the shoulder of the 50-times larger Si peak and, hence, P data have errors of ~30 %. The observed P concentrations in the Martian samples (~0.4 wt%) are very high when compared to terrestrial crustal concentrations of only ~0.1%, but these values are in line with high P concentrations predicted for the Martian mantle by Dreibus and Wanke [10] from data on Martian meteorites.
Fig. 4. **Top panels:** Linear regression lines for Si and Mg versus S of Pathfinder rocks and soils. The good fit of all the soil samples can be taken as an indication that the S is present in form of MgSO₄. **Lower Panels:** Linear regression lines for K and Cl. Although Cl in the soil is most likely to a large extent due to the presence of evaporates formed by HCl from volcanic exhalation, the Pathfinder rocks even seem to contain intrinsic Cl, too.

It became evident from the IMP-camera images at the Pathfinder Landing site that all rocks were covered by dust to a variable degree. Hence, the signal due to dust coatings had to be subtracted to derive rock compositions. The analyzed soils show sulphur concentrations of (2.72 ± 0.27) wt%, the measured S concentrations for the Pathfinder rocks, ranging from 0.75 to 1.96 wt%, by far exceed the concentrations accommodated in magmas or igneous rocks. When plotted against S most elements form linear arrays for the rocks (Fig. 4). Extrapolation of the regression lines to zero S content yields the approximate composition of soil-free unaltered rocks [Riedel et al., 23]. The rocks Shark and Barnacle Bill most closely approximate the composition of the soil-free rock. Minimal contamination by dust relative to other analyzed rocks is also evident from high-resolution IMP-camera images, which for these rocks exhibit higher red to blue reflectance ratios [McSween et al., 24].

As shergottites contain S in amounts between 0.13 to 0.28%, it might be more appropriate to assume for the Pathfinder rocks, which are more fractionated than shergottites, S concentrations of about 0.3 wt%. Hence, in a new approach Wänke et al. [25] extrapolated to 0.3% S rather to zero to find the composition of a soil-free rock. This procedure yields differences in the amounts for Si and Fe of <5% relative, but raises the MgO concentration from 0.6 to 1.5%. The good fit of all rock data to the regression lines (Fig. 4) indicates an almost identical composition of the 4 rocks analyzed. Sample A8 (Scooby Doo), although of rock-like appearance, is identified by its composition as cemented soil. The lower P content of Scooby Doo compared to all other analyzed samples seems to be real.
The composition of the rocks relative to the soil yielded a surprise. The Martian soil with its considerable Mg concentration was taken as evidence for a mafic crust of Mars. The Martian meteorites have mafic to ultramafic composition, too. In contrast, the rocks at Ares Vallis turned out to represent highly fractionated crustal material, rich in SiO₂ and K, but low in Mg. This holds regardless of the not-yet-solved nature of these rocks, i.e. igneous or sedimentary. The composition of the Pathfinder rocks, together with that of the soil in which they are embedded, as well as the average composition of the Martian meteorites, is illustrated in Fig. 5. The huge compositional difference between soil and rocks cannot be explained by diminution of these rocks, even considering weathering and interaction with volcanic gases SO₂ and HCI. Addition of material richer in Mg and Fe, but poorer in K and Cr as observed by the Martian meteorites, seems unavoidable.

Fig. 5. Histogram of some element concentrations in Mars Pathfinder rock (MPF) (soil-free), MPF soil and the mean composition of 14 Martian meteorites.

Taking the almost identical soil composition at the three landing sites as representative for the whole surface soil of Mars, the Mainz group [Brückner et al. 26] has shown that all elements fit into a two-component mixing diagram with the Pathfinder rocks on the high K-low Mg side and the Martian meteorites on the opposite side (Fig. 6). Hence, it was concluded that large geologic units of andesitic (Pathfinder rocks) as well as of basaltic (Martian meteorites) composition must exist on Mars and cover about equal areas. Bandfield et al. [27] recently confirmed this model with Thermal Emission Spectrometer (TES) data from Mars Global Surveyor (MGS). The soil composition and its relation to specific rock types, respectively to their weathering products, has been discussed by a number of authors [e.g. Bell et al., 28; McLennan, 29; McSween and Keil, 30]. McSween and Keil favor the view that basalts, chemically similar to basaltic shergottites, and evaporitic salts dominate the surface geology of Mars. We believe that admixture of the andesitic component is needed in addition, especially to account for K.

Fig. 6. Two component mixing diagram of the Martian soil. Iron is assumed to be present in the soil as Fe₂O₃, but in form of FeO for the end member rocks. MPF (Mars Pathfinder) Mean Rock is corrected for soil free.

In the two-component mixing diagram (Fig. 6), MgSO₄ and MgCl₂ have been subtracted. There is a reasonable fit for all elements except for Fe. The definite difference in the case of Fe between the concentration predicted from the two-component mixing diagram suggests that in addition to the above-mentioned major components, a Fe rich component has been admixed to the soil, too. Areas rich in hematite have recently been observed at the surface of Mars [Christensen et al., 31]. Hence, addition of Fe from hematite deposits seems plausible. Table 3 shows an attempt to account for the Martian soil by adding MgSO₄ and hematite (plus small amounts of ilmenite) to the basaltic and andesitic component. However, one should bear in mind that both MgSO₄ and hematite were originally supplied by weathering of basaltic material. Hence, the conclusion of an ~1:1 abundance of andesitic and basaltic material is still justified. Pathfinder rock compositions, extrapolated to 0.3 wt% S to account for the adhering soil, contain 57.0% SiO₂ and thus fall on the borderline between basaltic andesites and andesites, according to the chemical classification of volcanic rocks after Le Bas et al. [32].
Table 3. Composition of Martian soil: Comparison of measurements with the composition derived from a mixing model, which yields the following components from a least squares fit: Andesitic component (soil free Pathfinder rocks) = 54 %, basaltic component (mean composition of SNC meteorites) = 28 %, Mg-sulfate = 10 %, hematite = 7 %, and ilmenite = 0.9 %.

<table>
<thead>
<tr>
<th>Soil</th>
<th>Calculated</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgO</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>7.4</td>
<td>8.0</td>
</tr>
<tr>
<td>SiO₂</td>
<td>42.6</td>
<td>42.3</td>
</tr>
<tr>
<td>SO₃</td>
<td>6.9</td>
<td>6.8</td>
</tr>
<tr>
<td>K₂O</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>CaO</td>
<td>6.2</td>
<td>6.5</td>
</tr>
<tr>
<td>TiO₂</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>22.3</td>
<td>22.3</td>
</tr>
</tbody>
</table>

The next missions to the Martian surface will be Beagle 2 Lander of ESA Mars Express mission, and the two NASA Rover missions, all to be carried out in 2003. The two NASA MER (Mars Exploration Rover) missions will each bring two German instruments to the surface of Mars dedicated to mineralogical and chemical analyses of surface rocks and soils. Aside these and other instruments, like the Panoramic Camera and the Thermal Emission Spectrometer as well as a close-up camera, the rovers will also carry a rock abrasion tool with which it will be possible to remove the surface layers of the rocks so that the instruments can look on fresh rock material. The APX-spectrometer for the chemical analyses of soils and rocks will be considerably improved compared to that on Pathfinder. The energy resolution of the x-ray spectrometer will be about a factor of two higher for the low energy peaks. The sensitivity will be about a factor of three higher which in turn can reduce the measuring time. With the Mössbauer-spectrometer it will be possible for the first time to exactly determine the Fe²⁺ / Fe³⁺-ratio in the soil and rocks as well as to analyze iron-containing minerals.

3. REFERENCES


Exospheres of Earth-like Planets and Moons

G. Cremonese
INAF-Osservatorio Astronomico, vic.Osservatorio, 35122 Padova, Italy, cremonese@pd.astro.it

ABSTRACT

Surface-bounded exospheres have been detected at the Moon, Mercury and Europa and almost certainly exists about other objects. Historically, the first of these systems to be observed was the lunar exosphere, where He and Ar were detected by the Apollo spacecraft, but the most important discovery was the detection of Sodium and Potassium on 1988. The same discovery was made for Mercury on 1985. Soon after the discoveries several observers concentrate their efforts on the sodium emission lines, having a high cross section for resonant scattering making them very easy to detect, providing interesting data on the exosphere behaviour and its interaction with the surface and the interplanetary medium.

The sources that maintain these exospheres are of considerable interest, but direct information on sources is difficult to obtain. In the last few years several efforts have been concentrated on the interaction of the meteor showers and the lunar exosphere and surface, we will report the most important observations and results on this particular source.

We will review the current state of knowledge of the exospheres of Mercury and the Moon and discuss the important data that future space missions, as BepiColombo and Messenger, may provide.

Then we will briefly describe the exosphere of Europa, discovered on 1996, more than 20 years after the study of the Io sodium cloud, having the same main element, as sodium, but being characterized by some different sources and processes.

1. GENERAL DESCRIPTION OF LUNAR AND MERCURY EXOSPHERES

The atmospheres of Mercury and the Moon have been observed for the first time by some space missions as the Mariner 10 for the planet and the Apollo 14 and 17 for our satellite. In the first case only helium, hydrogen and oxygen were discovered using an airglow spectrometer on board the spacecraft [4], while sodium (5889.95 and 5895.92 Å) and potassium (7664.90 and 7698.96 Å) were discovered later on with groundbased instrumentation [23]. The first elements cannot be studied from the Earth, but only through a space mission orbiting around the planet, while the alkali elements, Na and K, can be observed using ground based instrumentation. This is the main reason why the atmospheric studies are performed on the alkali components in the visible spectral range.

A similar history occurred to the Moon when the Apollo spacecrafts revealed the presence of hydrogen, argon, helium and neon [11], while the sodium and potassium were discovered only in the 1988 using groundbased telescopes [24]. The atmospheres of Mercury and the Moon are collisionless to a good approximation. They differ from exospheres that exist above an atmosphere, such as those of the Earth and Venus, in that the exobase is a solid surface. The behaviour is therefore dominated by interactions of the gas atoms with the surface, rather than each other.

In the case of the Earth the exosphere is the outermost layer of the atmosphere and it goes from about 460 km high to about 1280 km. The lower boundary of the exosphere is called the critical level of escape, where atmospheric pressure is very low (the gas atoms are very widely spaced) and the temperature is very low. The lower boundary of the Mercury and Moon exospheres is in thermodynamic equilibrium, and the upper boundary can, to some approximation, be described as a loss-free diffuse boundary.

The atoms composing the Moon and Mercury exospheres must be continuously resupplied, as neither body can retain the atoms for more than few hours. The mechanisms proposed to explain the resupply include photon-stimulated desorption, electron-stimulated desorption, ion sputtering, meteorite vaporization, and chemical sputtering. Chemical sputtering is a term loosely applied to a process in which chemical reactions on the surface produce a product with enough kinetic energy to thermally desorb from the surface. Up to now there are few data and no general agreement about which processes dominate.

The exospheres contain particles in a Maxwell-Boltzmann kinetic energy distribution, and are barometric, even if in a surface bounded exosphere there is no reason to expect that the particles will be in thermodynamic equilibrium. Source processes such as photon-stimulated desorption and ion sputtering may produce a non-Maxwellian velocity distribution.

Sources and sinks should work in a similar ways on both bodies, even considering the smaller distance of Mercury from the Sun, and one of the most significant questions for the exospheres is the extent to which different source processes contribute. The main differences between the two bodies is the faint magnetic field present on Mercury that influences the exospheric atoms ionized and the possibility to recycle them hitting the surface, the higher temperature on the day side regions and the high orbital eccentricity with the corresponding strong variation of the solar flux.

These exospheres are generated by the vaporization of surface material, mainly present in the regolith, that is...
light and fluffy. Three things strike the regolith more or less continuously: a flow of micro-meteorites, a stream of plasma (the solar wind) and light.

The easily observed elements as sodium and potassium are the main targets of the exospheric observations using ground-based instruments working in the visible spectral range. Furthermore they are extremely efficient at scattering sunlight. They trace the behaviour of other sputtered products that are not directly observable and they provide important insights on the main processes working in the exosphere.

Laboratory simulations suggested that photon-stimulated desorption of sodium from surfaces, simulated by lunar silicates, can contribute substantially to the Moon's exosphere [31]. The sodium source rate needed to maintain the observed column density of \( \sim 10^9 \) atoms cm\(^{-2} \) above the lunar surface is variously estimated [13,20] to be in the range \( \sim 5 \times 10^9 \) to \( \sim 2 \times 10^{10} \) atoms cm\(^{-2} \) s\(^{-1} \). Reference [31] finds \( 4 \times 10^8 \) atoms cm\(^{-2} \) s\(^{-1} \) considering only the photon-stimulated desorption, that is in good agreement with the upper limit of sodium source rates indicated above.

In the case of Mercury thermal desorption is believed to contribute to the sodium exosphere because of the high day temperature (700 K at the sub-solar point), but photon-stimulated desorption is also likely to be important, especially at different latitudes.

In the last few years some interesting observations have been performed during a meteor shower impacting the lunar surface, as the Leonids, showing an increase in the sodium emission intensities and temperatures. A similar process is evident also during the total lunar eclipse that should shield the Moon from solar photons and ions. In Tab.1 we report the atmospheric densities of the two bodies with the Earth's mesopause and the Solar System for comparison. The sodium column densities for Moon and Mercury exospheres are in the range of \( 0.3 \times 10^9 \) and \( 1.5 \times 10^{11} \) atoms cm\(^{-2} \) respectively, while for potassium they are \( 0.3 \times 10^8 \) and \( 0.1 \times 10^8 \) atoms cm\(^{-2} \) respectively.

<table>
<thead>
<tr>
<th>Substance</th>
<th>Solar Si = 10(^6)</th>
<th>Earth (90km) cm(^3)</th>
<th>Mercury cm(^3)</th>
<th>Moon cm(^3)</th>
<th>Lifetime (1 AU) ( 10^4 ) sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>--</td>
<td>1.0 \times 10(^9)</td>
<td>200</td>
<td>&lt;17</td>
<td>2000</td>
</tr>
<tr>
<td>He</td>
<td>--</td>
<td>--</td>
<td>600</td>
<td>2000-40000</td>
<td>1400</td>
</tr>
<tr>
<td>O</td>
<td>2.1 \times 10(^7)</td>
<td>n.a.</td>
<td>( &lt;4\times10^4 )</td>
<td>&lt;500</td>
<td>250</td>
</tr>
<tr>
<td>Na</td>
<td>6.0 \times 10(^7)</td>
<td>1000</td>
<td>20000</td>
<td>70</td>
<td>16.9</td>
</tr>
<tr>
<td>K</td>
<td>4.2 \times 10(^7)</td>
<td>67</td>
<td>500</td>
<td>16</td>
<td>3.7</td>
</tr>
<tr>
<td>Ar</td>
<td>--</td>
<td>--</td>
<td>( &lt;3\times10^4 )</td>
<td>4 \times 10(^4)</td>
<td>200</td>
</tr>
</tbody>
</table>

A further item to be considered are some ground-based images of Mercury showing that the sodium is usually not uniform over the surface, often concentrated in regions at high or mid latitudes [25]. These authors suggested that sputtering by magnetospheric particles was the origin for the sodium. A problem with this proposal is that the magnetic field of Mercury is strong enough that it is believed to shield the surface from the solar particles much of the time, although particle precipitation at the magnetospheric cusps could deposit particles to the surface at high latitudes. Moreover some observations collected during a long period of 1.5 years, from June 1986 to January 1988, revealed an enhanced abundance of K above the longitude range containing the Caloris Basin [29]. This enhancement was consistent with an increased source of K from a well-fractured crust and regolith associated with this large impact basin. Then few bright spots observed in radar data have been associated with possible Na enhancement. Furthermore a sodium increase by a factor of about 3 has been observed in November 1997 [26], but in this case there are no obviously outstanding geologic features at the observed longitudes. Some explanations have been given, related to the solar activity and to some regions possibly with an higher abundance of Na and K, but they are hypothesis still to be well verified and much more observations are needed.

The last element discovered in the Mercury exosphere is the Calcium [3] observed at 4226.74 Å, showing a very low column density of \( 0.4 \times 10^9 \) atoms cm\(^3\) and an apparently very high temperature of about 12000 K. This observation is not yet confirmed, but the atmospheric calcium may arise from surface sputtering by ions, which enter the Mercury's auroral zone, explaining the decreasing intensity from the poles to lower latitudes.

2. GENERAL DESCRIPTION OF EUROPA'S EXOSPHERE

Europa, the second large satellite out from Jupiter, is roughly the size of Earth's Moon, and it has long been thought to be a dormant icy body. Recently a specific observation by the Hubble Space Telescope reported the detection of atomic oxygen emission, that has been interpreted as being produced by the simultaneous dissociation and excitation of atmospheric O2 [10]. This detection has been made in the UV spectral range (130.4 and 135.6 nm).

Even in this case we can define it as a exosphere with the exobase represented by the Europa surface, but the processes working on it may be very different with respect to the Moon and Mercury as it lies deep within the Jupiter's magnetosphere and very close to the Io's plasma torus. Europa is continuously bombarded by energetic ions which modify the surface ices.
One year after the oxygen detection an extended sodium atmosphere, at least 25 times Europa’s radius, has been discovered suggesting a strong correlation with the Io’s sodium cloud [6]. The discovery paper suggested that the sodium atoms are originally released by Io’s volcanoes, after which they are ionized in the strong Jupiter’s magnetosphere and implanted into Europa’s surface ice. Recently Brown [5] mentioned a factor-of-two calibration error in that paper pointing at higher sodium abundance in the exosphere and making difficult to explain this new value assuming Io as major source. In other words most of the sodium atoms observed should be intrinsic to the surface of Europa suggesting the presence of salty water ice. This hypothesis could be further supported by the discovery of single ionized chlorine in the Io plasma torus at 8579 Å [15] and salt may be, as a solid or gas, an important source of both Na and Cl.

As for any other body where the sodium has been detected even the potassium has been observed in the Europa’s exosphere by [5].

Taking into account the Europa environment, as the Jupiter’s magnetosphere and the Io’s plasma torus, the sputtering process is the likely source of Na atoms observed in the exosphere, but the ultimate origin of the sodium is not fully understood [16]. A minor source can be represented by the micrometeor impacts that can contribute to the alkali abundances, but an exhaustive model is still missing and the composition of the Europa surface is not known as for Mercury.

3. RATIO OF SODIUM TO POTASSIUM IN THE EXOSPHERES

Sodium and potassium have to be mainly derived from surface materials, so that we might expect the ratio Na/K to reflect the ratio of these elements in the surface crust. This expectation is approximately borne out for the Moon, where the ratio Na/K in the exosphere averages to be about 6 [13], not too far from the ratio in lunar rocks 2 to 7 [18]. However, the ratio in the Europa and Io exospheres was found to be slightly higher, and much higher in the Mercury case as reported in Tab.2.

<table>
<thead>
<tr>
<th>Object</th>
<th>Na/K</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>80-190</td>
<td>Potter and Morgan (1997)</td>
</tr>
<tr>
<td>Mercury</td>
<td>98</td>
<td>Potter et al. (2001)</td>
</tr>
<tr>
<td>Mercury</td>
<td>400</td>
<td>Hunten and Sprague (1997)</td>
</tr>
<tr>
<td>Lunar exosphere</td>
<td>6</td>
<td>Hunten and Sprague (1997)</td>
</tr>
<tr>
<td>Earth atmosphere</td>
<td>20-150</td>
<td>Gault and Rundle (1969)</td>
</tr>
<tr>
<td>Io</td>
<td>10</td>
<td>Brown (2001)</td>
</tr>
<tr>
<td>Europa</td>
<td>25</td>
<td>Brown (2001)</td>
</tr>
<tr>
<td>Cosmic abundance</td>
<td>20</td>
<td>Allen (1991)</td>
</tr>
<tr>
<td>Meteorites</td>
<td>13</td>
<td>Allen (1991)</td>
</tr>
<tr>
<td>Lunar rocks</td>
<td>2-7</td>
<td>Allen (1991)</td>
</tr>
<tr>
<td>Earth crust</td>
<td>2</td>
<td>Lide (1996)</td>
</tr>
</tbody>
</table>

The analysis of this ratio allows to study the production efficiencies and the loss rates of sodium and potassium in the exospheres and the possible relation with the composition of the surface rocks. For instance the value for the Mercury exosphere is higher than 100 and it seems unlikely to be in agreement with the surface rocks, it suggests that either production efficiencies or loss processes for the two elements are not equivalent. Assuming that the source processes should not be different on the Moon and Mercury by an order of magnitude the loss processes are the cause of the difference between the two. Reference [27] collected, over a decade, sodium and potassium data of the Mercury exosphere, at irregular intervals, and plotting the corresponding column densities they found a mean ratio of 98+20.

They suggest that the scatter in the ratio is due to sodium and potassium densities not closely correlated to one another, pointing at geographic differences of the ratio in the surface rocks and confirming some ground-based observations.

In the case of Europa the calculated ratio Na/K is higher than for Io ruling out Io as the source for the Europa trace species. Moreover, the analysis of the Europa Na/K value might be useful for trace the elements intrinsic to the ices and the elements deposited in liquid water resurfacing events.

More observations are needed to better define the processes working on sodium and potassium and the relations with the surface rocks. Future space missions orbiting around these bodies will help us very much
having the chance to get the surface composition at the same time of the exosphere observation.

4. MICROMETEOR IMPACTS

The micrometeor impacts have been taken into account from the beginning as a source of the exosphere of these bodies, but it has often been assumed negligible. Other sources have been considered as main contributors to the exosphere assuming that the alkali atoms are already present in the regolith.

In the case of the micrometeor impacts the exospheric components can be present in either the projectile or the target regolith, or both. Only in the last few years a major attention has been payed to this source mainly because some lunar exosphere observations have been associated to meteor streams.

On October 1990 [12] claimed to have observed the lunar exosphere during a possible meteor shower not yet identified. Their spectra showed a 60% increase of the sodium abundance at the South pole, while at the equator no substantial change has been measured, over a 3-day period. In these observations appeared that the strength of the dominant source doubled in the South polar region over a period of 2 days. If meteoritic impact was this dominant source for alkalis, the most likely explanation of an increase was a meteor shower with a radiant near the South pole. Later on [8] reported, 4 days before the Leonids peak in 1995, a significant sodium brightness enhancement translated into unprecedented abundance and scale height increases of the lunar exosphere. The statistical correlation of this event with the Leonids has been considered not very strong, but further observations obtained during the 1997 Leonids maximum [14] showing a small enhancement suggested a real correlation with the meteor shower. Following these results [28], using a bare CCD all-sky imaging system, revealed an extended region of neutral sodium in the direction of the anti-solar/lunar points.

These observations were performed during the Leonids passage, on 18-20 November 1998, and assuming the Moon as the most likely source these was the first detection of the lunar sodium tail out to a distance of hundreds of lunar radii. The important result was the greater sodium brightness observed on 19 November 1998 and attributed to the Leonid meteor shower having the peak on the Moon two days before, just the time to travel at the observed point for the neutral atoms.

Furthermore optical detection of meteoroidal impacts on the Moon have been observed during the Leonids 1999 passage [22]. Five bright and very brief flashes (duration <0.02 seconds) have been detected and they can be regarded as measurements of direct impacts, valuable in characterizing the current population of meteoroids in the vicinity of the Earth. At the same time the flashes might be correlated with the sodium emission enhancements observed in the lunar exosphere during meteor showers, even if there is no correlation, up to now, between the flashes and the sodium atoms.

Some calculations suggested that the degree of ionization increases with the meteoroid velocity and the total number of sodium ions depends from the mass flux of the shower [7]. The exosphere enhancements observed on the Moon are related to the parameters of the meteor showers. The Leonids in the past years had an higher mass flux and the geocentric velocity is the highest for a shower, it could explain the observations above mentioned if we also assume a higher fraction of sodium present in the meteoroids.

Other meteor showers have been suggested as candidates to stimulate exosphere enhancements as the Quadrantids having a much higher mass flux, but with a geocentric velocity of 41 km/s.

High resolution spectroscopy was performed during the Quadrantids on January 2000, but no emission enhancements were measured [30]. A further parameter to take into account is the fraction of sodium present in the meteoroids, that can be very different between the meteor showers. Fig.1 reports the sodium emissions intensity as a function of the impact velocity and the fraction of sodium assumed [7].

![Graph](image)

Fig.1, sodium emissions intensity as a function of the impact velocity of the meteor showers. The curve has been plotted for three different values of the fraction of sodium present in the meteoroid f.

Similar studies should be applied even to Mercury being interested from a much higher flux of micrometeoroids, up to now there are only some general models.

5. FUTURE SPACE MISSIONS

NASA approved and is going to build up the payload for the Discovery mission Messenger, designed to orbit around Mercury along a very eccentric orbit. Messenger will be launched on 2005 and will include a wide spectral range spectrograph, from UV to NIR,
that will observe also the exosphere. The cameras will work in the visible range, but the filters mounted are optimized for geologic and mineralogic features and they will not contribute substantially to the exospheric studies.

On October 2000 also ESA approved a mission to Mercury, the cornerstone n.s. BepiColombo that will include a planetary orbiter along a quasi circular orbit, a magnetospheric orbiter along a very eccentric orbit, similar to the Messenger one, and it is still a subject of debate a lander. The magnetospheric orbiter will be realized by the Japanese agency ISAS and will include some instruments interesting to study the exosphere and its interaction with the magnetic field and the solar wind.

The planetary orbiter will include an UV spectrograph devoted to observe the exosphere, to look for new elements, as for instance OH, as water dissociation product, magnesium, calcium and so on, and to monitor a sodium emission. Then, there will be a Wide Angle Camera that will provide a global mapping of the Mercury surface, but it should be able to get data on the exosphere, according to the scientific objectives identified for the payload by the ESA committees. It would be extremely interesting to get wide field images of the exosphere in the sodium and potassium light at the same time of the surface observations using the same instrument. It would help to study the possible relations between some surface regions and the exospheric emissions, as suggested by ground-based observations.

A further optional instrument has been proposed to study the exosphere, the Neutral Particle Analyzer called SERENA [21]. It will provide neutral particles measurements for getting a comprehensive picture of the solar wind-planet interaction and it will allow comparative investigations of evolution and dynamics of the planet magnetosphere.

Future space missions devoted to study the Moon don't appear to have instruments observing the exosphere. A big step forward on the lunar exosphere may come from observations performed by a small telescope orbiting around the Earth monitoring the sodium and potassium components [2]. In so doing it would be possible to provide continuity of observations required in order to relate observed morphologies to processes of Na and K ejection from the surfaces.

This continuity is impossible by using ground based telescopes, due to the extremely high level of scattered light at gibbous and full Moon, the sky background and low-elevation-angle scattering at new moon, the need for a continuously clear and photometric sky and the Na diffuse emission from terrestrial sources.

6. REFERENCES

Session 5
Impact Interactions with Planets and Moons
Chairs: L. Colangeli & G. Schwehm
EXOGENOUS MATERIAL DELIVERY TO EARTH-LIKE PLANETS AND MOONS

O. Botta (1), D. P. Glavin (2), E. Pierazzo (3), P. Ehrenfreund (4) and J. L. Bada (5)

(1) Soft Matter/Astrobiology Group, Leiden Institute of Chemistry, P. O. Box 9502, 2300 RA Leiden, The Netherlands. Email: o.botta@chem.leidenuniv.nl.
(2) Max Planck Institute for Chemistry, Cosmochemistry, P. O. Box 3060, 55020 Mainz, Germany. Email: glavin@mpch-mainz.mpg.de.
(3) Planetary Science Institute, 620 N. Sixth Ave, Tucson, AZ 85716, USA. Email: betty@psi.edu.
(4) Leiden Observatory, P O Box 9513, 2300 RA Leiden, The Netherlands. Email: pascale@strw.leidenuniv.nl.
(5) Scripps Institution of Oceanography, University of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0212, USA. Email: j.bada@ucsd.edu.

ABSTRACT

Asteroids, comets, meteorites and IDPS, are generally believed to have contributed to the volatile inventory of the terrestrial planets. About 1 weight-% of carbonaceous meteorites, which are thought to be fragments of asteroids, can be extracted with solvents in the form of separable compounds. Some of these compound might have been important starting materials for the chemical evolution of the early Earth (and potentially Mars). Interplanetary dust particles (IDPs), sub-millimeter sized particles that contain up to 40% organic carbon, are associated with comets and are thought to have contributed a large fraction to the inventory of volatiles found on the terrestrial planets. We summarize the organic composition of carbonaceous chondrites, and discuss the role that IDPs, carbonaceous chondrites, asteroids and comets may have played in the delivery of these organic compounds to the early Earth and other solar system bodies.

1. INTRODUCTION

The origin of life on Earth, and possibly on other solid bodies of the solar system, such as Mars and Europa, would have required the presence of liquid water and a continuous supply of pre-biotic organic compounds from which the first primitive forms of life could emerge. Various organic compounds including amino acids, the building blocks of proteins and enzymes, as well as purines and pyrimidines, essential components of replicating systems in all known organisms, have been synthesized abiotically in laboratory experiments from simple precursors including aldehydes, hydrogen cyanide, ammonia and water using electric spark discharge as an energy source (these experiments are commonly known as Miller-Urey-Experiments [1,2,3,4]). It is possible that if the atmosphere of the early Earth was reducing (i.e., based on CH4 and NH3), lightning and/or UV radiation could have triggered similar abiotic chemical reactions producing a large inventory of organic compounds thought to be required for the origin of life. However, it is now believed that the early Earth’s atmosphere was non-reducing, composed mainly of carbon dioxide [5,6]. Under a carbon dioxide atmosphere, spark discharge experiments yield only very low levels of amino acids [7,8], which makes it difficult to explain how a significant “home-grown” synthesis of amino acids could have taken place on the early Earth.

Abiotic synthesis in hydrothermal vents, sometimes referred to as “black smokers”, provide another possible terrestrial source of organic material [9]. These volcanically active regions on the ocean floor spew out highly mineralized water at temperatures of up to 350°C into the cold surrounding water. The main problem with this possible source, however, is the fact that complex organic molecules are not stable under these extremely high temperatures and would be decomposed shortly after their formation [10,11]. Alternatively, the exogenous delivery of organic matter by asteroids, comets and micrometeorites could have played a significant role in seeding the early Earth, and possibly Mars, with the prebiotic compounds considered to be necessary for the origin of life [12,13] (Fig. 1).

In this paper, we address the organic composition of carbonaceous chondrites, in particular of those complex organic molecules found in the water-soluble fraction that are important for life as we know it today. We will also discuss the role that IDPs, carbonaceous chondrites, asteroids and comets may have played in the delivery of these organic compounds to the early Earth and other solar system bodies.


© European Space Agency • Provided by the NASA Astrophysics Data System
2. ORGANIC COMPOUNDS IN CARBONACEOUS METEORITES

2.1. Overview

Carbonaceous chondrites, a rare class of stony meteorites, are considered to be the most primitive objects in the solar system in terms of their elemental composition, yet they feature a high abundance of organic carbon, more than 3 weight-% in some cases. The most extensively analyzed meteorites for organic compounds include the CMs Murchison (fell in 1969 in Victoria, Australia) and Murray (1950, Kentucky, USA) and the CI Orgueil (1864, France). The carbon phase in these meteorites is dominated by an insoluble fraction, which contains more than 80 % of the total mass of this phase, with the rest being soluble compounds. Polycyclic aromatic hydrocarbons (PAHs) make up the majority (up to 80 %) of this soluble fraction, followed by the carboxylic acids and fullerenes, which are about an order of magnitude less abundant [14].

2.2. Nucleobases

Purines and pyrimidines, which are one- or two-ring aromatic compounds containing several nitrogen atoms in their rings, play a major role in terrestrial biochemistry. They are central components of DNA and RNA, molecules that are used in the storage, transcription and translation of genetic information in all terrestrial organisms. Unlike amino acids, nucleobases do not exhibit molecular chirality (see below), which makes it difficult to distinguish between abiotic and biotic origins of these compounds. The concentrations of the pyrimidine uracil and the purines adenine, guanine, xanthine and hypoxanthine have been determined in Murchison, Murray and Orgueil in total abundances of about 1.3 parts per million (ppm) [15,16]. Although the isotopic abundances for carbon and nitrogen, which provide unambiguous evidence for the extraterrestrial origin of organic compounds in meteorites, have not yet been measured for the N-heterocyclic compounds, based on the very low contamination levels for amino acids in these meteorites [17,18], a low terrestrial contamination for the nucleobases can be inferred.

2.3. Amino Acids

Several amino acids which are extremely rare on Earth, such as α-aminoisobutyric acid (AIB) and isovaline, have previously been detected in the Murchison meteorite [19,20]. Altogether more than 70 different amino acids have been identified in this meteorite, most of which do not occur naturally on the Earth. The total abundance of amino acids in this meteorite is approximately 10 parts per million (ppm), which is ≥ 0.1 % of the total soluble organic carbon in this carbonaceous chondrite [17]. In all organisms on Earth, only the L-enantiomers (left-handed forms) of chiral amino acids are incorporated into proteins and enzymes. In contrast, the abiological synthesis of chiral amino acids always yields a 1:1 mixture of the D- and L-enantiomers (a racemic mixture). Therefore, the molecular architecture of these compounds provides a powerful tool to discriminate between biological and non-biological origins of amino acids in meteorites. Until recently, all chiral amino acids (e.g. alanine or isovaline) in meteorite extracts were found to be present as racemic mixtures, which indicates an abiotic origin and therefore the presence of indigenous extraterrestrial amino acids. In 1997, Cronin and Pizzarello [21] analyzed Murchison hot-water extracts and found enantiomeric excesses (EE's) of the L-enantiomer of the two diastereoisomers of 2-amino-2,3-dimethylpentanoic acid (DL-α-methylisoleucine and DL-α-methylalloisoleucine) as well as isovaline, which are non-biological amino acids that, due to their molecular architecture, are not prone to racemization (the conversion of an enantiomerically pure compounds into a racemic mixture). Similar, but not identical EE's were found in Murray [22]. Although it has been suggested that polarized UV radiation could produce the enantiomeric excesses observed in meteorites [23], no laboratory experiments have yet been able to reproduce these findings, assuming the amino acids were formed inside the meteorite and not in the gas phase in the interstellar medium (where they are not stable against the harsh UV-radiation [24]). In addition, it is still not clear at all if chirality was a prerequisite or a consequence of life.
3. ATMOSPHERIC ENTRY SURVIVAL IN IDP’S

3.1. Introduction

Although the present day contribution of organic compounds by carbonaceous chondrites (centimeter sized or larger) is estimated to be insignificant, only a few kilograms of organic carbon per year, interplanetary dust particles (IDPs), tiny sub-millimeter sized meteorites that contain up to 40\% organic carbon [25], supply the bulk of extraterrestrial debris accreted to the Earth each year [12]. The present day flux of carbonaceous matter by IDPs to the Earth is estimated to be $3 \times 10^5$ kg/yr, but this flux may have been as high as $5 \times 10^7$ kg/yr during the heavy bombardment period before 3.8 Gyr ago [13]. A present-day IDP flux has also been estimated for Mars [26]. Scaled to the carbon content of the carriers, these fluxes are approximately three orders of magnitude higher than the kilometer-sized objects such as asteroids and comets, and five orders of magnitude higher than for centimeter-to-meter-sized meteorites in the corresponding periods [13]. Thus, IDPs, which are called micrometeorites if they reach the Earth’s surface, could have been a major exogenous source of the ingredients of the primordial soup on the early Earth and potentially even on Mars.

One problem associated with the delivery of organic compounds by micrometeorites is that these grains, depending on their mass and size, can suffer full-depth heating to very high temperatures during atmospheric entry deceleration. Several atmospheric entry models and analyses of Antarctic micrometeorites (AMMs) predict that depending on the particle size, density, entry velocity and angle, micrometeorites can be heated to temperatures of $\sim 200^\circ C$ to $1200^\circ C$ for 5-15 seconds during atmospheric entry [27,28,29], and that less than 1\% of the total mass of infalling dust transits the atmosphere at temperatures lower than $600^\circ C$ [30]. In fact, an analysis of a melted IDP (L2005B22) collected in the lower stratosphere, indicated that this grain had been exposed to a temperature of $\sim 1200^\circ C$ in a flash-heating episode [31].

Sublimation has been proposed as a mechanism by which volatile organic compounds could escape and survive atmospheric entry heating by vaporizing off the surface of IDPs and even larger meteorites before they are melted and destroyed [32]. Sublimed organic compounds could re-condense as ice crystals in the cold upper atmosphere (Fig. 2), descend through the mesosphere by atmospheric transport and gravitational settling, coagulate with stratospheric aerosols, and then be softly deposited on the Earth’s surface. Single-particle analysis of stratospheric aerosols collected at 5 to 19 km altitude confirms that about half of the particles contain ablated material from meteorites [33].

Amino acids have relatively high vapor pressures under reduced pressure and at temperatures above 150$^\circ C$; therefore, it is possible that these compounds could sublime from IDP grains during atmospheric entry.

![Hot micrometeorite surface (200°C to 1200°C)](image-url)

Fig. 2. Sketch of a micrometeorite entering the Earth’s atmosphere.

3.2. Sublimation Experiments and Results

The available amount of micrometeorites is far too low to be used in the sublimation experiments. However, the analysis of AMMs has shown that they are chemically and petrologically similar to the CM and CR type carbonaceous chondrites [34], which makes it possible to use crushed samples of these meteorites as analogues for micrometeorites (for these purposes). In order to test the sublimation of amino acids and nucleobases from actual meteorite grains, the samples were sealed inside a glass sublimation apparatus (SA, Fig. 3) at a pressure equivalent to atmospheric pressure at 80 to 100km altitude (region of maximum frictional heating) and then the entire apparatus was placed in a furnace set at 1100$^\circ C$. A thermocouple was placed near the sample to estimate the actual temperature the sample reached during the experiment [18]. After heating, the cold finger was rinsed with water and analyzed for organic compounds. The amino acids were analyzed using OPA/NAC derivatization [35] combined with High Performance Liquid Chromatography (HPLC) with fluorescence detection. Nucleobases were analyzed with HPLC and UV absorption detection. Also, the residual meteorite grains were analyzed to determine if any of these compounds had survived the heating [36].
In the Murchison samples that had been heated for 5 seconds to a temperature of \(-220^\circ\text{C}\), 44% of the total amino acids originally present in the sample were recovered from the meteorite grains intact, even though most of these amino acids did not sublime from the meteorite in this experiment (data not shown). In contrast to the 5 second heating, only 3% of the total amino acids were recovered from the Murchison sample that had been heated for 30 seconds to a temperature of 550°C. The pyrolysis of most of the amino acids at 550°C is consistent with the solid phase thermal decomposition temperature (200°C-600°C) measured for several \(\alpha\)-amino acids [37]. No amino acids survived in the meteorite grains that had been melted after laser heating at a temperature of \(-1200^\circ\text{C}\) [18].

Even though \(-25\%\) of glycine in the meteorite sample sublimed onto the cold finger during the experiments, most of the amino acids in Murchison were not sublimed directly from the meteorite at elevated temperatures (Fig. 4).

This result is very surprising given that several of the amino acids detected in Murchison, including alanine, AIB and isovaline, were found to readily sublime from a pure standard mixture under the same heating conditions as Murchison with recoveries ranging from 97 to 100% (Fig. 4). No nucleobases could be detected on the coldfinger after the Murchison meteorite grains were heated to 450°C for 5 minutes, which indicates that these compounds did not sublime from the meteorite during the experiment. However, under the same conditions adenine, uracil and xanthine were sublimed from a dried formic acid extract of Murchison with recoveries between 15 and 35% (Fig. 5). These results indicate that the sublimation of these compounds directly from micrometeorites during atmospheric entry heating will probably not occur, and that some insoluble, but not acid resistant material inside the meteorite probably prevents the sublimation [18].

Fig. 5. Sublimation recoveries of adenine (A), guanine (G), uracil (U) and xanthine (X) from a pure standard mixture, a powdered sample of the Murchison meteorite, and a formic acid extract of Murchison after heating under reduced (500 mTorr) in the SA at 450°C for 5 minutes. Taken from Ref. 18.

Based on these results, it is apparent that most of the amino acids, with the exception of glycine, do not sublime and are almost completely decomposed at a temperature of 550°C. None of the nucleobases present in the meteorite grains sublimed at elevated temperatures up to 450°C. The sublimation of glycine from Murchison meteorite powder at temperatures above 150°C provides the only evidence that amino acids could sublime from micrometeorites and survive atmospheric entry heating. Although it is not fully understood why most of the amino acids, purines and pyrimidines do not sublime from Murchison, experimental evidence suggests that divalent cations such as \(\text{Ca}^{2+}\) and \(\text{Mg}^{2+}\), and/or the presence of kerogen-type organic polymers in Murchison may inhibit the

Fig. 4. Sublimation recoveries of amino acids from the Murchison meteorite and a pure mixture of amino acids (standard) after heating in the SA in a 1100°C furnace for 5 sec. and for 30 sec. Taken from Ref. 18.
sublimation of these compounds [18]. Because micrometeorite samples were not tested directly, it is still not known whether or not the sublimation of these compounds from micrometeorites would behave like Murchison.

4. COMETARY DELIVERY OF AMINO ACIDS

4.1. Impacts on the Earth

Earlier theoretical studies about impact processes on the Earth suggested that most organic compounds contained in the impactor would be destroyed by the high temperatures produced in these collisions [38]. These conclusions have recently been challenged by field evidence. Extraterrestrial amino acids were detected in the sedimentary layers that were formed when the bolide that formed the Chicxulub crater impacted the Earth 65 Myr ago [39]. More recently, extraterrestrial helium was found trapped inside fullerenes delivered by another large impactor, which created the Sudbury impact structure in Canada almost 2 Gyr ago [40]. In addition, new laboratory experiments that simulate the physical conditions in shock events indicate not only survival of amino acids under these conditions, but also the formation of dimers [41].

Experimental studies of impact events in the laboratory can be extended to planetary scales of hundreds of kilometers by computer simulations. New high-resolution hydrocode modeling simulations of asteroid and comet impacts, which trace the impactor’s thermodynamic evolution, coupled with recent experimental data for amino acid pyrolysis in the solid phase [37] suggest that amino acids would survive the shock heating of large (kilometer-radius) cometary impacts at the percent level [42]. These calculations investigated the effects of impact velocities, projectile size and material, as well as impact angles, on amino acid survival and delivery to the Earth’s surface. In particular, the effect of very low angle impacts has been shown to be enormously important, increasing the survival of certain amino acids by an order of magnitude [42,43]. The steady-state concentrations of some of the exogenously delivered amino acids in the Earth’s early ocean could have equaled or substantially exceeded concentrations obtained from Miller-Urey-type experiments in realistic non-reducing atmospheres. However, these resulting concentrations would still be extremely low, too low for any polymerization reactions to occur without some sort of catalyst. The occasional grazing cometary impacts – delivering large boosts to amino acid levels – offers one possible solution to the concentration problem. There may have been times in the Earth’s early history, right after a grazing comet impact, in which the oceanic concentrations of certain amino acids would have exceeded the steady state value by up to ten times.

4.2. Commentary Delivery to Europa

Europa, one of the larger satellites of Jupiter, shows strong evidence for the presence of a liquid water ocean under its icy crust [44]. Since Europa’s formation conditions in the circum-jovian nebula are poorly known, its inventory of biogenic elements, particularly of carbon, in this ocean or in the ice could be less than chondritic, which some models assume for Europa’s bulk composition [45].

More than 90% of the craters on the Galilean satellites are estimated to be due to Jupiter-family comets. The distribution of the cumulative impact velocity suggests that on Europa the comets’ median impact velocity is around 26.5 km/s, with 10% of them striking at velocities below ~16 km/s [46].

Again, hydrocode modeling simulations were applied to investigate the exogenous contribution to Europa’s biogenic inventory [47]. In this case, a range of comet densities (1.1 to 0.6 g/cm³, corresponding to porosities from 0 to 45%) and impact velocities (16 to 30.5 km/s) were investigated. The results indicate that at typical European impact velocities most impactor material reaches escape velocity, and is lost to space. In particular, the very favorable effect of a dramatic increase in amino acid survival at very low impact angles on Europa is counteracted by the low escape velocity. While for a porous comet, a significant fraction (~25%) of the impactor material is retained only at the lowest impact velocity modeled (16 km/s), significant amounts are retained in all cases for a high-density comet (Fig. 6).

Estimating an integrated incident mass on Europa of 8.2 x 10¹⁷ g over the age of the solar system [42] and using the results of the impact simulations, total masses of accreted cometary material can be calculated for the different comet densities.
Fig. 6. Fraction of projectile material that reaches Europa's escape velocity in the impact simulations as a function of projectile density (x-axis) and impact velocity (various lines/symbols) for a comet 1 km in diameter. Taken from Ref. 47.

By using the appropriate cometary elemental abundances, these calculations suggest that the masses of biogenic elements accreted from cometary impacts on Europa range from 0.9 to 10 Gt for C (1 Gt = 10^{15} g), 0.2 to 3 Gt for N, 0.2 to 2 Gt for S and 0.02 to 0.3 Gt for P, which corresponds to between 0.003 and 0.3 % of the estimated total biomass in Earth's oceans today [42].

The situation looks different if one considers the delivery of complex organic molecules to Europa's surface. The same approach used for the delivery of amino acids to the early Earth [42], that is a combination of hydrocode modeling and thermodynamic parameters for amino acids pyrolysis, has been applied to Europa to determine the impact survival of these compounds [47]. Although the survival fraction for some of the amino acids at all impact velocities is at the percent level, only a small component of these organics will actually be deposited on the surface because a large fraction of the impactor (see Fig. 6) is lost to space. Fig. 7 shows the fraction of selected protein-building amino acids that would be delivered intact to Europa for a comet 1 km in diameter at various impact velocities. Integrated over the age of the solar system, it is found that the overall exogenous contribution of organic compounds is several orders of magnitude lower than that estimated on Earth's oceans [42]. In addition, it is not easy to estimate the fraction of that material that would be ultimately cycled into the ocean [48].

![Graph showing fraction of amino acids delivered intact to Europa as a function of impact velocity for a comet 1 km in diameter.](image)

Fig. 7. Surviving fraction of selected amino acids (relative to the initial concentration on the comet) that would be delivered intact to the surface of Europa as a function of impact velocity for a comet 1 km in diameter. For glycine, only two data points were obtained, and the dashed line indicates an appropriate extrapolation. Data taken from Ref. 47.

CONCLUSIONS

Exogenous delivery by asteroids, comets and IDPs appears to have been an important mechanism for seeding the terrestrial planets with complex organic compounds. These compounds probably triggered the increase in molecular complexity that may have been necessary for the origin of life to occur. The analyses of carbonaceous chondrites, which are related to asteroids and possibly comets, have revealed that these rocks contain a wide variety of organic compounds, including amino acids, nucleobases, PAHs, fullerenes as well as sugar-related compounds. [49]

Although these compounds easily survive atmospheric entry and impact in centimeter- to meter-sized objects, their survival in micrometeorites and in massive impact events is much less understood. Atmospheric entry models and analyses of Antarctic micrometeorites indicate that IDPs are heated to temperatures of up to 1200°C during atmospheric entry. This is too high for any organic compound to survive. The possibility that amino acids and other organic compounds such as nucleobases would be able to survive atmospheric entry by subliming off the hot surface of a micrometeorite grain have been tested through experimental studies. Although these organic compounds have been shown to sublime in high yields from pure standard mixtures, they do not sublime from actual meteorite grains, which was used as an analogue for micrometeorites. Only glycine was found to sublime in low yields (~25%) after the sample was heated to temperatures above ~220°C. All other amino acids did not sublime, and almost all were decomposed at temperatures of ~550°C. No nucleobases were found to sublime from the meteorite powder at any of the temperatures in the experiments, but some of them could be recovered in sublimation experiments from the formic acid extract. Experimental evidence suggests that divalent cations such as Ca^{2+} and Mg^{2+}, and/or the presence of kerogen-type organic polymers in Murchison may inhibit the sublimation of these compounds. However, it is not fully understood yet which chemical or structural interactions within the meteorite are responsible for this inhibition. In any case, even though the total carbon flux from IDPs and micrometeorites is orders of magnitudes higher than for larger objects, the efficiency of their delivery for complex organic compounds is probably severely reduced.

Recent field evidence has suggested that the survival of organic compounds in massive impacts is probably higher than estimated in earlier studies. The thermodynamic evolution of the impactor in large
cometary impacts into the surfaces of Earth-like planets and moons can be simulated using high-resolution hydrocode modeling. This information can then be combined with laboratory thermodynamic data to determine the survival of organic compounds present in these bodies. The results of modeling massive impacts into the Earth have shown that the steady-state concentrations of some of the exogenously delivered amino acids in the early ocean could have equaled or substantially exceeded concentrations obtained from Miller-Urey-type experiments in realistic non-reducing atmospheres. Furthermore, grazing cometary impacts could have provided an increase of the steady state concentrations of certain amino acids by up to ten times. The delivery of such complex molecules to the surface of Europa is much less efficient. However, the cometary delivery of biogenic elements to the surface of Europa has shown to be able to supply total masses equivalent to 0.003 to 0.3% of the estimated total biomass in Earth's oceans today.

In summary, even though the fluxes and the delivery efficiencies for biogenic elements such as carbon is high for IDPs and massive impacts to Earth-like planets and moons, the delivery efficiencies for complex organic compounds important for life, such as amino acids, are much lower. In the case of the IDPs, the temperatures reached during atmospheric entry are extremely high, and since most of the amino acids apparently will not sublime from the meteorite grains, these organic compounds will probably be destroyed. In the case of cometary delivery to Europa, most the organic compounds are lost to space, even though their survival in the impact is probably significant.

ACKNOWLEDGMENT

This work was supported by the NASA Specialized Center for Research and Training in Exobiology, located at the University of California at San Diego. O. Botta is supported by an external ESA post-doctoral fellowship. E. Pietrazzo is supported by a grant from the SETI Institute (Mountain View, CA).

REFERENCES

Session 6
Comparing Atmospheres and Fluids
(with emphasis on Earth, Mars, Venus, Titan, Europa)

Chairs: J.-P. Lebreton, L. Becker & T. Encrenaz
ATMOSPHERIC STRUCTURE, COMPOSITION AND DIAGNOSTICS

Thérèse Encrenaz(1)

(1) LESIA, Observatoire de Paris, F-92195 Meudon, France, Email: therese.encrenaz@obspm.fr

ABSTRACT

Atmospheric properties of terrestrial planets and outer satellites are reviewed, with special emphasis on their composition, thermal structure and cloud structure. These properties are analysed in the frame of their formation and evolution. Infrared spectra of planets and satellites are used to retrieve element abundances and isotopic ratios which are diagnostics of formation and evolution processes. In conclusion, the main characteristics of the infrared spectra of terrestrial exoplanets are also discussed in order to estimate the kind of information which can be retrieved from these data.

1. EARTH-LIKE PLANETS AND MOONS: ATMOSPHERIC PROPERTIES

Earth-like planets and moons show a remarkable variety of atmospheres. Among the terrestrial planets, all objects but Mercury show a stable atmosphere; however, Venus, the Earth and Mars exhibit drastic difference in terms of surface temperatures and pressures. In the inner solar system, no satellite has a stable atmosphere. In contrast, three outer satellites – Io, Titan and Triton – have stable atmospheres, but again with very different conditions, since their surface pressures range from 1.5 bar in the case of Titan down to a few nanobars in the case of Io. Titan’s atmosphere shows analogies with the Earth’s one in surface pressure and global composition, in spite of its very low temperature. The stability of Io’s very tenuous atmosphere is maintained by a permanent volcanism, due to tidal effects generated by Jupiter’s neighborhood. Pluto does exhibit a stable atmosphere when the planet is close to perihelion, at 29.5 AU from the Sun. The period of Pluto’s orbit is 248 years and last perihelion occurred in September 1989. During the past decade, Pluto’s atmosphere was remarkably similar to Triton’s one, which is at a comparable heliocentric distance, both in pressure and global composition. Other solar-system bodies exhibit transient atmospheres, generated either by high-energy particle impacts (Mercury, galilean satellites) or by ice sublimation (comets). Table 1 shows the main properties of the stable atmospheres of Earth-like planets and moons.

<table>
<thead>
<tr>
<th>Object</th>
<th>Surface pressure (bar)</th>
<th>Surface temperature (K)</th>
<th>Composition</th>
<th>Nature of clouds</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venus</td>
<td>98</td>
<td>730</td>
<td>CO₂, N₂</td>
<td>H₂SO₄</td>
<td>Strong greenhouse effect</td>
</tr>
<tr>
<td>Earth</td>
<td>1</td>
<td>288</td>
<td>N₂, O₂</td>
<td>H₂O</td>
<td>Moderate greenhouse effect</td>
</tr>
<tr>
<td>Mars</td>
<td>0.006</td>
<td>220</td>
<td>CO₂, N₂</td>
<td>CO₂, H₂O</td>
<td>Weak greenhouse effect, H₂O in oceans</td>
</tr>
<tr>
<td>Io</td>
<td>1.4x10⁻⁴</td>
<td>110</td>
<td>SO₂</td>
<td></td>
<td>Non-uniform atmosphere (volcanism)</td>
</tr>
<tr>
<td>Titan</td>
<td>1.5</td>
<td>93</td>
<td>N₂, CH₄</td>
<td>CH₃ hydrocarbons</td>
<td>Continuous CH₄ outgassing</td>
</tr>
<tr>
<td>Triton</td>
<td>1.8x10⁻⁸</td>
<td>58</td>
<td>N₂, CH₄</td>
<td>Hydrocarbons</td>
<td>Cryovolcanism</td>
</tr>
<tr>
<td>Pluto</td>
<td>10⁻⁷</td>
<td>40</td>
<td>N₂, CH₄</td>
<td>Ps and Ts</td>
<td>Variable with Rh</td>
</tr>
</tbody>
</table>

Fig. 1. Thermal profiles of planets and satellites


© European Space Agency • Provided by the NASA Astrophysics Data System
Figure 1 shows the various thermal profiles as a function of pressure for all solar-system objects having a dense atmosphere. Giant planets are considered in this study as they all exhibit, at a pressure in the range 1-0.1 bar, a cloud layer which, in some ways, acts as a surface which reflects (or scatters) the solar radiation. If the range of temperatures and pressures, all profiles exhibit, above the surface (or the cloud layer), a "tropospheric" convective region where the temperature decreases with pressure down to a minimum (the tropopause). Above this level, two kinds of profiles can be considered: a quasi-isothermal mean mesospheric profile (Mars, Venus) with strong diurnal fluctuations, or a temperature inversion with an increase of the temperature in the stratosphere (Earth, giant planets, Titan). This increase is due to the absorption of the solar flux by an atmospheric constituent, gas or aerosol (O$_3$ for the Earth, CH$_4$ and hydrocarbons for Titan and the giant planets).

The global composition of the planetary atmospheres can be understood in the frame of their formation scenario (see below). While the atmospheres of terrestrial planets are dominated by CO$_2$ and N$_2$, the giant planets' atmospheres are mostly made of hydrogen and helium, with smaller quantities of CH$_4$ and other reduced elements. Titan's atmosphere is dominated by N$_2$ and CH$_4$, as in the case of Triton and Pluto. Io's atmosphere, in contrast, is mostly made of sulfur dioxide SO$_2$ generated by its permanent volcanism.

The cloud structure is very different from a planet to another, as it is due to the condensation of species which are often very minor. This is the case of the H$_2$SO$_4$ clouds of Venus, which result from chemical reactions involving in particular SO$_2$ and H$_2$O, two trace species in Venus' atmosphere. Mars' tenuous clouds are mostly made of H$_2$O ice, another minor atmospheric constituent. In Jupiter and Saturn, the main cloud observed in the visible range is NH$_3$, a very minor constituent. In the case of Uranus and Neptune, the temperature is low enough for methane to condense. In addition, a stratospheric haze is present, due to the condensation of hydrocarbons produced by methane photochemistry. The situation is the same on Titan, with the addition of nitrile condensates, which result from the dissociation of nitrogen by high-energy particles.

2. FORMATION AND EVOLUTION

Why do we observe two classes of planets in the solar system? This dichotomy naturally derives from the fact that the planets accreted from solid planetesimals within a protosolar rotating disk. In the vicinity of the Sun (R$_S$ < 2 AU) the temperature was such that only metals and silicates were in condensed form. In contrast, beyond about 2 AU ("snow line"), most of the elements (including C, N, O...) were in condensed form; according to the cosmic abundances, most of the mass (apart from H and He) was thus in solid form. Big solid, icy cores were formed with masses as high as 10-15 terrestrial masses (M$_E$), while at smaller heliocentric distances, the solid cores had masses of 1 M$_E$ or less. This led to a major difference in their subsequent formation, as, at large R$_S$, the massive cores were able to accrete the surrounding nebula, mostly composed of hydrogen and helium, thus forming the giant planets [1, 2]. In contrast, at R$_S$ < 2 AU, the gravity field of the smaller cores was not sufficient to keep the lightest gases. As a result, the terrestrial planets have a "secondary" atmosphere, partly outgassed from the globe and partly accumulated from meteoritical impacts. This simple scenario accounts for the two classes of planets observed in the solar system. With the discovery of protoplanetary disks around nearby stars, it was thought for some time that this scenario could apply also on other stellar systems. This is why the recent discovery of many giant exoplanets in the close vicinity of their stars raises an unexpected problem and requires different scenarios of planetary formation. With regard to "Earth-like" exoplanets, it can be noted that their radius cannot exceed 2.5 terrestrial radii, if their density is close to the Earth's one, or 3.5 R$_E$ at maximum, if they have an icy core; if not, they would become massive enough to accrete the protosolar nebula and they would become giant exoplanets (like Neptune, which has a radius of 3.8 R$_E$ for a total mass of 17 M$_E$).

The global chemical composition of planetary atmospheres can also be understood as a result of thermochemical equilibrium in the protosolar nebula. At low temperature and high pressure, carbon and nitrogen tend to form CH$_4$ and NH$_3$, while the reactions favor CO, N$_2$ and H$_2$ at high temperature and low pressure. In the subnebulae of the giant planets, methane and ammonia were the dominant species as observed in their atmospheres today. In the case of the terrestrial planets, H$_2$ escaped because of the low gravity field and CO subsequently formed CO$_2$. In the case of Titan, formed in Saturn's subnebula, NH$_3$ was probably photolyzed to form N$_2$, which explains the present (N$_2$, CH$_4$) atmosphere.

In the case of the terrestrial planets, H$_2$O, CO, O$_2$ and O$_3$ are found as minor species in the atmospheres of Mars and Venus, with, in addition, traces of sulfur species in the case of Venus (SO$_2$, H$_2$S, OCS) which lead to the formation of H$_2$SO$_4$ clouds. The case of the Earth's atmosphere is different: H$_2$O being in liquid form, CO$_2$ was trapped in the oceans in the form of CaCO$_3$, and O$_2$ built up in the atmosphere as a consequence of the
apparition of life. The main difference between the present terrestrial atmosphere and the other terrestrial planets' ones is thus the large abundance of water on Earth. Most likely, water was also present in large amounts in the early atmospheres of Mars and Venus; understanding its history on both planets is a major question for understanding the comparative evolution of the terrestrial planets.

In the giant planets, most of the minor species are in reduced form (CH₄, NH₃, PH₃, H₂O, H₂S...). Apart from CH₄, these species are not detected in Uranus and Neptune as they are trapped in clouds at deeper tropospheric levels due to the low atmospheric temperature. Hydrocarbons (C₂H₂, C₂H₆...), formed from the methane photolysis, are detected in the stratospheres of the giant planets as well as Titan. In addition, nitriles (HCN, C₂N₂, CH₃CN...) are found in Titan’s stratosphere as a result of the dissociation of N₂ by energetic particles.

3. THE INFRARED SPECTRA OF PLANETS AND SATELLITES

A solar photon penetrating a planetary atmosphere can either be reflected (or scattered), or converted into thermal heat and re-radiated at longer wavelengths. In the first case, the observed spectrum is the reflected solar spectrum with, in absorption, molecular signatures of the planetary atmospheric constituents; the information which is retrieved is the identification of the absorber and a measurement of its column density. In the second case, the thermal emission of the planets peaks at a wavelength which ranges from 7 µm (Mercury) to 70 µm (Neptune); the observed flux is strongly function of the temperature profile which generates either emission lines (in the stratosphere) or absorption lines (in the troposphere). In particular, in the case of the exoplanets, it is essential to have information about the thermal structure in order to disentangle emission and absorption features and to obtain reliable identifications. Information can then be retrieved upon the vertical distributions of the atmospheric constituents. It can be noted that the thermal emission range is especially suited for the study of neutral atmospheres, as molecules exhibit their strongest spectroscopic signatures (rotational and ro-vibrational bands) in this wavelength range, extending from the near-infrared to the millimeter range.

Figure 2 shows the near-infrared (day-side) spectra of Venus, the Earth and Mars [3, 4, 5]. The reflected sunlight component prevails at wavelengths shorter than 3 µm. In the case of Venus, the solar light is reflected above the H₂SO₄ cloud level at a pressure level of about

![Near-infrared spectra of Venus, the Earth and Mars. Figures are taken from Refs. [3], [4] and [5]. In the case of Venus, the reflectivity is shown.](image)

![Near-IR spectrum of the night-side of Venus. The figure is taken from Ref. [6].](image)
1 bar, so that the lower troposphere cannot be probed. It is possible, however, to probe this region at these wavelengths by observing the night side of Venus, between the very strong CO$_2$ absorption bands (Fig. 3; [6]). In the case of the Earth, the near-IR spectrum is dominated by the signatures of H$_2$O and CO$_2$, so that the surface cannot be observed. In contrast, as a consequence of the low Martian surface pressure, the Mars spectrum exhibit some signatures of the surface emissivity between the CO$_2$ absorption bands.

Figure 4 shows the infrared spectrum of Jupiter, Saturn and Neptune between 2 and 16 μm [7]. Thermal emission dominates at wavelengths larger than 4 μm. It can be seen that signatures of stratospheric species appear in emission (CH$_4$ at 7.7 μm, C$_2$H$_6$ at 12 μm, C$_2$H$_2$ at 13.7 μm) while tropospheric species (NH$_3$ at 10.5 μm in Jupiter, PH$_3$ at 8.9 μm in Saturn) exhibit absorption bands. In the case of the terrestrial planets (Fig. 5; [8]), the strong CO$_2$ 15-μm band, formed in the troposphere, appear in absorption in all cases. Water is present in the Earth spectrum and also marginally on Mars. In the case of Venus, the thermal radiation comes from the H$_2$SO$_4$ cloud level outside the CO$_2$ bands. Ozone is present in the Earth spectrum (although formed in the stratosphere, the band is seen in absorption because the stratospheric temperature is still lower than the surface temperature); if detected on an exoplanet, such a signature, associated to H$_2$O and CO$_2$, might be the signature of biogenic activity.

Fig. 4. The infrared spectra of Jupiter, Saturn and Neptune as observed by ISO-SWS. In the thermal regime (λ > 4 μm), stratospheric hydrocarbons exhibit emission features, while tropospheric species (NH$_3$, PH$_3$) appear in absorption. The figure is taken from Ref. [7].

Fig. 5. Thermal emission from the terrestrial planets in the mid- and far-infrared range (5-100 μm). The figure is taken from Ref. [8].

4. ELEMENT ABUNDANCES AND ISOTOPIC RATIOS

Diagnostics of formation and evolution processes can be found in the measurement of element abundances and isotopic ratios of volatiles. In the case of the giant planets, in particular, the nucleation model described above predicts an enrichment of C/H, N/H, O/H...with respect to the cosmic values, because these ratios were strongly enriched in the ices which made the initial cores and were later outgassed in the outer atmospheres of the giant planets following the accretion of the surrounding protosolar gas. To a less extent, the same comment applies to D/H, as deuterium is known to be enriched in ices as a result of ion-molecule reactions at low temperature, as observed in the interstellar medium [9]. In the case of Jupiter, the Galileo probe has provided us with a clear confirmation of the nucleation theory by measuring a factor 3 enrichment for C, S, N, Ar, Kr and Xe, but has also raised unanswered questions about the low-temperature formation of the planetesimals which formed Jupiter [10].

In the case of Earth-like planets, element abundances and isotopic ratios can be used for (1) identifying
possible reservoirs of volatiles (protosolar nebula, solar winds, meteorites, icy planetesimals...), (2) identifying possible mechanisms of gain and loss (outgassing, accretion, giant impact, meteoritic/cometary infall...), and (3) identifying possible fractionation processes (hydrodynamical escape, adsorption of nebular gases...). Key measurements are, in particular, the abundances of C, N and noble gases, and the isotopic ratios.

4.1 Element Abundances

Figure 6 shows the relative abundances of C, N and the rare gases normalized to $^{84}$Kr, with respect to their solar value. It can be seen that C and N exhibit an enrichment for all terrestrial planets, which illustrates that these species came on the terrestrial planets at least partly in solid form. The rare gases exhibit a regular increase from Ne to Kr. This behaviour can be interpreted by adsorptive fractionation (implantation of rare gases from the solar wind in the surface grains; [11]). This process also explains the larger rare-gas abundance observed in Venus, which, being closer to the Sun, must have been more affected by solar-wind implantation [12].

4.2 The D/H ratio

In the Big Bang Standard Model, deuterium is entirely formed by primordial nucleosynthesis, and is later continuously destroyed in stars where stellar nucleosynthesis converts it into $^3$He. The protosolar value of D/H thus gives us a measurement of this parameter 4.6 Gy ago, while its value in the local interstellar medium refers to present.

Figure 7 shows the measurements of D/H in planets and satellites, compared to its protosolar and interstellar values. Several means have been used to determine the D/H ratio: $^3$He in the solar wind (protosolar value), HD infrared lines (giant planets), CHD infrared lines (giant planets and Titan), HDO lines (terrestrial planets and comets) and finally in-situ mass spectroscopy of HDO (Jupiter and comet Halley).

As mentioned above, D/H is expected to be enriched in the ices. As the relative mass fraction of the initial icy core, following the nucleation scenario, increases from 3% (Jupiter) to more than 50% (Uranus and Neptune), the D/H enrichment is expected to increase accordingly. In the case of Jupiter, the expected enrichment is only a few percent, so that the Jovian value can be considered as representative of the protosolar value, as confirmed by the Galileo [13] and ISO [14] results. Uranus and Neptune exhibit a deuterium enrichment, as measured by ISO, in agreement with the expectations [15]. It is
interesting to note that the D/H value in proto-neptunian ices, inferred from these measurements through a modelling of the planets' interiors, is about twice lower than what is observed in the three Oort-cloud comets for which D/H measurements have been made, using the main volatile H2O (Halley, Hyakutake and Hale-Bopp; [16]). If confirmed, this result will provide constraints upon the formation conditions of the planetesimals which formed the giant planets; it will be also very important in the future to measure D/H in Kuiper-belt comets and TNOs.

D/H in Titan, measured in CH4, also shows a significant enrichment as compared to the protosolar value; this illustrates that, as expected, its atmosphere has not been accreted from the surrounding protosolar subnebula, but has been outgassed from the satellite's interior. The terrestrial value of D/H, measured in the oceans, is also significantly higher than the protosolar value, but is also about twice lower than the cometary value. This implies that the terrestrial water cannot come entirely from comets, but must have at least partly another origin (outgassing, asteroidal impacts). Finally, the strong deuterium enrichment observed on Mars (5 times terrestrial; [17]) and even more on Venus (over 100 times terrestrial; [18]) is currently interpreted as the signature of differential escape, which implies that large abundances of water were present in the early history of the two planets. How water disappeared from Venus is still an open question. H2O might have been photodissociated, implying subsequent hydrogen escape; however the escape of oxygen atoms is not clearly understood. In the case of Mars, other indices suggest the presence of an early ocean of water, which also implies that the early atmosphere of Mars was significantly warmer and more massive than today. Water is expected to be present under the surface in form of water ice or permafrost. The Mars Odyssey mission has recently confirmed this hypothesis by detecting water ice under the surface of the southern polar cap.

4.3 The 15N/14N ratio

The 15N/14N ratio was measured on Mars by Viking, using in-situ mass spectrometry; with respect to the terrestrial value, an enrichment by a factor 1.6 was found. This was interpreted as the signature of an early outgassing of the N2 Martian atmosphere [19], which is consistent with the D/H result mentioned above.

An even stronger 15N enrichment, by a factor 5 was measured in Titan, from millimeter heterodyne spectroscopy measurements of HCN and HC15N (Fig. 8; [20, 21]). Similarly, the 18O/16O ratio in Titan, measured from millimeter CO lines, was found to be twice the terrestrial value [22]; its interpretation, however, is more ambiguous as CO is probably, at least partly, of external origin [23, 24]. Both result again suggest an early massive atmospheric escape; however, non-thermal escape mechanisms might be needed to account for the high fractionation factor [25].

It should be mentioned that recent measurements of 15N/14N in Jupiter, both by ISO and Galileo, have shown that this ratio is depleted by a factor 2 with respect to the terrestrial value [26, 27]. The current interpretation is that the Jovian value is representative of the protosolar value, which implies that protosolar nitrogen was in the form of N2, while the terrestrial isotopic ratio (as well as the ratio measured in comet Hale-Bopp) refers to N-compounds (NH3, HCN) in which 15N/14N is enriched, as observed also in the interstellar medium [27].

5. EARTH-LIKE EXOPLANETS

In this section, we will try to use our knowledge of the solar-system planetary atmospheres in order to investigate the possible atmospheric properties of exoplanets, their expected spectra, and the kind of information which we can hope to retrieve from these spectra.

Infrared spectra of exoplanets are expected to show two components, a stellar reflection spectrum and a thermal emission. The shapes and intensities of these two components will depend upon four parameters: (1) the spectral type of the star, (2) the stellar luminosity; (3) the distance of the planet to the star; (4) the albedo of the planet. In particular, the effective temperature of the planet will be inversely proportional to the square root
of the asterocentric distance, and directly proportional to
the fraction of stellar energy absorbed by the planet
(1-a). Figure 9 displays a series of spectral profiles
corresponding to various heliocentric distances, for a
solar-type star of 1 solar mass and a planetary albedo of
0.3. It can be seen that, for an asterocentric distance $R_a$
of 0.05 AU (which corresponds to the case of many
 giants exoplanets recently detected), the thermal
emission starts to dominate at wavelengths above 1.25
$\mu$m.

The composition and structure of giant exoplanets has
been investigated by several authors for various
asterocentric distances. Assuming thermochemical
equilibrium, models predict, in the upper atmosphere,
a composition dominated, apart from H$_2$ and He, by (CO,
N$_2$) at $R_a < 0.06$, (CH$_4$, N$_2$) between 0.06 and 0.2 AU,
and (CH$_4$, NH$_3$) at $R_a > 0.2$ AU [28]. Other processes,
however, like photodissociation and photochemistry,
will have to be taken into account, especially in the case
of Earth-like exoplanets. We have to remember the case
of Titan, where the early outgassed atmosphere was
probably made of CH$_4$ and NH$_3$, but NH$_3$ was later
photodissociated and converted into N$_2$.

Fig. 10. Synthetic transmission spectra of molecular species in
the infrared range, calculated for a column density of 10 cm-
am. From top to bottom: H$_2$O, CO$_2$, CO, CH$_4$, NH$_3$.

Molecular signatures expected to be found in the spectra
of Earth-like exoplanets are thus CH$_4$, NH$_3$, CO, as well
as H$_2$O and possibly CO$_2$. Are there spectral
wavelengths where the surface could be probed? Figure
10 shows transmission spectra of H$_2$O, CO$_2$, CO, CH$_4$
and NH$_3$ between 2 and 20 $\mu$m, calculated for a column
density of 10 cm-am. The surface of an Earth-like
exoplanet could be probed around 4 $\mu$m, 5 $\mu$m, 8.5 $\mu$m,
13 $\mu$m and 19-20 $\mu$m. If NH$_3$ is absent (as in the case of
a N$_2$-dominated atmosphere), the whole 8-13 $\mu$m range
becomes free of gaseous absorptions. This opportunity
is used in the concept of the DARWIN mission, which
is designed to search for the ozone signature in the
atmospheres of exoplanets at 9.5 $\mu$m.

Fig. 9. Reflected stellar component (left side, maximum flux
at 0.5 $\mu$m) and thermal emission (right side) of exoplanets, for
various asterocentric distances. From top to bottom: $R_a =
0.05, 0.3, 1.0, 5.0, 20.0$ AU.
6. REFERENCES


TITAN WIND EFFECTS ON THE DESCENT TRAJECTORY OF THE ESA HUYGENS PROBE

B. Kazeminejad1, J.-P. Lebreton1, M.K. Bird2, and D.H. Atkinson3

1ESA Research and Scientific Support Department, ESTEC/SCI-SB, 2200 AG Noordwijk, The Netherlands
2Radioastronomisches Inst., Univ. Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
3Dept. of Electrical and Computer Engineering, University of Idaho, Moscow, ID 83844-1023, USA

ABSTRACT

The Huygens Probe is the ESA-provided element of the joint NASA/ESA Cassini/Huygens mission to Saturn and Titan. The Cassini/Huygens spacecraft was launched on 15 October 1997 and will arrive at Saturn on the 1st of July 2004. The Huygens probe will be released on 24 December 2004 and enter the atmosphere of Titan on 14 January 2005. A recently discovered design flaw in the Huygens radio receiver onboard Cassini led to a significant redesign of the mission geometry by both the Huygens and Cassini project teams. In this new scenario the Orbiter will pass Titan at high altitude (i.e., 60,000 km) on the retrograde side of Titan and will trail the Probe by only about 2.1 hours instead of the originally planned 1250 km flyby altitude on the prograde side of Titan and a 4 hour delay time. Among the factors governing the duration and quality of the Cassini/Huygens communication window during the descent is the Probe drift caused by zonal winds. Existing Titan wind models have been reevaluated and compared to recent ground-based observations. Simulations of the Probe entry and descent show a drift from ~300 km up to ~430 km away from the “no wind” landing point, depending on the wind model. At the end of the nominal mission this difference in wind drift (assuming prograde winds) causes a difference of up to 1.7 dB (within a margin of 3 to 4 dB, resulting from the receiver design flaw) in the received signal-to-noise ratio. The high sensitivity of the received signal strength to zonal winds and their directions is due to the steep decrease of the Probe antenna gain when the Cassini spacecraft (with the Huygens receiver) as seen from the Probe moves to increasingly higher elevation angles. A simulation of the Probe atmospheric entry phase shows that the zonal wind direction also impacts the shape of the deceleration profile and its peak value. The deceleration profile will be accurately measured during the entry phase by accelerometers onboard the Probe. From this data set the density profile of the upper Titan atmosphere will be inferred. This will complement the orbiter instrument measurements planned during the early Titan flybys for validation of the upper Titan atmosphere model and evaluation of the drag force that will act on the Cassini spacecraft at subsequent low altitude Titan flybys (~950 km).

The atmosphere and surface of Saturn’s moon Titan are primary targets of the Cassini mission and its Huygens Probe (Lebreton & Matson 1997). Titan and the planet Venus stand together as a limiting case in planetary meteorology and atmospheric dynamics, slowly rotating bodies with dense atmospheres, unique among objects in our solar system. Titan’s rotation period of 16 days implies an equatorial surface rotation velocity of about 11.7 m/s (40 times slower than the Earth). Measurements of thermal gradients on Titan (Flasar et al. 1981) as well as an occultation of the star 28 Sgr (Sicardy et al. 1990) can be interpreted as indicators of zonal wind velocities ~10 times the speed of surface motion. Such winds will clearly transport the Huygens Probe horizontally during its parachute descent sequence, thereby shifting the landing site and possibly shortening the duration of the communication window to the Orbiter. This depends critically on the relative geometry of the Huygens transmitting antennas and the Orbiter high gain antenna.

The present work compares different wind models and their impact on the redesigned Huygens mission scenario (Huygens Recovery Task Force 2001). Theoretical wind models are unable to determine the wind direction. However, first results from direct wind measurements on Titan using spectroscopy measurements of Doppler-shifted ethane (C2H6) emission lines from the stratosphere (Kostiuk et al. 2001) claim, with 94% statistical confidence, a prograde zonal circulation, i.e., in the direction of global rotation. Another goal of this work is to evaluate the influence of the wind direction on the entry deceleration profile.

2. THE LELLOUCH-HUN TEN TITAN ATMOSPHERE MODEL

The Voyager radio occultation data provided the first vertical pressure and temperature profiles (ingress and egress


© European Space Agency • Provided by the NASA Astrophysics Data System
The measurements) between the ground and ca. 200 km altitude for a pure $N_2$ atmosphere (Lindal et al. 1983). Above 200 km, the only available data give the temperature and number density at 1270 km (Smith et al. 1982). Hunten (1987) developed an atmosphere model that consists of the average of the two Lindal profiles for the range of altitudes from 0-200 km and an isothermal interpolation obtained by fitting the hydrostatic equation to the data points at 200 and 1270 km. The discontinuity at the 200 km level was removed by multiplying all the densities by a factor of 1.18.

The Lellouch-Hunten (1987) model (Lellouch & Hunten 1987) is based on a reanalysis of the radio occultation measurements (Lindal et al. 1983) considering uncertainties due to experimental noise and chemical composition. For altitudes under 200 km a nominal and two extreme profiles were derived. For the 200-1250 km range the Hunten (1987) model was adapted with some modifications to get better consistency with the methane emission at 7.7 $\mu$m and to be in qualitative agreement with aeronomical models predictions (cp. Lellouch et al. 1990). Fig. 1 and 2 show the extreme temperature and density profiles. The minimum and maximum temperature are, level by level, equal to min ($T_{\text{rec}}$, 175K) - 30$K$ and max ($T_{\text{rec}}$, 175K) + 30K, where $T_{\text{rec}}$ is the temperature value of the recommended profile. These extreme profiles are not consistent with the UVS densities from Smith et al. (1982), but provide the estimated range of acceptable temperatures for each level. The uncertainty on the density increases rapidly with altitude in parallel with the uncertainty in temperature, reaching a factor of 10 at 800 km (minimum and maximum density in Fig. 2). The Lellouch-Hunten (1987) recommended model was the nominal Titan atmosphere model that was used for mission analysis within the framework of the redesign of the Huygens mission (Huygens Recovery Task Force 2001). We therefore adopted it as standard atmosphere for all numerical simulations presented here.

3. ZONAL WIND MODELS

The meridional pressure gradient is related to the zonal winds by the gradient wind equation

$$u \left[ 2 \Omega \sin \Lambda + \frac{u \tan \Lambda}{r} \right] = -\frac{1}{\rho} \frac{\partial \rho}{\partial \Lambda},$$

where $\Lambda$ is the latitude and $\Omega$ is the angular rotation velocity. The pressure gradient is taken along a surface of constant geopotential in the reference state. The first term on the left of Eq. (1) is the Coriolis force which dominates in rapidly rotating planets, such as the Earth, Mars and the outer planets. The balance of the Coriolis force and the pressure gradient is termed geostrophic. On the other hand, for slow rotators such as Venus and Titan, the second term is thought to dominate throughout much of the atmosphere. In this case the balance of the centrifugal force and the pressure gradient is called cyclotrophic. Assuming hydrostatic equilibrium (i.e., $\partial p/\partial z = -\rho \ddot{g}$, where $\ddot{g}$ is the effective gravity), Eq. (1) can be written as

$$\frac{\partial}{\partial \ln p} \left[ u \left( 2 \Omega \sin \Lambda + \frac{u \tan \Lambda}{r} \right) \right] = \frac{R}{\mu} \left( \frac{1}{r} \frac{\partial T}{\partial \Lambda} \right) \rho,$$

where $T$ is the temperature, $R$ the universal gas constant, and $\mu$ the molecular weight. Eq. (2) relates the vertical gradient in the zonal wind to the meridional gradient in temperature (cp. Flasar & Conrath 1992).

The Flasar Wind Model

Knowing the meridional temperature gradient, one gets the zonal wind height profile by integrating Eq. (2). This was done by Flasar et al. (1981) using the horizontal temperature profiles derived from observations by the Voyager infrared spectroscopy experiment (IRIS). These observations covered most latitudes between 60° N and 60° S, with some zonal (i.e., east-west) coverage at certain
Figure 3. Vertical profile of model zonal wind speed (at 45° latitude) for the equator-to-pole temperature contrasts taken from Voyager IRIS brightness temperatures (thick solid line). Other curves: Flasar 2 × ΔT envelope (dashed line); Aerospatiale wind model (thin blue line).

Figure 4. The Aerospatiale zonal wind model for three different altitudes. The upper line (219 km) corresponds to a pressure of 0.5 mbar.

Figure 5. The HRTF wind model for the altitudes shown in Fig. 4.

The Maximum Wind Envelope

Given the limited spatial coverage in latitude and altitude and the errors in the retrieved temperatures, application of the thermal wind equation Eq. (2) can provide only a rough determination of the zonal winds. If the equator-to-pole temperature variations are doubled to account for the uncertainties (referred to as 2 × ΔT profile; see Fig. 3: dashed line), a zonal wind height profile is derived with zonal velocities a factor of the order of √2 larger than the best estimate (Lunine et al. 1991; Flasar et al. 1997). Lunine et al. (1991) and Flasar et al. (1997) suggested a simple engineering model, based on an integration of Eq. (2), which describes this maximum envelope. The temperature gradient increases roughly linearly with height according to the formula

$$|u - u_0| \leq 200 \text{ m s}^{-1} \left[ 1 + \frac{1}{8} \ln \left( \frac{0.5 \text{ mbar}}{p} \right) \right] \cos \Lambda \quad (3)$$

where Λ is the latitude and $u_0$ the velocity of the surface wind (which can be neglected due to surface friction). The latitudinal dependence in Eq. (3) corresponds to superrotation at constant angular velocity (rigid rotation). Eq. (3) was used for mission analysis by the Huygens prime contractor Aerospatiale (now: Alcatel Space) in Cannes, France. This Aerospatiale Model, defined by Eq. (3), is seen to differ from the Flasar maximum envelope latitude in Fig. 3. Note that for pressures lower than 0.5 mbar (heights < 219 km) the Aerospatiale model assumes wind speeds equal to those at the 0.5 mbar level given by Eq.(3). The Aerospatiale model thus implies a maximum wind speed of 200 m/s which is attained at the equator at pressure levels ≥ 0.5 mbar (see Fig. 4).

The HRTF Wind Model

The HRTF Model was proposed in the framework of mission analysis done by the Huygens Recovery Task Force (Lebreton 2001). Wind speed profiles for three different altitudes are shown in Fig. 5. The HRTF model is based on the equations used for the Aerospatiale model, but the wind velocities are reduced by a factor of √2.
maximum wind speed at 0° latitude and \( p \geq 0.5 \, \text{mbar} \) thus becomes 140 m/s, which is about the value proposed by the Flasar 2 × \( \Delta T \) model in the upper stratosphere. The more important aspect of the HRTF model was to account for two different versions: one representing prograde zonal wind direction (i.e., in the direction of the rotation of the satellite) and a second for retrograde winds. This was an important step for mission analysis to better understand the impact of the wind direction on important mission parameters.

The Engineering Wind Envelope

Flasar et al. (1997) developed an engineering wind envelope using the Flasar 45° wind profile. Taking theoretical studies (Allison et al. 1994) and general circulation modelling (Del Genio et al. 1993) into account, the variation of the winds to lower and higher latitudes is based on the limiting cases of either constant angular momentum \( (m) \) or constant angular speed \( (\omega = u/r \cos \Lambda) \) on isobaric surfaces, depending on the value of the Richardson number \( R_e \).

The zonal wind speeds are assumed to vary in the troposphere as \( \cos \Lambda (\text{constant} \omega) \) along isobars at latitudes equatorward of 45°, and with \( m = \text{const.} \) along isobars poleward of 45° up to 60° latitude. From the surface to the tropopause (\( \sim 100 \, \text{mbar} \)), the envelope winds thus vary with latitude as \( u_{\text{trop}}(p, \Lambda) = \)

\[
\begin{align*}
&\begin{cases}
  u(p, 45°) \sqrt{2} \cos \Lambda & : |\Lambda| \leq 45° \\
  u(p, 45°) + \frac{\Omega}{r} \sqrt{2} \cos \Lambda & : 45° < |\Lambda| \leq 60°
\end{cases}
\end{align*}
\]

The upper stratospheric data from Voyager’s IRIS near Titan’s northern spring equinox and from the 28 Sgr occultation (Sicardy et al. 1990) near northern summer solstice both suggest that the winds satisfy a constant \( m \) scaling on isobaric surfaces from low to mid latitudes up to 60°. At these altitudes, therefore, it is assumed that the meridional variation of the winds is one of constant \( m \) on isobars from the equator to 60°. For \( p \leq 0.5 \, \text{mbar} \) and \( 0° < |\Lambda| < 60° \), \( u \) varies

\[
u_{\text{upstrat}}(p, \Lambda) = \frac{u(p, 45°) + \Omega r / \sqrt{2} \cos \Lambda}{\sqrt{2} \cos \Lambda} - \Omega \cos \Lambda.
\]

(5)

At intermediate pressure levels (0.5 \( < p(\text{mbar}) \) \( < 100 \)) Eqs. (4) and (5) are crudely weighted and the mean zonal wind is determined by \( u_{\text{trop}}(p, \Lambda) = \)

\[
\frac{1}{\ln 200} \left[ \frac{u_{\text{trop}} \ln \left( \frac{p}{0.5 \, \text{mbar}} \right)}{p} + \frac{1}{\ln 200} \left[ u_{\text{upstrat}} \ln \left( \frac{100 \, \text{mbar}}{p} \right) \right] \right]
\]

(6)

Wind velocities for latitudes from the equator to ±60° are shown in Fig. 6. The engineering model described in Eqs. (4), (5) and (6) is referred to in the following sections as the Engineering Wind Envelope (EWE).

---

Theoretical Consideration of Wind Direction

From existing data, the thermal wind analysis of the previous sections says nothing about the direction of the zonal winds. The gaps in the IRIS vertical resolution preclude any such conclusion. On Venus, the one planet known to have a cyclotrophic atmospheric global circulation, the winds do blow in the direction of the planetary rotation. Why this might be so follows from a consideration of the “spin-up” of a planetary atmosphere from a state of rest. One approach is to consider Kelvin’s Circulation Theorem for a barotropic fluid (Flasar & Conrath 1992). For the case of axially symmetric motion, this reduces to conservation of the axial angular momentum:

\[ u \, R + \Omega \, R^2 = u_0 \, R_0 + \Omega \, R_0^2 \]

(7)

where \( R \) is the axial distance from the rotation axis, and the subscripts ‘0’ denote initial values. Taking the above relation, it can be shown that planetary rotation establishes a bias such that even initially retrograde winds (\( u_0 < 0 \)) eventually spin up to prograde winds with sufficient decrease in scale (\( R / R_0 \) \( \to 0 \)). Retrograde winds can be maintained only if they are sufficiently strong or the vortex sufficiently small (\( u_0 / \Omega \, R_0 < -1 \)). For example, assuming an initial velocity of \( u_0 = 1 \, \text{ms}^{-1} \) and a contraction of \( R / R_0 = 0.5 \), Flasar & Conrath (1992) compute a critical initial scale \( R_0 = 290 \, \text{km} \) for Titan, above which motions are dominated by planetary rotation rather than initial conditions. Because the critical scale on Titan is much smaller than the planetary radius, one would thus expect Titan’s cyclotrophic zonal flow to be prograde. It must be kept in mind, however, that atmospheric dynamical systems are nonlinear and not always amenable to theoretical predictions. An independent source of information comes from recent ground-based observations which also suggest prograde zonal winds in Titan’s upper atmosphere (Kostiuk et al. 2001) (see Sec.4).
4. EARTH-BASED WIND MEASUREMENTS

The first direct measurements of the global circulation of Titan have been determined by measuring the Doppler shift of ethane lines emitted from Titan's stratosphere (~0.1-7 mbar) (Kostiuk et al. 2001). A prograde global wind direction was determined at the 94% confidence level with speeds estimated to be 210±150 m/s. A limited signal-to-noise ratio for these difficult observations, however, precluded discrimination among different wind field models.

Further Earth-based observations of Titan's winds are planned with the VLT UV-Visual Echelle Spectrograph (UVES) at ESO (Courtin 2001). These observations attempt to measure the differential Doppler shift introduced by the zonal wind flow in the back-scattered solar radiation from the opposite limbs of Titan (East and West). Realizing that the expected wind velocities are too low to produce a Doppler shift on individual solar lines large enough for the spectral resolution of UVES, it is planned to apply a retrieval scheme developed for stellar accelerometer to a large number of well-suited solar lines at the available resolution. Both limbs are to be observed simultaneously with a 0.3" entrance slit placed along the equator. This will eliminate most unwanted spurious Doppler shifts (Earth rotation, motion with respect to Titan, pointing jitter, instrumental drift, etc), resulting in significant reduction of the measurement error.

5. THE PROBE TRAJECTORY

Cassini/Huygens is a collaborative ESA/NASA mission. It is NASA's responsibility to deliver the Probe to the ESA/NASA interface point, which was defined as 1270 km above the surface of Titan (GMV 2001). As part of the interface definition a set of files with initial state vectors of Probe and Orbiter as well as delivery uncertainties and Titan and Saturn ephemerides are communicated to ESA for further propagation of the Probe entry and descent trajectory.

The entry and descent trajectory of the Huygens probe and the orbit of Cassini are calculated by means of numerical integration of the equations of motion, using a 7th order Runge-Kutta-Fehlberg method. The Orbiter and Probe state vectors at the interface altitude (defined as 1270 km above the Titan surface) were provided by Cassini Navigation (NASA/JPL) together with their uncertainties and propagated over a time span of 180 min. We simulated the atmospheric entry of the Probe using the Lellouch-Hunten atmosphere model (Lellouch & Hunten, 1997), and the wind models described in Sec.3. The acceleration due to wind drag $a_d$ was taken from Flury (1986) to be:

$$a_d = -\frac{\rho(z)}{2C_B} V_{PW} V_{PW}$$

where $\rho(z)$ is the atmospheric density, $V_{PW} = V_P - V_W$, is the relative velocity of the Probe with respect to the atmospheric wind, and $C_B$ is the probe ballistic coefficient defined by

$$C_B = \frac{m}{C_D A}$$

with $m = $ Probe mass, $C_D = $ drag coefficient, and $A = $ effective Probe area to the flow.

Third body perturbations due to Sun and Saturn were taken into account. Both Probe and the Orbiter were considered to be point masses. No modifications were introduced to simulate either the Probe's specific attitude or angle of attack. Probe aerodynamics were taken into account in a simplified model, where the drag coefficient is considered to be a function of only the Mach number (i.e., $C_d = C_d(M)$) during the entry phase and the Reynolds number (i.e., $C_d = C_d(R_e)$) during the descent.

The impact of errors on various input parameters was evaluated by means of a specially developed Monte Carlo simulation technique (also referred to as "Mini Monte Carlo" (cp. Huygens Recovery Task Force 2001,p48+)), which uses the numerically propagated system covariance matrix. This 23 x 23 matrix comprises errors of the Probe and Orbiter state vectors (12 variables), the Saturn and Titan gravitational constants (2 variables), errors of Titan atmosphere and wind models, Probe entry mass error, Titan radius and topography errors and, finally, icing mass and acceleration limit detection errors (cp. Mora & Nogales 2000,p32+). The covariance matrix is defined at entry altitude and then propagated from entry time down to surface impact by evaluating the system transition matrix $\Phi_{t,j}$ using a numerical differentiation scheme. This readily allows one to obtain the perturbed state vector of Probe and Orbiter during various mission times and thereby accumulate statistics on dependent mission parameters. Furthermore, it is used to determine the dispersion ellipse around the nominal landing point on the surface.

![Figure 7. Deceleration profiles during the Huygens entry phase assuming the recommended Lellouch-Hunten atmosphere and nominal entry angle $\gamma = -64^\circ$; Important acceleration detections are shown: $T_a =$ arming timer triggering; $S_0 =$ POSW mission time start; $T_0 =$ triggering of Parachute Deployment Device (PDD);](image-url)
The Entry Phase

The entry phase is assumed to start at 1270 km above Titan surface (Lellouch-Hunten Model "reference" altitude) with a nominal Probe entry angle of -64° and a Probe mass of 319 kg. The nominal entry sequence comprises the following important Probe accelerometer detections:

- $T_a$ detection at 80 m/s² (positive slope)
- $S_0$ detection at 10 m/s²
- $T_0$ detection ($T_0 = S_0 + 6.375$ sec)

where $T_0$ designates the start of the parachute deployment sequence, $T_a$ the arming timer trigger, and $S_0$ the starting epoch of the Probe Onboard Software (POSW). These events are indicated in the entry acceleration profile shown in Fig. 7. The impact of the wind direction on the profile was investigated by propagating the Probe entry trajectory for both prograde and retrograde winds of the same model. Fig. 8 shows the deceleration profiles for the nominal entry angle ($\gamma = -64^\circ$) and two extreme values of $-60^\circ$ and $-68^\circ$ assuming the HRTF zonal wind. The differences of the deceleration peaks between prograde and retrograde wind ($\Delta a_p$) for three wind models are listed in Table 1. Although the differences increase with the wind velocities assumed in the model, they tend to be quite moderate. The HRTF cases show that the deceleration peak difference is higher ($\Delta a_p \approx 3.25$ m/s²) for the low entry angle (i.e., $-60^\circ$) compared to the nominal value with $\gamma = -64^\circ$ ($\Delta a_p \approx 2.98$ m/s²) or the steep entry angle with $\gamma = -68^\circ$ ($\Delta a_p \approx 2.76$ m/s²).

The Huygens Atmospheric Structure Instrument (HASI) will be the only instrument onboard the Probe that will operate during the entry phase (Fulchignoni et al. 1997). The atmosphere's density profile $\rho(z)$, is proportional to the Probe entry deceleration profile that will be measured by the HASI 3-axis accelerometer. Direct measurements of the direction and strength of zonal winds will only be available in the altitude range from ~160 km down to the surface (Bird et al. 1997). The impact of high stratospheric winds on the deceleration profile during the Probe entry phase will therefore introduce an uncertainty in the inferred density profile.

Simulation of Descent and Landing

Parachute aerodynamics are implemented by 3 different configurations: main chute plus decelerator, main chute only and stabilizing drogue only. Each configuration is represented by a different value for the cross section, drag coefficient and mass (Mora & Nogales 2000). Fig. 9 compares the Probe trajectory (i.e., east longitude in a Titan centered bodyfixed frame) for prograde and retrograde wind (solid lines). The dashed line results from the simulation without wind. Once the parachute sequence has started (begin of descent phase, see Fig. 9) the Probe is, depending on the wind direction, carried to higher (in case of prograde wind) or lower (in case of retrograde wind) East longitudes. Surface impact is ~140 minutes after entry. Fig. 10 shows the atmospheric pressure as a function of descent time assuming the recommended Lellouch-Hunten atmosphere model as depicted in Fig. 1. The Probe enters the tropopause ~47 minutes after entry at 1270 km Fig. 11 shows the zonal wind velocity at the Probe during entry and descent for the three different wind models explained in Sec.3. The crossing of the tropopause is marked with the black dashed line. The Aerospatiale model has the highest wind speeds (up to 200 m/s) in the stratosphere, but the Engineering Wind Envelope dominates in the troposphere. The HRTF model assumes the lowest wind speeds in both cases. The Flasar 2 $\times$ $\Delta T^\circ$ profile (see Fig. 3) advocates a wind speed of 150 m/s in the upper stratosphere, which, to first order, would justify the 140 m/s given by the HRTF model. However, in view of recent direct earth-based wind measurements (see Sec.4), which suggest wind speeds of 210 m/s at pressures ~0.1-7 mbar (~300-120 km), the HRTF model would seem to underestimate the zonal winds in that range of altitudes.

Table 1 gives the Probe landing site east longitude and the wind drift (i.e., the distance from the Probe impact point computed without wind drag: LON = 160.48° E; LAT = 10.74°S) for the three different wind models. Note that the wind drift of the Probe is slightly smaller for retrograde winds. This is due to the atmosphere, which was assumed to rotate in a prograde direction with the same angular velocity as Titan.

Table 1 demonstrates how the different wind models affect the landing coordinates of the Probe. The impact points of the Aerospatiale and the EWE wind models were very close together. Even though the Aerospatiale model wind velocities are considerably higher in the stratosphere, this effect seems to be compensated by slightly higher EWE velocities in the troposphere. This makes sense, because the troposphere descent phase lasts about twice as long as the descent in the stratosphere (see Fig. 11). The Probe drifts ~127 km further (to the East

![Figure 8. Deceleration profiles for $\gamma = -68^\circ$, $\gamma = -64^\circ$ and $\gamma = -60^\circ$.](image)
Table 1. Titan body-fixed landing site coordinates for the three wind models and the corresponding differences $\Delta \alpha_p$ of deceleration peaks for prograde and retrograde wind; nominal entry angle $\gamma = -64^\circ$:

<table>
<thead>
<tr>
<th>Wind Model</th>
<th>Lon [deg]</th>
<th>Drift [km]</th>
<th>$\Delta \alpha_p$ [m/s$^2$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRTF pr.</td>
<td>167.4 E</td>
<td>306.8 E</td>
<td>2.98</td>
</tr>
<tr>
<td>HRTF ret.</td>
<td>153.6 E</td>
<td>306.2 W</td>
<td>4.41</td>
</tr>
<tr>
<td>Aerosp. pr.</td>
<td>170.3 E</td>
<td>434.3 E</td>
<td></td>
</tr>
<tr>
<td>Aerosp. ret.</td>
<td>150.7 E</td>
<td>433.1 W</td>
<td></td>
</tr>
<tr>
<td>EWE pro.</td>
<td>170.2 E</td>
<td>429.9 E</td>
<td>2.32</td>
</tr>
<tr>
<td>EWE ret.</td>
<td>150.8 E</td>
<td>429.2 W</td>
<td></td>
</tr>
</tbody>
</table>

or West depending on wind direction) using the Aerospatiale model as opposed to the HRTF model.

6. THE PROBE - ORBITER COMMUNICATION LINK

During its descent phase the Probe will establish a one-way radio link to the Orbiter via two redundant transmitters on the Probe and the Cassini high gain antenna (HGA) (Jones & Giovagnoli 1997). In the pre-HRTF baseline mission, the Orbiter was targeted to pass Titan on the prograde side at a flyby altitude of 1228 km (cp. Huygens Recovery Task Force 2001). In this scenario, lateral Probe drift due to wind had to be taken into account, because of its potential impact on the Probe-Orbiter link. Strong winds could cause a significant miss pointing of the narrow-beam Orbiter HGA, particularly during the later stages of the descent and after landing. Without knowledge of the zonal wind direction, it was decided that the HGA should point to the landing coordinates of the Probe calculated without accounting for any wind.

In the recovered Huygens mission, the Orbiter passes on the retrograde side of Titan at a periapsis distance of 60,000 to 65,000 km (Huygens Recovery Task Force 2001; Strange 2001). This new link geometry reduces significantly the received signal sensitivity to HGA off-pointing, but implies a steadily increasing Probe Aspect Angle (PAA), i.e., the angle between the $-X_p$ axis in the Probe reference frame (nominally vertical on Titan) and...
Our simulations resulted in a dispersion ellipse with a semi-major axis of ~114.6 km and a semi-minor axis of ~10.9 km (1σ values) for the baseline mission (i.e., HRTF prograde wind). Probe delivery errors and Orbiter navigation errors were estimated by the Cassini navigation team. For the Aerospatiale and the EWE wind model these values did not change much for the Aerospatiale and EWE wind models, clearly indicating that only the assumed errors, but not the specific models, affect the size of the dispersion ellipse.

The received telemetry symbol rate onboard Cassini will differ from the nominal transmitted rate $f_0$ of the Probe during its descent due to the Doppler effect caused by the relative velocity $v$ (range rate) between the Probe and Orbiter and the bias and drift of the clock in the Probe which generates the telemetry stream. The received telemetry symbol rate $f_r$ is therefore given by

$$f_r = f_0 + \Delta f_{\text{Doppler}} + \Delta f_{\text{clock}}$$

(10)

As determined during a receiver test using a simulated Probe signal generated on Earth, the specially designed Huygens receiver has a much too narrow bandwidth for its symbol-synchronizer whenever the Doppler shift becomes large. This design flaw was discovered very late (but not too late), because all previous testing of telemetry decoding had been performed when $\Delta f_{\text{Doppler}} = 0$. For a given signal-to-noise ratio $E_s/N_0$ and a given transition probability $P_t$ in the data stream, it was determined that unacceptable data loss occurs whenever the frequency offset ratio, defined as $(\Delta f_{\text{Doppler}} + \Delta f_{\text{clock}})/f_0$ exceeds a certain limit. This limit is shown in Fig. 13 as a sawtooth curve. The sawtooth effect comes from another uncorrectable feature in the receiver which actually further narrows the telemetry bandwidth as the received signal strength increases. The sawtooth curves (for various values of $P_t$) were calculated using an analytical model that simulates the symbol timing recovery of the Data Transition Tracking Loop (DTTL) of the receiver onboard Cassini (L. Popken 2001). The solid and dashed lines represent the frequency offsets and corresponding signal-to-noise ratios for the minimum and maximum Probe transmitting antenna gain values during the baseline descent trajectory taking into account the various uncertainties in the two extreme combinations (statistical 1 and 99 percentiles of a 10,000 shot Monte Carlo simulation). If these curves cross into the region above the black sawtooth curves in the upper part of the figure, the receiver will have problems locking on the data stream and begin to lose frames.

Fig. 13 depicts the link from start of transmission until the end of the nominal mission for a simulation assuming the current baseline HRTF prograde wind (solid lines) and the stronger Aerospatiale wind (dashed lines). The stronger wind clearly weakens the signal by ~1.7 dB at the end of the nominal mission (~16 minutes after nominal touchdown) for the 1% line (left line). With the given frequency offset at that point one can readily see that the receiver would not be able to maintain the data link without any frame losses due to cycle slips in the DTTL loop. For the case with the HRTF-wind the signal remains strong enough to avoid this fate. The frequencies shown in the example of Fig. 11 were calculated
assuming that the Probe is powered about 4 hours prior to entry. This “pre-heating” would partially compensate the frequency offset $\Delta f_{\text{lock}}$, a strategy beneficial mainly at the beginning of the mission (presently part of the current baseline scenario).

7. CONCLUSIONS AND RECOMMENDATIONS

A recently discovered design flaw in the Huygens receiver onboard Cassini led to a significantly different mission geometry designed and implemented by the Huygens and Cassini project teams. In this new scenario the Orbiter will pass Titan at high altitude (65,000 - 60,000 km) on the retrograde side and trail the Probe by only about 2.1 hours instead of the planned 4 hours after its separation in December 2004 (Huygens Recovery Task Force 2001; Strange 2001). Due to this new geometry the communication window between Orbiter and Probe strongly depends on the zonal wind strength and its direction.

The only observations presently available for a determination of wind velocities are the infrared brightness temperature measurements from the IRIS instrument onboard Voyager 1, from which equator-to-pole temperature gradients were inferred and the associated zonal wind speed profile was derived for mid latitudes (Flasar et al. 1981). Certain physical assumptions (e.g., constant angular momentum; constant angular speed) were used to extend the wind model to lower and higher latitudes and to develop engineering models for future mission analysis and planning (Lunine et al. 1991; Flasar et al. 1997). We have used these models in a simulation of the Probe entry and descent down to Titan’s surface and found that the more conservative model (i.e., the Aerospatiale model) produces a Probe drift of $\sim 127$ km further than the model proposed as a baseline within the framework of the Huygens Recovery Task Force (Lebreton 2001). Due to the large value of the Probe Aspect Angle (PAA > 70°) and associated steep gain decrease of the Probe transmitting antennas at the end of the mission, a drift of $\sim 127$ km would lower the signal strength by as much as 1.7 dB, thereby possibly endangering maintenance of the communication link.

We could also confirm sensitivity of the Probe deceleration to the wind direction during its atmospheric entry phase. Even if the effect is considerably small and is to first order negligible for mission analysis purposes, it will introduce an error in the density profile of the upper atmosphere as this is going to be inferred from Probe accelerometer data measured during the entry phase.

Finally, we encourage any Earth-based observations that could increase the reliability of the current engineering models used for our ongoing key parameter optimization of the recovered Huygens mission. Meanwhile, we recommend using the HRTF prograde wind as an optimistic case and the Aerospatiale model as the conservative counterpart.

REFERENCES

Courtin R., 2001, private communication
TECTORIC ACTIVITY OF THE PLANETS AND SATELLITES: MECHANISM AND NATURE OF CYCLICITY

Yu.V. Barkin

(Sternberg Astronomical Institute, Universitetskii pr-4, 13, Moscow, 119899, Russia; Email: barkin@sai.msu.ru)

ABSTRACT/RESUME

On the base of new conception of shell-dynamics of the planets and satellites the qualitative evaluations of their inner activity were obtained in good agreement with planetology date and with date about natural processes of the Earth.

1. CONCEPTION

Gravitational interaction of the Moon and Sun with non-spherical, non-homogeneous shells of the Earth generates very big additional mechanical forces (and moments) of the interaction of the neighboring shells (rigid core, liquid core, mantle and its layers, lithosphere and separate plates and others). Action of these forces in the different time scales on the corresponding shells generates cyclic perturbations of the tensional state of the shells, their deformations, small relative translational displacements and slow rotation of the shells, formation of the planetary crack system, redistribution of the plastic and fluid masses, the planetary redistribution of the liquids and others. In geological period of time it leads to a fundamental tectonic reconstruction of the Earth.

In accordance with our conception global processes of the planets and satellites, time evolution and cyclicity have celestial mechanical nature and in first are caused by relative translational and rotational displacements of their shells. In given report we analyze the force interaction between neighboring non-spherical shells cause by different gravitational influence of the external celestial bodies, energetic budgets of these gravitational interactions and their possible component for inner (endogenous) processes. In given paper I consider some upper evaluations of above-mentioned energetic characteristics in assumption that all energy of gravitational interaction of the bodies caused by their non-sphericity goes on the inner process.

Due to celestial-mechanical nature the inner additional forces between shells are cyclic and change in wide diapason of frequencies (with periods from hours to millions years). These forces produce deformations of the all layers of the body and organize and control practically all natural processes. The analytical expressions of the components of these forces in the inertial and in body reference system were obtained and their structure was studied.

From the more important mechanisms of endogenous activity of celestial body (in particular of the Earth) we point out:
- small relative rotation (nutation) of the shells in different time-scales;
- small relative translational motions of the shells (displacements of their center of mass);
- relative displacements and rotations of the shells due to eccentricity of their center of mass positions.

The first and second mechanisms are caused by different non-sphericities of the shells (by differential interactions of the shells with external attracting body). Third mechanism is defined by eccentricity in the relative positions of the shell centers of mass. Of course for real celestial bodies all these mechanisms work with different contributions into inner mechanical processes [1]. In given paper we will concentrate our attention on the analysis of the force interaction between non-spherical shells caused by attraction of the external body.

2. SHELL INTERACTIONS

Due to different oblatenesses of the shells a external body produces different accelerations of their centers of mass. Giant tensions appear in result between neighboring shells and they deform all layers of the shells. This effect has place even in the case when centers of mass of the shells coincide. For planets and satellites similar forces are appeared between the core, mantle, external shell and their non-homogeneous layers. For the Earth for example the giant forces are appeared between core and mantle, between shells with boundary at 670 km, between lithosphere and separate plates and others. So additional cyclic force of mutual interaction of the liquid core and mantle (caused by the Moon attraction) is about $10^6$ from the full value of the force of the gravitational attraction between the Earth and the Moon. This force in a few orders bigger then classical tide force in the Earth-Moon system. Of course this force is cyclic and characterized by very wide basis of frequencies typical for orbital motions (of the Sun, Moon, planets and Solar system in Galaxy), for rotational motion of the Earth, Moon and Sun and for many from observed natural processes.

To understand the role of the inner interaction we will consider simplest model of celestial body, consisting


© European Space Agency • Provided by the NASA Astrophysics Data System
from two axisymmetric shells (TSB). Inner shell ($S_1$) and external shell ($S_2$) can execute small relative translatory-rotatory displacements due to thin elastic layer between them. In unperturbed motion of celestial body axes of its shells are coincided. The shells are subjected by different gravitational action from every external celestial body. And as result small relative displacements of the shells and wide number of their geodynamical consequences must be observed.

2.1 Force between shells

Let us to consider gravitational interaction of the shells $S_1, S_2$ with the external celestial body $S$, which moves on the unperturbed (or perturbed) Keplerian elliptic orbit around TSB. Body $S$ (satellite) we will consider as material point. In unperturbed motion the TSB rotates as rigid body relatively of axis of symmetry $O \zeta$. $r$ is a module of radius-vector $\mathbf{r}$ of the satellite $S$ with respect to the center of mass of the planet TSB. $\gamma$ is an angle between radius-vector $\mathbf{r}$ and polar axis of inertia of satellite $O \zeta$. This model can be used also for analysis of the shell interaction of the planet (TSB) and the Sun (material point $S$), see p. 3.

From differential equations of the relative translatory-rotatory motion of the shells it follows that a module of additional force acting to the external shell from the side of inner shell is characterized by formula:

$$ F(r, \gamma) = k \sqrt{1 - 2\gamma^2 + 5 \gamma^4} / r^4 $$  \hspace{1cm} (1)

where $k$ is a constant parameter depending from the mass of the shells $m$ and from their moments of inertia ($C_i$ is a polar moment of inertia, $A_i$ is a equatorial moment of inertia, $i = 1, 2$).

$$ k = \frac{3}{2} f \left[ C_2 - A_2 - \frac{m_2}{m_1} (C_1 - A_1) \right] $$  \hspace{1cm} (2)

$f$ is a gravitational constant. For upper evaluations of the forces and powers of the inner interaction of the shells we will use some model value of parameter (2):

$$ F_0 = \frac{3}{2} \mathcal{F} J_2 \times m R^2, $$  \hspace{1cm} (3)

where $J_2$ is a coefficient of second harmonic of TSP potential, $m = m_1 + m_2$ and $R$ is a full mass and mean radius of this planet. Formula (3) will be correct for example in the case of the small spherical core of TSP.

Obviously in dependence from character of the satellite orbit the force $F(r, \gamma)$ is a cyclic function of time. Even in the case of circular orbit this force is subjected by big variations. In reality this direction cosine is changed in diapason $\gamma \in (-\sin \rho, \sin \rho)$, where $\rho$ is an angle between the axis of rotation of the TSB and normal to the plane of satellite orbit. Orientation of the force in the TSB axes (latitude $\varphi_F$) depends only from $\gamma$:

$$ \sin \varphi_F = \frac{\gamma (3 - 5 \gamma^2)}{\sqrt{1 - 2 \gamma^2 + 5 \gamma^4}}. $$  \hspace{1cm} (4)

Force characteristics (1), (4) are very important. They define direction and intensity of the action of the inner shell to the external, control processes of variations of the tension state of interacting shells and dictate variations of the many natural processes into shells and on the surface of the TSB.

2.2 Shell interactions and inclinations of the axes of the planets

In the case of circular orbit of the planet a minimal possible value of the force $F(r, \gamma)$ is achieved by $\gamma = \pm 1/\sqrt{3}$. Corresponding inclination of the axis of planet rotation will be $26^\circ 6$. By more smaller inclination the force (1) will lightly increase and by $\gamma = 0$ ($\rho = 0$) will be about $112\%$ from minimal value. Practically it means that for all diapason of the values $\rho \in (0^\circ - 39^\circ)$ the force of interaction of the planet shells will be minimal $F(r, \gamma) \leq F_0 = k / r^4$ (maximal value is $2 F_0$ by $\gamma = \pm 1$).

It is very interesting to note that for planets of the solar system condition of the minimal interaction of their shells is fulfilled. This condition is not satisfied only for Uran and for double external planet Pluto-Charlo. Obtained result can be considered as confirmation of the important dynamical evolutionary role of the discussed mechanism of shell-dynamics for the planets and satellites. Probably the shells use special inclined orientations of own axes of rotation for minimization of their mutual interaction. In another words observed inclinations of the planet axes of rotation correspond to more quiet joint life of the shells. But for full decision of considered problem the new treatments of the mechanical problems about relative oscillations of the planet shells and their evolution must be considered.
Here we will not consider more detailed evaluations of the forces of shell interaction for the planets and satellites but in p. 3.2 similar evaluations will be obtained for powers of the inner interactions of the shells.

3. PLANET ENDOGENOUS ACTIVITY

3.1 Power of the endogenous processes

In accordance with our approach definite part of the potential energy stored in the gravitational interaction of the shell and described by the second harmonic of the gravitational potential of the non-spherical planet and the Sun transformed to the endogenous energy of the deformable planet due to action of mentioned mechanism of the relative displacements of the planet shells. For upper evaluations of the inner planet energy we will consider full expression of the second harmonic of the force function of the system «Sun-planet». For simplicity we will consider axisymmetric planet with principal moments of inertia \( A_p \) and \( C_p \) (polar):

\[
U_p = \frac{f m_{\text{Sun}}}{2r_p^3} \cdot (C_p - A_p) \cdot (1 - 3\gamma^2) \tag{5}
\]

Here \( f \) is a gravitational constant, \( m_{\text{Sun}} \) is a mass of the Sun, \( r_p \) is a module of the radius vector \( \vec{r}_p \) of the center of mass of the planet in heliocentric reference system, \( \gamma \) is a cosine of the angle between the polar axis of inertia of the planet and vector \( \vec{r}_p \). Let us assume that orientation of the polar axis of the planet are given by the constant Euler angle of precession \( \Psi = \Psi_o \) and nutation \( \Theta = \rho \).

Power of the gravitational action on the non-spherical planet we define as time derivative of the potential (5)

\[
W = \dot{U}_p = \frac{3f m_{\text{Sun}}}{2r_p^4} \cdot (C_p - A_p) \times
\]
\[
\times [r_p (1 - 3\gamma^2) + 2r_p \gamma \dot{\gamma}] \tag{6}
\]

Here we will be restricted by the case of the elliptic unperturbed orbit. In this case we will have \( \gamma = \sin \Theta \sin(\Psi - \nu_p) \). Where \( \nu_p \) is a true anomaly of the Keplerian motion. Assuming also that the axis of the planet is fixed in the space (the angle \( \Psi \) and \( \Theta \) are constant) we obtain:

\[
\dot{r}_p = \frac{2\pi}{T_p} \cdot \frac{e_p a_p}{\sqrt{1 - e_p^2}} \sin \nu_p, \tag{7}
\]

\[
\gamma = -\sin \Theta \cos(\Psi - \nu_p) \dot{\nu}_p, \tag{8}
\]

\[
\dot{\nu}_p = \frac{2\pi}{T_p} \cdot \frac{1}{(1 - e_p^2)^{3/2}} (1 + e_p \cos \nu_p)^2 \tag{9}
\]

Here \( a_p \) and \( e_p \) is a major semi-axis of the planet orbit and eccentricity, \( T_p \) is a period of the orbital motion of the planet. If unperturbed planet orbit is a circular (by \( e_p = 0 \)), so we have main part of the power in the following form:

\[
W_p = \dot{U}_p = \frac{3(2\pi)^3}{2T_p^3} \cdot (C_p - A_p) \times
\]
\[
\times \sin^2 \rho \sin 2(\Psi - M_p) \tag{10}
\]

Power (10) is a definite function of time with maximum by \( \Psi - M_p = \pm \pi / 4, \pm 3\pi / 4 \). It means that the maximum of the power of gravitational action on the planet shells has place before \( T_p / 8 \) and \( 3T_p / 8 \) days and after intersection of the Sun of the planet equator. Pointed moments must be considered as a days of the potential activation of the planet endogenous activity. In the table 1 the evaluations of the power of this activity of the planets are given. They were obtained as upper evaluations of this characteristic on the base formula:

\[
\overline{W}_p = 0.372 \cdot 10^3 \cdot (J_2)_p \cdot \frac{m_p R_p^2}{T_p^3} \sin^2 \rho \tag{11}
\]

where \( (J_2)_p \) is a coefficient of second harmonic of the planet (TSP) potential.

Another part of the power (6)-(9) is caused by the eccentricity effect. In the case when the axis of the planet rotation is orthogonal to the orbit plane the extreme value of considered component of power is defined as

\[
\overline{W}_e = 0.372 \cdot 10^3 \cdot (J_2)_p \cdot \frac{m_p R_p^2}{T_p^3} \cdot e_p \tag{12}
\]

This formula was used for evaluation of the Mercury activity.

Table 1. Tectonic activity of the planets due to Sun stimulation \( \overline{W}_{p,\text{Sun}} \).
<table>
<thead>
<tr>
<th>Planet</th>
<th>((J_2)_P)</th>
<th>(m_P) (\times 10^{26}) kg</th>
<th>(T_s) (day)</th>
<th>(W_{P\text{Sun}}) (Wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>0.600 (\times 10^{-3})</td>
<td>0.3302</td>
<td>87.97</td>
<td>2.06 (\times 10^{11})</td>
</tr>
<tr>
<td>Venus</td>
<td>0.597 (\times 10^{-2})</td>
<td>4.8685</td>
<td>224.7</td>
<td>1.15 (\times 10^{11})</td>
</tr>
<tr>
<td>Earth</td>
<td>1.083 (\times 10^{-2})</td>
<td>5.9736</td>
<td>365.2</td>
<td>4.94 (\times 10^{11})</td>
</tr>
<tr>
<td>Mars</td>
<td>1.959 (\times 10^{-1})</td>
<td>0.6419</td>
<td>686.9</td>
<td>4.68 (\times 10^{12})</td>
</tr>
<tr>
<td>Jupiter</td>
<td>1.470 (\times 10^{-2})</td>
<td>1898.6</td>
<td>4330.6</td>
<td>3.02 (\times 10^{12})</td>
</tr>
<tr>
<td>Saturn</td>
<td>1.631 (\times 10^{-3})</td>
<td>568.46</td>
<td>10746.9</td>
<td>2.87 (\times 10^{12})</td>
</tr>
<tr>
<td>Uran</td>
<td>3.513 (\times 10^{-2})</td>
<td>86.831</td>
<td>30588.7</td>
<td>3.94 (\times 10^{12})</td>
</tr>
<tr>
<td>Neptune</td>
<td>3.539 (\times 10^{-2})</td>
<td>102.43</td>
<td>59799.9</td>
<td>1.35 (\times 10^{11})</td>
</tr>
<tr>
<td>Pluto</td>
<td>1.300 (\times 10^{-4})</td>
<td>0.0124</td>
<td>90739.0</td>
<td>1.19 (\times 10^{10})</td>
</tr>
</tbody>
</table>

In Table 1 also are given values of the planet parameters: mass \((m_P)\), orbit period \((T_P)\), coefficient of the second zonal harmonic of planet potential \((J_2)\).

### 3.2 Stimulation of the planet endogenous activity by its own satellites

Massive satellites also very actively act on the shells of their planet and produce additional (sometimes very high) tectonic activity. In first it is concerned the following pairs of celestial bodies: Earth-Moon, Neptun-Triton, Pluto-Charo.

In similar way we can obtain expressions for power of the gravitational influence on the satellites on their mother planet. Considering the planet as system of the axisymmetric shells (TSB) for a power we will obtain evaluations similar to (11), (12):

\[
\overline{W}_\rho = 0.372 \cdot 10^{-3} \cdot (J_2)_P \cdot \frac{m_s R_p^2}{T_s^3} \sin^2 \rho, \quad (13)
\]

\[
\overline{W}_\epsilon = 0.372 \cdot 10^{-3} \cdot (J_2)_P \cdot \frac{m_s R_p^2}{T_s^3} \cdot e_s. \quad (14)
\]

Here \(m_s\) is a mass of satellite, \(e_p\) is an eccentricity of satellite orbit, \(T_s\) is a period of the orbital motion of the satellite.

Formula (14) were used for evaluations of the stimulation powers of the planets by their own synchronous satellites (see Table 2). For the Moon both formulae (13), (14) were used for evaluations of the eccentricity and inclination component of Earth stimulated powers.

From results of the Table 2 it follows that satellite stimulation of the planets plays more important role than the Sun stimulation. The table 2 presents very high (upper) evaluations of the endogenous energy of planets. Definite part of these energies in reality goes to the inner planet processes. These results explain the more high tectonic activity of the Earth comparatively with the Mars and Venus activity and more high Neptun endogenous activity then Uran activity [3], [4] (see Table 3).

### Table 2. Power of planet stimulation by own satellites.

<table>
<thead>
<tr>
<th>Planet</th>
<th>Power (Wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUPITER</td>
<td>3.14 (\times 10^{11})</td>
</tr>
<tr>
<td>Io stimulation</td>
<td>2.80 (\times 10^{11})</td>
</tr>
<tr>
<td>Europa stimulation</td>
<td>2.98 (\times 10^{10})</td>
</tr>
<tr>
<td>Ganymede stimulation</td>
<td>3.50 (\times 10^{10})</td>
</tr>
<tr>
<td>Amalthea stimulation</td>
<td>7.54 (\times 10^{10})</td>
</tr>
<tr>
<td>Callisto stimulation</td>
<td>7.05 (\times 10^{9})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>3.02 (\times 10^{9})</td>
</tr>
<tr>
<td>SATURN</td>
<td>7.59 (\times 10^{10})</td>
</tr>
<tr>
<td>Titan stimulation</td>
<td>3.45 (\times 10^{10})</td>
</tr>
<tr>
<td>Mimas stimulation</td>
<td>3.14 (\times 10^{9})</td>
</tr>
<tr>
<td>Enceladus stimulation</td>
<td>0.50 (\times 10^{9})</td>
</tr>
<tr>
<td>Diona stimulation</td>
<td>0.39 (\times 10^{9})</td>
</tr>
<tr>
<td>Tethya stimulation</td>
<td>0.14 (\times 10^{9})</td>
</tr>
<tr>
<td>Rhea stimulation</td>
<td>0.09 (\times 10^{9})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>2.87 (\times 10^{9})</td>
</tr>
<tr>
<td>NEPTUN</td>
<td>1.95 (\times 10^{10})</td>
</tr>
<tr>
<td>Triton stimulation</td>
<td>1.93 (\times 10^{10})</td>
</tr>
<tr>
<td>Proteus stimulation</td>
<td>0.14 (\times 10^{9})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>1.35 (\times 10^{9})</td>
</tr>
<tr>
<td>URAN</td>
<td>1.30 (\times 10^{9})</td>
</tr>
<tr>
<td>Titania stimulation</td>
<td>0.80 (\times 10^{9})</td>
</tr>
<tr>
<td>Ariel stimulation</td>
<td>0.36 (\times 10^{9})</td>
</tr>
<tr>
<td>Umbriel stimulation</td>
<td>0.12 (\times 10^{9})</td>
</tr>
<tr>
<td>Miranda stimulation</td>
<td>0.02 (\times 10^{9})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>3.94 (\times 10^{9})</td>
</tr>
<tr>
<td>EARTH</td>
<td>1.45 (\times 10^{9})</td>
</tr>
<tr>
<td>Moon stimulation ((e))</td>
<td>0.50 (\times 10^{9})</td>
</tr>
<tr>
<td>Moon stimulation ((\rho))</td>
<td>0.90 (\times 10^{9})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>4.94 (\times 10^{9})</td>
</tr>
<tr>
<td>MARS</td>
<td>3.93 (\times 10^{9})</td>
</tr>
<tr>
<td>Phobos stimulation</td>
<td>3.46 (\times 10^{10})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>0.47 (\times 10^{13})</td>
</tr>
<tr>
<td>PLUTO</td>
<td>0.90 (\times 10^{7})</td>
</tr>
<tr>
<td>Charo stimulation</td>
<td>0.90 (\times 10^{7})</td>
</tr>
<tr>
<td>Sun stimulation</td>
<td>1.19 (\times 10^{7})</td>
</tr>
</tbody>
</table>

### Table 3. Relations of activity of the planets.

<table>
<thead>
<tr>
<th>Observed / theoretical relations of planet tectonic activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth activity &gt; Venus activity (1.45 \times 10^{10} \text{ Wt} &gt; 1.15 \times 10^{11} \text{ Wt})</td>
</tr>
<tr>
<td>Earth activity &gt; Mars activity (1.45 \times 10^{10} \text{ Wt} &gt; 3.93 \times 10^{9} \text{ Wt})</td>
</tr>
<tr>
<td>Mars activity &gt; Venus activity (3.93 \times 10^{9} \text{ Wt} &gt; 1.15 \times 10^{11} \text{ Wt})</td>
</tr>
<tr>
<td>Neptun activity &gt; Uran activity (1.95 \times 10^{10} \text{ Wt} &gt; 1.30 \times 10^{9} \text{ Wt})</td>
</tr>
</tbody>
</table>
4. MECHANISM OF THE ENDOGENOUS ACTIVITY OF THE EARTH

4.1 Source of energy

Relative translatory-rotary oscillations of the Earth shells due to differential gravitational influence of the external celestial bodies (the Moon, the Sun and others) is a main mechanism of the endogenous activity and in particular is a main mechanism of the ordering of positions of the geological formation centers [1].

4.2 Power budget of the Earth natural processes

Gravitational energy of the interaction of the Earth with the Moon and the Sun (caused by non-sphericity of the Earth) is evaluated as \( U_{\text{moon+Sun}} = 0.85 \times 10^{27} \) Erg. This energy is transformed to the mechanical energy of translatory-rotary motion of the Earth-Moon-Sun system (effects of non-sphericity of the Earth) and to elastic and warm energy of the intermediate layer between shells. In general case of course all layers of the shells are subjected to the similar changes. Remarkable part of energy \( U_{\text{moon+Sun}} \) goes to the wide list of inner process including tectonic processes, redistribution of the liquids, cracks formation processes, warm flows and to many others. For energy of above mentioned process we suggest as high evaluation \( U_{\text{endogenous}} = 10^{28} \) Erg.

In pp. 3.1, 3.2 the upper evaluations of the powers of the endogenous activity of the Earth due to Sun and Moon perturbations of the Earth shells were obtained. These values are given in the Table 4 in comparison with powers of some natural processes of the Earth [2].

Table 4. Earth power budget.

<table>
<thead>
<tr>
<th>Moon and Sun influence</th>
<th>( W_{\text{moon+Sun}} = 1.45 \times 10^{26} \text{ Wt} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous processes</td>
<td>( W_{\text{end}} = 10^{14} \text{ Wt} )</td>
</tr>
<tr>
<td>Warm convection</td>
<td>( W_{\text{convection}} = 10^{15} \text{ Wt} )</td>
</tr>
<tr>
<td>Tide processes</td>
<td>( W_{\text{tide}} = 4 \times 10^{11} \text{ Wt} )</td>
</tr>
<tr>
<td>Plate tectonics</td>
<td>( W_{\text{plate}} = 10^{11} \text{ Wt} )</td>
</tr>
<tr>
<td>Seismic processes</td>
<td>( W_{\text{earthquake}} = 3 \times 10^{10} \text{ Wt} )</td>
</tr>
<tr>
<td>Volcanic processes</td>
<td>( W_{\text{volcano}} = 10^{10} \text{ Wt} )</td>
</tr>
</tbody>
</table>

From presented results it follows that considered mechanism of the shell-dynamic can provide by energy for all natural processes on the Earth.

5. GEODYNAMICAL CONSEQUENCES

Force variations are reflected in variations of the tensional states of the body shells in different timescales from hours to hundred millions years. Above mentioned shell interactions lead to the considerable variations of tensional state of the planet, to its planetary geological and geophysical reconstruction's and processes characterized by the properties of cyclicity, polarity, asymmetry, inversion and others. These fundamental properties of the planet and satellite life can be explained as consequences of the shell oscillations in definite geological periods of time. The Galaxy cycles play here the main role. They are caused by the perturbed orbital motion of the Solar system in Galaxy and by Galaxy gravitational attraction. Radial character of the shell displacements and their non-permanent drift are the dynamical reasons of the phenomena of asymmetry and inversion of the geodynamical states. On the basis of the developed approach a mechanical interpretation can be given for following observed phenomena and processes: 1. Cyclicity of the natural processes on the planets and satellites. 2. Galaxy cycles of geoevolution; 3. Origin and existence of the latitudinal and longitudinal ordering of the planet and satellite formations (grid phenomenon). 4. Correlation of positions of the biggest formations of the planets (satellites) and orientation of their axes of rotation with peculiarities of the Solar system trajectory in Galaxy. All these phenomena are by reflections of the small and slow relative displacements of the shells in geological periods of time.

6. NATURAL CYCLES

Relative oscillations of the two shells of the planet under action of the gravitating external bodies were studied on the base some simple model problems. We have suggested that shells are rigid axisymmetric bodies separated by very thin elastic or inelastic layer. Due to this layer the sells can execute small relative translational and rotational motions. They are subjected by definite additional reactions from intermediate layer. In first we take into account elastic forces acting from the layer to the shells.

It was shown that orbital perturbations of the external celestial bodies (the Moon, the Sun and planets) are reflected directly in the relative displacements of the Earth shells. Hierarchic role of the orbital perturbations...
of the Moon, the Sun and planets in the shell dynamics was established. Relative oscillations of the shells are characterized: by the short periodic perturbations with periods comparative with orbital periods of the planets in tens and hundred years (P); by the secular perturbations (S) with periods of the secular orbital motions of the solar system in tens and hundred thousand of years; by galactic perturbations with periods in tens and hundred millions of years (G). All pointed orbital perturbations give similar class of relative oscillations of the Earth shells. Different combinations of these perturbations generate new class shell oscillations with intermediate periods. Amplitude of the perturbations increase in the consequence P - S - G. Studies of the shell oscillations show the increasing of the geodynamic role of the mentioned orbital perturbations that is confirmed by the materials of the observations and geological data.

Galaxy influence on the shells leads to the more remarkable variations of their tensional states, to planetary tectonic, geological and geography reconstruction’s of the Earth which are characterized by the properties of cyclicity, polarity, asymmetry, inversion and others [1].

The main properties of the considered shell displacements are: 1. ordering, cyclicity and directricity of their motions (the main cycles of the shell oscillations are characterized by periods of the perturbed galactic motion of the Sun); 2. radial character of displacements leading to antisymmetry phenomenon in the change of the geodynamical states (in the opposite hemispheres of the Earth); 3. non-uniform displacements of the shells, leading to the saw-forming variations of the activity of the geodynamical and geophysical processes.

8. TECTONIC ACTIVITY OF THE SATELLITES

As in p.4 we will assume that definite part of the potential energy stored in the second harmonic of the gravitational potential of the non-spherical satellite and its central planet is transformed to the endogenous energy of the satellite due to action of mentioned mechanism of the relative displacements of the satellite shells. For upper evaluations of the inner planet energy we have used full expression of the second harmonic of the force function of the system «satellite-planet» and expression of its time derivative. For simplicity we will consider plane motion of non-spherical satellite in the gravitational field of the mother planet on Keplerian elliptic orbit. Satellite axis of rotation is remained orthogonal to the orbit plane. Main satellites of the planets are synchronous and their motion are similar to the Moon resonant rotation.

Let us introduce well known coefficients of the gravitational potential

\[(J_{2})_{S} = \frac{2C_{S} - A_{S} - B_{S}}{2m_{S}R_{S}^{2}}, \quad (C_{22})_{S} = \frac{B_{S} - A_{S}}{4m_{S}R_{S}^{2}}\]

where \(A_{S}, B_{S}\) and \(C_{S}\) are the principal moment of inertia of satellite \((C_{S} > B_{S} > A_{S}).\)

In similar manner to p.4 we have been obtained the following simple formula for upper evaluation of the power of the gravitational action of the planet to the satellite:

\[
\bar{W}_{e} = 0.372 \cdot 10^{3} \cdot J_{S} \cdot \frac{m_{S}R_{S}^{2}}{T_{S}^{3}} \cdot e_{S} \quad (15)
\]

Here \(T_{S}\) is a period of the orbital motion of the satellite, \(J_{S} = (J_{2})_{S} + 6(C_{22})_{S}\) is a reduced gravitational coefficient of the satellite. Formula (15) is a native to (13), (14) and describe the effect of the influence of the orbit eccentricity of satellite on the inner behavior of its shells. Evaluations of the powers of the endogenous processes on the satellites are given in the Table 5 for present epoch. For calculations we have used known date about moments of inertia of the satellites from the paper [5].

Table 5. Satellite endogenerous activity (in Wt).

<table>
<thead>
<tr>
<th>N</th>
<th>Satellite</th>
<th>(W_{e}) (Wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Io</td>
<td>8.25 \times 10^{7}</td>
</tr>
<tr>
<td>2</td>
<td>Europe</td>
<td>2.32 \times 10^{16}</td>
</tr>
<tr>
<td>3</td>
<td>Titan</td>
<td>4.52 \times 10^{14}</td>
</tr>
<tr>
<td>4</td>
<td>Ganymede</td>
<td>2.08 \times 10^{18}</td>
</tr>
<tr>
<td>5</td>
<td>Mimas</td>
<td>1.36 \times 10^{8}</td>
</tr>
<tr>
<td>6</td>
<td>Miranda</td>
<td>4.24 \times 10^{4}</td>
</tr>
<tr>
<td>7</td>
<td>Ariel</td>
<td>3.56 \times 10^{4}</td>
</tr>
<tr>
<td>8</td>
<td>Enceladus</td>
<td>1.43 \times 10^{8}</td>
</tr>
<tr>
<td>9</td>
<td>Diona</td>
<td>1.24 \times 10^{4}</td>
</tr>
<tr>
<td>10</td>
<td>Moon</td>
<td>1.18 \times 10^{5}</td>
</tr>
<tr>
<td>11</td>
<td>Callisto</td>
<td>0.93 \times 10^{4}</td>
</tr>
<tr>
<td>12</td>
<td>Umbriel</td>
<td>4.46 \times 10^{5}</td>
</tr>
<tr>
<td>13</td>
<td>Triton</td>
<td>2.27 \times 10^{3}</td>
</tr>
<tr>
<td>14</td>
<td>Amalthea</td>
<td>9.40 \times 10^{12}</td>
</tr>
<tr>
<td>15</td>
<td>Rhea</td>
<td>8.15 \times 10^{12}</td>
</tr>
<tr>
<td>16</td>
<td>Tephys</td>
<td>2.80 \times 10^{12}</td>
</tr>
<tr>
<td>17</td>
<td>Titania</td>
<td>2.40 \times 10^{12}</td>
</tr>
<tr>
<td>18</td>
<td>Oberon</td>
<td>7.95 \times 10^{16}</td>
</tr>
<tr>
<td>19</td>
<td>Phobos</td>
<td>4.82 \times 10^{16}</td>
</tr>
<tr>
<td>20</td>
<td>Iapetus</td>
<td>1.72 \times 10^{16}</td>
</tr>
<tr>
<td>21</td>
<td>Hyperion</td>
<td>1.62 \times 10^{4}</td>
</tr>
<tr>
<td>22</td>
<td>Deimos</td>
<td>0.58 \times 10^{7}</td>
</tr>
</tbody>
</table>

Obtained results are in good agreement with well
known planetology date about tectonic activity of the satellites of the big planets [3], [4]. In particular the following relations of satellite tectonic activity have been explained in terms of numerical power characteristics (Table 5).

Table 6. Relations of activity of satellites.

<table>
<thead>
<tr>
<th>Observed / theoretical relations of satellite tectonic activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ariel activity &gt; Umbriel activity</td>
</tr>
<tr>
<td>$3.56 \times 10^{14} \text{ Wt} &gt; 4.46 \times 10^{13} \text{Wt}$</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Diona activity &gt; Rhea activity</td>
</tr>
<tr>
<td>$1.24 \times 10^{14} \text{ Wt} &gt; 8.15 \times 10^{13} \text{Wt}$</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Titania activity &gt; Oberon activity</td>
</tr>
<tr>
<td>$2.40 \times 10^{13} \text{ Wt} &gt; 7.95 \times 10^{12} \text{Wt}$</td>
</tr>
</tbody>
</table>

In paper [3] authors discussed the questions of tectonic activity of the planets and satellites in their geological past on the base of the planetology date about surface formations and different planetary structures. The base of our study is a dynamical analysis of the shell interactions of the planets (satellites). Although for calculations we have used present values of the parameters of considered bodies our results are in good agreement with geological constructions [3]. But our results predict sufficiently high tectonic activity for Titan. Authors [3] also discuss possible high activity of Titan. But in their general scheme Titan occupies position for non-active celestial bodies. From another hand the comparatively no high activity of Triton follows from our study.

Obtained results confirm a reality of the mechanism of the relative swinging of the planet and satellite shells and its main role in tectonic activity of the all celestial bodies.

The work was accepted by RFBR grant N 02-05-64176 and Spain grant for Sabatico year at Alicante University (N SAB 2000-0235).

References


INTERIOR OF MARS FROM NONRIGID NUTATIONS

T. Van Hoolst¹, V. Dehant¹, and J.P. Barriot²

¹Royal Observatory of Belgium, Ringlaan 3, B-1180 Brussels, Belgium
²Observatoire Midi-Pyrénées, 14 av. E. Belin, 31400 Toulouse, France

ABSTRACT

The interior of Mars is presently not well constrained by observational data. Especially information on the core is difficult to obtain. The NEIGE experiment of the Netlander mission is an experiment that allows to derive direct information on the core of Mars. A brief description of this experiment is given. It is in particular shown that the measurement of nutations of Mars yields information on the planet’s core. The state of the core (liquid or solid) can be derived and the density and radius of the core can be better constrained. Also, information on the possible existence of a solid inner core can be derived from the NEIGE experiment.

Key words: Mars; interior; nutations.

1. INTRODUCTION

Our present knowledge of the internal structure of Mars is based on geochemical and geophysical data. The geochemical data are principally derived from the SNC meteorites, meteorites found on Earth that have their origin on Mars. The study of these meteorites has led to a model for the composition and mineralogy of Mars. It has, for example, been derived that the Martian core contains about 14 wt% of the light element sulfur, besides the heavier elements iron and, to a much lesser extent, nickel (Longhi et al. 1992). The main geophysical data are, besides mass and radius of the planet, the gravitational field (Lemoine et al. 2001, Yuan et al. 2001) and the precession rate (Folkner et al. 1997). On the basis of these data, and using arguments of similarity with the Earth and laboratory data, models for Mars’ interior structure have constructed (see, e.g., Sohl and Spohn 1997). The data are however not sufficient to permit a modeling as precise as that for the Earth, for which seismology has enabled to “look” inside.

Here, we are mainly interested in the core, for which many uncertainties remain: neither the radius is known with good precision, nor is the state of the core established with certainty. Given the importance of the core for a wide variety of phenomena (see Stevenson 2001), it then seems important to better constrain the core models of Mars. Therefore new experiments are needed that measure quantities directly related to the core.

One of the aims of the NetLander mission to Mars is to resolve questions regarding the physical state of the core and to more precisely determine its density, temperature, and radius (Harri et al. 1999). In this mission, with launch foreseen in 2007, four smalllanders will be installed on the surface of Mars. Two experiments are particularly suited to address these issues: the seismological experiment SEIS, with a seismometer extended to tidal periods (Lognonné et al. 2000), and the geodetical experiment NEIGE (Barriot et al. 2001). In this paper, we describe how the NEIGE experiment of the NetLander mission offers a unique possibility to obtain core data.

One of the main goals of the NEIGE experiment is to determine the orientation and rotation variations of Mars. This will be done by measuring Doppler shifts of radio links. The four landers will be in radio contact with an orbiter, the CNES orbiter 2007, and the orbiter will send and receive radio signals to and from the Earth. The orbiter-landers link uses two frequencies, UHF (400 MHz) and S-band (2 GHz), to be able to correct for the effects of the Martian atmosphere, whereas the orbiter-Earth link is in X-band (about 8 GHz). For both radio links, Doppler shifts will be measured, from which the motion of the landers can be derived in inertial space. Since the NetLanders are firmly fixed to the surface, their motion reflects the rotation of Mars. The variations in rotation can be described as precession and nutation, which is the motion of the rotation axis with respect to an inertial reference frame, as polar motion, which is the motion of the rotation axis with respect to a reference frame fixed to Mars, and as rotation rate variations, usually referred to as length-of-day variations. Polar motion and length-of-day variations can be used to constrain the sublimation/condensation of the polar caps and the global dynamics of the atmosphere (Defraigne et al. 2000, Van den Acker et al. 2002). Nutation and precession is important for the deep interior, and will be studied in detail here.

In Sect. 2, a brief description will be given of the uncertainties on the core of the present-day models of the interior structure of Mars. In Sect. 3, the nutations of Mars are investigated, and the relation to the core is explained. Conclusions are presented in Sect. 4.

2. INTERIOR STRUCTURE OF MARS

For planets and moons, the moment of inertia is the main parameter for the determination of the profile of density and thus of the global structure of the body. The polar moment of inertia is defined as

\[
C = \int_V d^2 \rho \, dV,
\]

(1)

where \( \rho \) is mass density, \( d \) is the distance to the polar axis, and the integration is over the whole planet. For Mars, it can be derived from the observation of the precession rate and the degree two gravitational field. The precession rate is proportional to \( J_2 C^{-1} \), where the form factor \( J_2 = (C - A)M^{-1}R^{-2} \) is a coefficient of the degree two gravitational field. \( M \) and \( R \) are the mass and radius of Mars, respectively, and \( A \) is the mean moment of inertia with respect to an equatorial axis. Folkner et al. (1997) determined the polar moment of inertia, from observations of the precession rate by using Viking Lander and Pathfinder data, to be \( C M^{-1}R^{-2} = 0.3662 \), with an error of only 0.5%. Because the moment of inertia mainly depends on the mantle, through the distance squared dependence in Eq. (1), a very accurate determination is needed to infer the radius and density of the core. With the above quoted result for \( C \), and using reasonable assumptions on the temperature profile, Folkner et al. (1997) estimated the core to be about \( 1500 \pm 200 \) km. Clearly, a better determination is highly wanted.

Thermal evolution models of Mars result in an entirely liquid core when a large enough concentration of sulfur is supposed to be present in the core (Stevenson et al. 1980). Corroborative evidence also comes from the magnetic field measurements (Acuña et al. 1999), in particular the absence of a global dipole magnetic field at present, and evidence for such a field for the young Mars. In this scenario then, Mars initially had a vigorously convective liquid core, which through dynamo action created a global magnetic field. After a few 100 million years, the planet had cooled enough to allow all heat to be transported by conduction in the core, and, as a result, the dynamo was shut down (Stevenson 2001), a situation which persisted until now. However, the question of whether the core is entirely liquid or solid (or whether there exists a solid inner core inside a liquid core) is still not definitely resolved. The composition of the core and the temperature of the core are not sufficiently known to definitely conclude that an inner core could not have formed. On the other hand, an inner core is usually linked to the existence of a planetary dynamo through the release of light elements and the onset of convection in such a case. However, this need not to be the case in general. The outer core convection could, for example, not have the right geometry, or could be too weak, for instance in the case of a slow inner core growth, or the outer core could be too thin to sustain a dynamo. The knowledge of dynamo action is too limited to make definite conclusions in this respect. Even a completely solid core, although highly unlikely on grounds of indirect arguments, can not be excluded. Direct information on the state of the core and on a possible inner core would thus be very desirable to settle these questions.

In our study of the nutations of Mars, we have considered several models for Mars' interior. We started from model A of Sohl and Spohn (1997) and constructed additional models, with the same mantle profile as model A, by changing the core radius and correspondingly the core density to keep the total mass of Mars fixed (Fig. 1). The models in Fig. 1 cover the range of uncertainty in core radius and are compatible with the value of the moment of inertia \( C \) of Folkner et al. (1997).

3. NUTATIONS OF MARS

Nutations of a planet are periodic changes of the orientation of the rotation axis (or figure axis) with respect to an inertial reference frame and are due to the gravitational torque acting on the equatorial bulge of the flattened planet. The torque can be computed from the tidal potential which, apart from the masses of the various bodies, essentially depends
on the relative position of the planet with respect to the Sun, the moon(s) and the other planets as given by ephemerides. To derive the nutational motions of the rotation axis (or of the figure axis) in inertial space, the planet is usually, in a first step, considered as a rigid body. The rigid nutations of Mars have been calculated with high precision (Bouquillon and Souchay 1999, Roosbeek 2000). These studies give the nutation frequencies, stemming from the relative motions of the planet and the perturbing celestial bodies, and the rigid body amplitudes with a precision of 0.1 mas (milliarcsecond), corresponding to a displacement of the rotation axis at the Martian surface of 1.6 mm. In a second step, the non-rigid effects on the nutations are calculated for each frequency of the rigid nutation series. This approach of first calculating the rigid body nutations is commonly used in Earth nutation studies since it allows to separate the celestial mechanics problem of determining the tidal potential from the physics of the planetary interior. Generally, the non-rigid body amplitudes do not differ more than a few percent from the rigid body amplitudes, such that the rigid nutations give a very good estimate of the characteristic nutation amplitudes. The main nutations of Mars are due to the gravitational attraction of the Sun, with periods equal to integer divisors of a Martian year. The largest nutation is the prograde (meaning in the same direction as the rotation of Mars: counterclockwise) semi-annual nutation with an amplitude of about 500 mas (Roosbeek 2000). Remark that annual is used here to denote one Martian year, i.e. about 687 terrestrial mean solar days (here simply called a “day”) of 86400s.

3.1. Core and Free Core Nutation

If Mars has a (partially) liquid core, the amplitudes of several nutations can differ substantially from their rigid body nutation amplitudes because of the existence of a particular rotational normal mode, the Free Core Nutation. This mode is a relative rotation of the instantaneous rotation axis of the liquid core with respect to that of the ellipsoidal, deformable mantle and has a retrograde long period in an inertial frame. Because it has a quasi-diurnal period in a frame tied to the planet, it is also called the Nearly Diurnal Free Wobble. Its frequency lies in the main nutation frequency band, which makes it particularly useful for nutation studies. For a planet consisting of two layers (mantle+core), the FCN frequency $\omega_{FCN}$ in the celestial reference frame can approximately be written as

$$\omega_{FCN} = \frac{\omega}{A_m} \left( \frac{1 - \frac{A}{A_m^3}}{2\alpha_f} \right)$$

(see, e.g., Dehant et al. 1993, Sasao et al. 1980). Here, the rotation frequency of Mars around its axis is given by $\Omega$, and $\alpha_f$ is the dynamical flattening at the CMB. The equatorial moments of inertia of the whole planet and the mantle are denoted by $A$ and $A_m$, respectively. The Love number $\Lambda_1$ denotes the

Love number for deformation induced by pressure loading at the CMB, and $q_0$ is the ratio of the centrifugal and gravitational acceleration at the equator. A list of global properties of Mars containing values of quantities such as, e.g., $J_2$ can be found in Spohn et al. (1998) and Dehant et al. (2000b). The main parameter determining the FCN frequency is the CMB flattening $\alpha_f$. For a planet in hydrostatic equilibrium, the flattening can be calculated from the density profile and the rotation rate by Clairaut’s equation (see, e.g., Moritz, 1990). For a homogeneous core, which is a good first approximation for the Martian core, the flattening is proportional to the density $\rho_c$ of the core:

$$\alpha_f = \frac{15}{16\pi G \rho_c} \frac{\Omega^2}{A}$$

(3)

Therefore, from an observed FCN frequency, a core density estimate can be derived (Van Hoolst et al. 2000a).

The FCN frequency can be constrained from nutation observations. In Figs. 2 and 3, the amplification of nutation amplitudes is given as a function of period for models differing in core radius and density (see also Fig. 1). The FCN resonance clearly dominates the behaviour for the models with a liquid core. In the retrograde band (Fig. 2), three main nutations are affected: the semi-annual, ter-annual, and quater-annual retrograde nutations. In the prograde band (Fig. 3), the amplification of the rigid body amplitudes is much smaller. However, the prograde semi-annual nutation is by far the largest (see Table 1), and is therefore important for the NEIGE experiment. The amplification is about 5 mas, above the observational accuracy of the experiment (Barriot et al. 2001). Therefore, from the observation of this nutation amplitude, it can be determined directly whether the core is (partially) liquid or entirely solid. Such a direct information on the core is unique for the NEIGE experiment.

Because it is far from the FCN frequency, the largest semi-annual prograde nutation is not sensitive to the
2002ESASP.514D...7F

**Figure 3.** Nutation amplification as a function of period for prograde periods.

FCN frequency and cannot be used to determine it. The smaller retrograde nutations can, however, be largely amplified depending on the closeness to the FCN frequency. The amplifications in Table 1 for the retrograde nutations of model A of Sohl and Spohn (1997) are below the NEIGE precision, but, for FCN frequencies close enough to one of these frequencies, the effects can become observable, and will then allow to estimate the FCN frequency from the amplitudes of the retrograde nutations, and to get information on the core density and radius. More details can be found in Dehant et al. (2000a) and Van Hoolst et al. (2000a and 2000b). To better quantify the precisions with which the rotational and interior structure parameters can be determined from the NEIGE experiment, numerical simulations based on an integrated orbitography software have been performed. By using a methodology based on the determination of the resonance strength and FCN frequency, a good determination of the parameters is possible (Yseboodt et al. 2002).

Additionally to the nutations, the NEIGE experiment will measure the precession rate of Mars with an error 5 times smaller than the present estimate (Barriot et al. 2001). An estimate of the polar moment of inertia $C$ can then be derived with an error of 0.1 %, 5 times better than at present. This will allow to further constrain the range of possible interior structure models. For example, with the present error on the polar moment of inertia $C$, all eleven models of Fig. 1 are possible, whereas only three would fall within the error bars for the NEIGE determination of $C$.

### 3.2. Inner core and Free Inner Core Nutation

The Free Inner Core Nutation (FICN, also called the Prograde Free Core Nutation, PFNC) is a relative motion of the instantaneous rotation axis of the solid inner core with respect to those of the mantle and liquid outer core. It is a prograde long period mode in an inertial reference system and is quasi-diurnal in

<table>
<thead>
<tr>
<th>nutation</th>
<th>period</th>
<th>rigid</th>
<th>nonrigid</th>
</tr>
</thead>
<tbody>
<tr>
<td>prograde 1/2 yr</td>
<td>343</td>
<td>499.5</td>
<td>5.2</td>
</tr>
<tr>
<td>retrograde 1 yr</td>
<td>687</td>
<td>136.5</td>
<td>1.4</td>
</tr>
<tr>
<td>retrograde 1/2 yr</td>
<td>343</td>
<td>18.0</td>
<td>0.6</td>
</tr>
<tr>
<td>retrograde 1/3 y</td>
<td>229</td>
<td>4.7</td>
<td>0.5</td>
</tr>
<tr>
<td>retrograde 1/4 y</td>
<td>172</td>
<td>0.8</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The FICN can, as the FCN, resonantly enhance nutation amplitudes, and observation of this effect directly implies the existence of a solid inner core. The width of the resonant amplification curve is, however, smaller than that for the FCN (see Fig. 4), and only observable amplifications can be obtained if the FICN frequency lies close to one of the main prograde nutations, the annual, semi-annual, ter-annual, or quarter-annual nutation (see Dehant et al. 2000b). We have determined the resonant amplification due to the FICN resonance for two models of Mars' interior. One with a small solid inner core of 500 km radius, the other with a large solid inner core of 1100 km radius. A major difference between these models is that the small core contains no light elements (assumed to be sulfur), whereas the outer layers of the core have the same composition as the liquid outer core (eutectic composition). This difference implies a large difference in density jump across the inner core boundary between the two models. Because this density jump is an important parameter in the FICN frequency, these frequencies differ largely between the two models (see Fig. 4). The observation of an amplification of a prograde nutation due to the FICN resonance would then not only be direct evidence of the existence of a solid inner core, but would also put a constraint on the dimension of the inner core.

### 4. CONCLUSION

The NEIGE experiment of the NetLander mission is unique in the sense that it can yield direct information on the core. The experiment will observe nutations of Mars that can be amplified by two rotational normal modes of the planet if Mars has a liquid outer core and a solid inner core. The Free Core Nutation is the most important normal mode, and enhances the nutations when Mars has a liquid core. Its effects are observable by the NEIGE experiment, and can be interpreted directly in terms of the physical state of the core and in terms of density of the core.
The possible existence of a solid inner core, which is a crucial factor in thermal evolution models, and a constraint on its size could also be determined if a resonance effect of the Free Inner Core Nutation can be measured in the prograde nutations. NEIGE will also determine, by measuring the precession rate of Mars, the moment of inertia of Mars with an error of only 0.1 %, which will allow a more better determination of the profiles of density and temperature inside Mars.

REFERENCES

Acuña, M.H., et al., Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment, Science, 284, 790-793, 1999


Defraigne, P., de Viron, O., Dehant, V., Van Hoolst, T., HOURD, F., Mars rotation variations induced by atmospheric and ice caps, J. Geophys. Res., 105, E10, 24563-24570, 2000

Dehant, V., J. Hinderer, H. Legros, and M. Lefftz, Analytical approach to the computation of the Earth, the outer core and the inner core rotational motions, Phys. Earth Planet. Inter., 76, 259-282, 1993


Dehant, V., Van Hoolst, T., Defraigne, P., Comparison between the nutations of the planet Mars and the nutations of the Earth, Surveys in Geophysics, 21, 89-110, 2000


Moritz, H., The Figure of the Earth. Theoretical Geodesy and the Earth's Interior, Herbert Wichmann Verlag GmbH, Karlsruhe, 1990


Van Hoolst, T., Dehant, V., Defraigne, P., Sensitivity of the Free Core Nutation and the Chandler Wobble to changes in the interior structure of Mars, Phys. Earth Planet. Inter., 117, 397-405, 2000a
SOFT X-RAY EMISSIONS FROM PLANETS, MOONS, AND COMETS


(1)Space Physics Laboratory, Vikram Sarabhai Space Centre, Trivandrum 695022, India, spl_vssc@vssc.org
(2)Southwest Research Institute, San Antonio, TX 78228, USA
(3)NASA Marshall Space Flight Center, Huntsville, AL 35812, USA
(4)University of Michigan, Ann Arbor, MI 48109, USA
(5)LPAP, University of Liege, Belgium
(6)University of Kansas, Lawrence, KS 66045, USA
(7)University of Wyoming, Laramie, WY 82071, USA
(8)Jet Propulsion Laboratory, Pasadena, CA 91109, USA
(9)University of California at Berkeley, Berkeley, CA 94720, USA
(10)SRON National Institute for Space Research, Utrecht, The Netherlands
(11)University of Virginia, Charlottesville, VA 22903, USA
(12)Massachusetts Institute of Technology, Cambridge, MA 02139, USA
(13)Boston University, Boston, MA 02215, USA
(14)Pennsylvania State University, State College, PA 16802, USA
(15)Blackett Laboratory, Imperial College, London, England
(16)Max-Planck-Institut für Aeronomie, Kättenburg-Lindau, Germany

ABSTRACT

A wide variety of solar system bodies are now known to radiate in the soft x-ray energy (<5 keV) regime. These include planets (Earth, Jupiter, Venus, Saturn, Mars): bodies having thick atmospheres, with or without intrinsic magnetic field; planetary satellites (Moon, Io, Europa, Ganymede): bodies with thin or no atmospheres; and comets and Io plasma torus: bodies having extended tenuous atmospheres. Several different methods have been proposed to explain the generation of soft x-rays from these objects, whereas in the hard x-ray energy range (>10 keV) x-rays mainly result from the electron bremsstrahlung process. In this paper we present a brief review of the x-ray observations on each of the planetary bodies and discuss their characteristics and proposed source mechanisms.

1. INTRODUCTION

The usually-defined range of x-ray photons spans 0.1-100 keV. Of this wide energy extent the soft x-ray energy band (<5 keV) is an important spectral regime for planetary remote sensing, as a large number of solar system objects are now known to shine at these wavelengths. These include Earth, Moon, Jupiter, Saturn, comets, Venus, Galilean satellites, Mars, Io plasma torus, and (of course) the Sun. Since Earth’s thick atmosphere efficiently absorbs x-ray radiation at lower altitudes (<30 km, even for hard x-rays), x-rays can only be observed from space by high-altitude balloon-, rocket-, and satellite-based instruments. But to observe most of the soft x-ray band one has to be above ~100 km from Earth’s surface. Terrestrial x-rays were discovered in the 1950s. The launch of the first x-ray satellite Uhuru in 1970 marked the beginning of satellite-based x-ray astronomy. Subsequently launched x-ray observatories - Einstein, and particularly Rontgensatellit (ROSAT), made important contributions to planetary x-ray studies. With the advent of the latest and most sophisticated x-ray observatories - Chandra and XMM-Newton – the field of planetary x-ray astronomy is advancing at a much faster pace. Earth and Jupiter, as magnetic planets, are observed to emanate strong x-ray emissions from their auroral (polar) regions, thus providing vital information on the nature of precipitating particles and their energization processes in planetary magnetospheres [1,2,3]. X-rays from low latitudes have also been observed on these planets. Saturn should also produce x-rays in the same way as Jupiter, although the intensity is expected to be weaker. Lunar x-rays have been observed from the sunlit hemisphere; and a small number of x-rays are also seen from the Moon’s nightside [4]. Cometary x-rays are now a well-established phenomena; more than a dozen comets have been observed at soft x-ray energies [5,6]. The Chandra X-ray Observatory (CXO) has recently captured soft x-rays from Venus [7,8]. Martian x-rays are expected to be similar to those on Venus. More recently, using CXO [9] have discovered soft x-rays from the inner moons of Jupiter - Io, Europa, and probably Ganymede. The Io Plasma Torus (IPT) was also discovered recently.
by CXO to be a source of soft x-rays [9]. Though the x-rays from Jupiter were discovered in 1979 by Einstein observatory [cf. Ref. 2], the recent high spatial resolution observations by CXO/HRC-I have revealed a mysterious pulsating (period ~45 minutes) x-ray hot spot in the northern polar regions of Jupiter that have called into question our understanding of Jovian auroral x-rays [10]. In this paper we will present an overview of soft x-ray observations on planets, comets, and moons, and discuss the proposed emission production mechanisms. The Sun and heliosphere are the other known sources of soft x-rays in our solar system. The solar x-rays arise in the solar corona (which has a temperature of ~10^6 K), and consist of both line and continuum x-ray radiation that are produced by excitation of highly charged ions and thermal bremsstrahlung processes, respectively [11]. The heliospheric x-rays are observed as a part of the soft x-ray background [12,13], and are largely produced through charge transfer collision between highly stripped heavy solar wind ions and interstellar neutrals in the heliosphere [14]. The solar and heliospheric x-rays will not be discussed further as they are not covered by the topic of this paper.

2. EARTH
2.1. Auroral Emissions
Precipitating particles deposit their energy into the Earth's atmosphere by ionization, excitation, dissociation, and heating of the neutral gas. High-energy electrons or ions impacting the nucleus of atoms or molecules can lead to an emission of an x-ray photon by bremsstrahlung with an energy comparable to the energy of the incident particle. The x-ray bremsstrahlung production efficiency is proportional to 1/m, where m is the mass of the precipitating particle. This implies that electrons are 10^6 times more efficient than protons at producing x-ray bremsstrahlung. The production efficiency is a non-linear function of energy, with increasing efficiency for increasing incident energies. For example, for a 200 keV electron the probability of producing an x-ray photon at any energy below 200 keV is 0.5%, while the probability for a 20 keV electron to produce an x-ray photon below 20 keV is only 0.0057% [15].

The main x-ray production mechanism in Earth's auroral zones is electron bremsstrahlung, and therefore the x-ray spectrum of the aurora has been found to be very useful in studying the characteristics of energetic electron precipitation [16,17,18]. Since the x-ray measurements are not contaminated by sunlight, the remote sensing of x-rays can be used to study energetic electron precipitation on the nightside as well as on the dayside of the Earth [19]. Characteristic line emissions for the main species of the Earth's atmosphere, Nitrogen (Kα at 0.393 keV), Oxygen (Kα at 0.524 keV) and Argon (Kα at 2.958 keV, Kβ at 3.191 keV) will also be produced by both electrons and protons, but so far no x-ray observations have been made at energies where these lines are dominant compared to the x-ray bremsstrahlung.

Fig.1. Auroral x-ray image of the Earth from the Polar PIXIE instrument (energy range 2.81-9.83 keV) obtained on July 31, 1997. The red box denotes the PIXIE field-of-view. The red dashed line and solid black line represent the day/night boundary and local noon, respectively. The grid in the picture is in geomagnetic coordinates and the numbers shown in red are magnetic local time.

X-ray photons from bremsstrahlung are emitted dominantly in the direction of the precipitating electron velocity. Consequently, the majority of the x-ray photons in Earth's aurora are directed towards the planet. These downward propagating x-rays, therefore, cause additional ionization and excitation in the atmosphere below the altitude where the precipitating particles have their peak energy deposition [e.g., Ref. 20,21]. The fraction of the x-ray emission that is moving away from the ground can be studied using satellite-based imagers, e.g., AXIS on UARS and PIXIE on POLAR.

Auroral x-ray bremsstrahlung has been observed from balloons and rockets since the 1960s and from spacecraft since the 1970s [22,23,24,25,16,17]. Due to the detector techniques that have been used, only x-rays above ~3 keV have been observed from the Earth's ionosphere. The PIXIE (Polar Ionospheric X-Ray Imaging Experiment) aboard Polar [26] is the first x-ray detector that provides true 2-D global x-ray image at energies >=3 keV (cf. Fig. 1). Because of the high apogee of the Polar satellite (~9 R⊕), PIXIE is able to image the entire auroral oval with a spatial resolution of ~700 km for long duration when the satellite is around apogee. This has helped to study the morphology of the x-ray aurora and its spatial and temporal variation, and consequently the evolution of energetic electron precipitation during magnetic storms (days) and substorms (1-2 hours). Data from the PIXIE camera showed that the x-ray substorm brightens up in the midnight sector and has a prolonged and delayed
maximum in the morning sector due to the scattering of drifting electrons [27]. Statistically the x-ray bremsstrahlung intensity peaks in the midnight substorm onset, is significant in the morning sector, and has a minimum in the early dusk sector [28]. During the onset/expansion phase of a typical substorm the electron energy deposition power is 60-90 GW, which produces 10-30 MW of bremsstrahlung x-rays [29]. Combining the results of PIXIE with the UV imager aboard Polar, it is possible to derive the energy distribution of precipitating electrons in the 0.1-100 keV range [18].

Fig. 2. X-ray image of Earth from the Polar PIXIE instrument for energy range 2.9-10.1 keV obtained on August 17, 1998, showing the dayside x-rays during a solar x-ray flare. The grid in the picture is in geomagnetic coordinates, and the numbers shown in red are magnetic local time. The terminator at the surface of the Earth is shown as a red dashed line.

2.2 Non-Auroral Emissions

The non-auroral x-ray background above 2 keV from the Earth is almost completely negligible except for brief periods during major solar flares [28]. However, at energies below 2 keV soft x-rays from the sunlit Earth's atmosphere have been observed even during quite (non-flaring) Sun conditions [e.g., ref. 30,31]. The two primary mechanisms for the production of x-rays from the sunlit atmosphere are: 1) the Thomson (coherent) scattering of solar x-rays from the electrons in the atomic and molecular constituents of the atmosphere, and 2) the absorption of incident solar x-rays followed by the emission of characteristic K lines of Nitrogen, Oxygen, and Argon. Fig. 2 shows the PIXIE image of Earth demonstrating the x-rays (2.9-10 keV) production in the sunlit atmosphere during a solar flare of August 17, 1998. During flares, solar x-rays light up the sunlit side of the Earth by Thomson scattering, as well as by fluorescence of atmospheric Ar to produce characteristic x-rays at 3 keV, which can be observed by PIXIE camera. The x-ray brightness can be comparable to that of a moderate aurora. Ref. [28] examined the x-ray spectra from PIXIE for two solar flare events during 1998. They showed that the shape of the measured x-ray spectra was in fairly good agreement with modeled spectra of solar x-rays subject to Thomson scattering and argon fluorescence in the Earth's atmosphere.

3. JUPITER

3.1 Auroral Emissions

The first detection of x-ray emissions from Jupiter was made by the satellite-based Einstein observatory in 1979 [32]. The emissions were detected in the 0.2-3.0 keV energy range from both poles of Jupiter. Analogous to the processes on Earth, it was expected that Jupiter's x-rays might originate as bremsstrahlung by precipitating electrons [33]. However, the power requirement for producing the observed emission with this mechanism ($10^{17}$-$10^{19}$ W) is more than two orders of magnitude larger than the input auroral power available as derived from Voyager and IUE observations of the ultraviolet aurora [cf. Ref. 2]. Ref. [32] suggested a mechanism implying K-shell line emissions from precipitating energetic sulfur and oxygen ions from the inner magnetosphere, with energies in the 0.3-4.0 MeV/nucleon range. The heavy ions are thought to emit x-rays by first charge stripping to a highly ionized state, followed by charge exchange and excitation through collisions with H$_2$. The bremsstrahlung process was further ruled out by theoretical models [34,35] showing that primary and secondary precipitating electrons in the 10-100 keV energy range are inefficient at producing the observed x-ray emissions.

Fig. 3. Chandra x-ray image of Jupiter on 18 December 2000 generated from 10 hours of continuous observations. A jovianentric graticule with $30^\circ$ intervals is overplotted, along with the L=5.9 (orange lines) and L=30 (green lines) footprints of the magnetic field model. The image shows strong auroral x-ray emissions from high latitudes and rather uniform emissions from the disk. [from Ref. 10].

© European Space Agency • Provided by the NASA Astrophysics Data System
Furthermore, during its Jovian flyby, the Ulysses spacecraft did not detect significant emissions in the 27-48 keV energy range as would have been the case if electron bremsstrahlung was a major process. Observations of Jupiter x-ray's emissions by ROSAT supported the suggestion of [32] and the model calculations [34,35] that precipitating energetic (>700 keV per nucleon) S and O ions are most probably responsible for the x-ray emissions from Jupiter. A detailed modeling of the x-ray production [38,39] suggests that recombination lines from highly charged precipitating O and S ions mainly contribute to the soft x-rays detected by ROSAT.

Fig. 4. This composite image displays x-ray data from Chandra (magenta) and ultraviolet data from Hubble Space Telescope (blue) overlaid on an optical image of Jupiter. While Chandra observed Jupiter for an entire 10-hour rotation period on December 18, 2000, this image shows a 'snapshot' of a single 45-minute X-ray pulse. [from http://chandra.harvard.edu/photo/2002/0001/0001_xray_opt_uv/jpeg].

Recent high-spatial resolution observations of Jupiter with the Chandra telescope [10] reveal that most of Jupiter's northern auroral x-rays come from a "hot spot" located significantly poleward of the latitudes connected to the inner magnetosphere (cf. Figs. 3, 4). The hot spot is fixed in magnetic latitude and longitude and occurs in a region where anomalous infrared and ultraviolet emissions (the so-called "flares") have also been observed. Its location must connect along magnetic field lines to regions in the Jovian magnetosphere well in excess of 30 Jovian radii from the planet (cf. Fig. 3), a region where there are insufficient S and O ions to account for the hot spot [10]. More surprisingly, the hot spot x-rays pulsate with an approximately 45-min period, similar to that reported for high-latitude radio and energetic electron bursts observed by near-Jupiter spacecraft [cf. Ref. 10].

These results invalidate the idea that Jovian auroral x-ray emissions are mainly excited by steady precipitation of heavy energetic ions from the inner magnetosphere [cf. Ref. 2]. Instead, the x-rays seem to result from currently unexplained processes in the outer magnetosphere that produce highly localized and highly variable emissions over an extremely wide range of wavelengths. In any case, the power needed to produce the brightest ultraviolet "flares" seen in the same polar cap region as the x-ray hot spot is a few tens of TW, much less than the estimated power of a few PW needed to produce the observed x-rays by electron bremsstrahlung. Thus, electron bremsstrahlung still seems to fail in explaining the observed Jovian x-rays hot spot. One possible source of Jovian x-rays production is via charge-exchange of solar wind ions that penetrate down the atmosphere in the magnetic cusp region. But in this case the solar wind ions would have to be accelerated to much higher energies (100s of keV), probably by parallel electric field or wave particle interactions, to generate sufficient luminosity to account for the observations. In summary, at the present time the origin of the Jovian x-rays and its source is still an open issue.

3.2. Non-Auroral Emissions

Soft x-ray emissions with brightnesses of about 0.01-0.2 Rayleighs were observed from the equatorial regions of Jupiter using the ROSAT/HRI. It was proposed [40] that the equatorial emission, like the auroral emission, may be largely due to the precipitation of energetic (>300 keV/amu) sulfur or oxygen ions into the atmosphere from the radiation belts. Further evidence for a correlation between regions of low magnetic field strength and enhanced emission [41] lent additional support to this mechanism, since it can be assumed that the loss cone for precipitating particles is wider in regions of weak surface magnetic field. However, [42] showed that two alternative mechanisms should not be overlooked in the search for a complete explanation of low-latitude x-ray emission, namely elastic scattering of solar x-rays by atmospheric neutrals and fluorescent scattering of carbon K-shell x-rays from methane molecules located below the Jovian homopause. Modeled brightnesses agree, up to a factor of two, with the bulk of low-latitude ROSAT/PSPC measurements which suggests that solar photon scattering (~90% elastic scattering) may act in conjunction with energetic heavy ion precipitation to generate Jovian non-auroral x-ray emission. The solar x-ray scattering mechanism is also suggested from the correlations of Jovian emissions with the F10.7 solar flux and of the x-ray limb with the bright visible limb [41]. During the December 2000 observations by Chandra HRC-I, the disk-averaged emitted x-ray power was about 2 GW [10], but the signal-to-noise ratio of the disk emission was not adequate to show the limb brightening expected by solar photon-driven x-ray emission.
4. MOON

Though it is the Earth's nearest planetary body, the Moon has been relatively little studied at x-ray wavelengths. Other than the discovery observation by [4] using the ROSAT PSPC and a detection by Advanced Satellite for Cosmology and Astrophysics (ASCA) [43], most recent high-energy remote sensing of the Moon has been made at extreme- and far-ultraviolet wavelengths [e.g., Ref. 44,45,46]. However, as noted by [47], x-ray fluorescence studies could provide an excellent way to determine the elemental composition of the lunar surface by remote sensing, since the soft x-ray optical properties of the lunar surface should be dominated by elemental abundances (rather than mineral abundances, which determine the optical properties at visible and longer wavelengths). Although reflection of the strong solar lines likely dominates the soft x-ray spectrum of the Moon, the detection of weaker emissions due to L- and M-shell fluorescence would provide a direct measure of specific elemental abundances.

**Fig. 5.** ROSAT soft x-ray (0.1-2 keV) images of the Moon at first (left side) and last (right side) quarter. The dayside lunar emissions are thought to be primarily reflected and fluoresced sunlight, while the origin of the faint but distinct nightside emissions is uncertain. The brightness scale in R assumes an average effective area of 100 cm² for the ROSAT PSPC over the lunar spectrum.

Fig. 5 shows ROSAT data; the left image shows the [4] data, while the right image is unpublished data from a lunar occultation of the bright x-ray source GX5-1 (the higher energy x-rays from GX5-1 have been suppressed in this figure, but a faint trail to the upper left of the Moon remains). The power of the reflected and fluoresced x-rays observed by ROSAT in the 0.1-2 keV range coming from the sunlit surface was determined by [4] to be only 73 kW, making the Moon the faintest x-ray source in the sky (the flux measured was 2.5x10⁻¹⁰ erg cm⁻² s⁻¹).

While the dayside lunar soft x-rays are reflected and fluoresced sunlight, the faint but distinct lunar nightside emissions are a matter of controversy. Ref. [4] suggested that solar wind electrons of several hundred eV might be able to impact the night side of the Moon on the leading hemisphere of the Earth-Moon orbit around the sun. However, this was before the GX5-1 data were acquired, which clearly show lunar nightside x-rays from the trailing hemisphere as well. Another possible explanation is the accepted mechanism for comet x-rays, heavy ion solar wind charge exchange (SWCX) [e.g., Ref. 6]. In this case, however, the heavy ions in the solar wind would be charge exchanging with geocoronal H atoms that lie between the Earth and Moon but lie outside the Earth's magnetosphere.

Future observations of the Moon's x-rays by Chandra and XMM are likely, and there is a planned x-ray spectrometer D-CIXS on ESA's SMART-1 lunar mission [48] that will provide global coverage of ambient x-ray emission. This will greatly improve upon the elemental abundance maps produced by the x-ray spectrometers on Apollo 15 and 16 [49].

5. COMETS

X-ray emission from a comet was first discovered in 1996 with the ROSAT observations of comet Hyakutake [50]. Extreme ultraviolet (EUV) emission was also detected from this comet by the EUVE satellite [51]. Since the initial discovery of cometary x-ray emission, it has been shown that active comets are almost always EUV and soft x-ray sources [52]. A thorough review of this topic has just appeared [6], so only a brief summary is provided here.

**Fig. 6.** X-ray and visible images of comet C/LINEAR 1999 S4. Left: Chandra x-ray (0.2-0.8 keV) July 14, 2002 ACIS-S image. Right: Visible light image of comet taken on July 14, 2002 showing a symmetric coma and a long anti-solar tail. The plus sign mark the position of the nucleus [from Ref. 5].

The key observational features of cometary x-ray emission (cf. Fig. 6) are now summarized. The x-ray emission is "very soft" with photon energies of a few
hundred eV or less. High-resolution x-ray spectra of comets C/Linear 1999 S4 [5] and McNaught-Hartley (C/1999 T1) [53] have recently been measured by the CXO. Emission lines associated with highly charged ions (in particular, O\(^{6+}\)) are evident in these spectra. A spectrum of comet Hyakutake from the EUVE satellite [54] also displays line emission. Cometary x-ray luminosities are quite large and tend to correlate with the gas production rate (e.g., Ref. 55). The comet Hyakutake x-ray luminosity measured by ROSAT was about 1 GW. Cometary x-ray emission is spatially very extensive with observed emission out to radial distances from the cometary nucleus of 10\(^3\)-10\(^6\) km (e.g., Ref. 5,50). The emission is also time variable and has been shown to correlate with the solar wind flux (e.g., Ref. 56).

Several cometary x-ray emission mechanisms were proposed following the initial discovery. These include bremsstrahlung associated with solar wind electron collisions with cometary neutrals and ions, K-shell ionization of neutrals by electron impact, scattering of solar photons by cometary dust, and charge transfer of solar wind heavy ions with neutrals (cf. Ref. 6). The SWCX mechanism [57] has gradually won favor. In this mechanism highly charged solar wind heavy ions (e.g., O\(^{8+}\), O\(^{7+}\), C\(^{6+}\), Ne\(^{11+}\)) undergo charge transfer collisions when they encounter cometary neutrals. The product ions are invariably left in highly excited states and emit EUV and soft x-ray photons. This mechanism is able to explain the luminosity, spatial distribution [58], time variability [56], and spectrum [5,53,59,60] of the x-ray emission.

6. VENUS

In January 2001 the CXO discovered soft x-ray emissions from Venus [7,8]. Observations were performed with the CXO's ACIS-I high-resolution imaging camera (cf. Fig. 7) and with the LETG/ACIS-S high-resolution grating spectrometer. The x-ray emission originated from the sunlit hemisphere, exhibited noticeable limb brightening (when compared to visible disk), and consisted primarily of O-K\(\alpha\) at 530 eV. The C-K\(\alpha\) line emission at 280 eV was also detected, with marginal evidence for N-K\(\alpha\) emission at 400 eV. The carbon emission might also include a 290 eV line from CO\(_2\) and CO. The derived energy fluxes in the oxygen and carbon lines were 20\times10\(^{-14}\) erg cm\(^{-2}\) s\(^{-1}\) for O-K\(\alpha\) and 5\times10\(^{-14}\) erg cm\(^{-2}\) s\(^{-1}\) for C-K\(\alpha\). The total power in Venusian x-rays was estimated to be 30-70 MW. The Chandra data also indicated possible time variability on a timescale of a few minutes.

The observers [8] argue persuasively that the x-ray emission from Venus results from fluorescent scattering of solar x-rays, as predicted by [61], and not from charge exchange between heavy ions in the solar wind and neutral atoms in the Venusian atmosphere. Their detailed modeling of the interaction between solar x-rays and the planetary atmosphere showed that the fluorescent scattering occurred \(\approx 110\) km or higher above the planet's surface. The amount of limb brightening predicted by their models depended sensitively on the chemical composition and the density profile in Venus's upper atmosphere.

7. GALILEAN SATURN S

Recently the CXO has discovered [9] x-ray emission from the Galilean satellites (cf. Fig. 8). The CXO observations of the Jovian system were made on 25-26 November 1999 for 86.4 ks with the ACIS-S instrument and on 18 December 2000 for 36.0 ks with the HRC-I instruments. The time tagged nature of the CXO data makes it possible to correct for varying satellite motions, and with ACIS it is also possible to filter the data by energy for optimum sensitivity. During the ACIS-S observation, Io and Europa were detected with a high degree of confidence, and Ganymede at a lesser degree of confidence. Io was also detected with high confidence during the shorter HRC-1 observation. Over the nominal energies of 300-1890 eV range detected by ACIS-S, the x-ray events show a clustering between 500 and 700 eV, probably dominated by the oxygen K\(\alpha\) line at 525 eV. The estimated energy fluxes at the telescope and power emitted are 4\times10\(^{-16}\) erg cm\(^{-2}\) s\(^{-1}\) and 2.0 MW for Io, and 3\times10\(^{-16}\) erg cm\(^{-2}\) s\(^{-1}\) and 1.5 MW for Europa. Ganymede was roughly a third as luminous as Io. Callisto was not detected in either set of data.

The most plausible emission mechanism is inner (K shell) ionization of the surface (and perhaps incoming magnetospheric) atoms followed by prompt x-ray emission. Oxygen should be the dominant emitting atom in either a silicate or SO\(_2\) surface (Io) or in an icy one (the outer Galilean satellites). It is also the most common heavy ion in the Jovian magnetosphere. The extremely tenuous atmospheres of the satellites are transparent to x-ray photons with these energies, and also to much of the energy range of the incoming ions. However, oxygen absorption of the 525 eV line is such that the x-rays must originate in the top \(\approx 10\) microns of the surface in order to escape. Simple estimates suggest that excitation by
incoming ions dominates over electrons and that the x-ray flux produced is sufficient to account for the observations.

![Image of Io and Europa](image1)

Fig. 8. Chandra ACIS-I image (0.2-2 keV) of Io and Europa obtained on November 25-26, 1999. The solid circle shows the size of the satellite (the radii of Io and Europa are 1821 km and 1560 km, respectively), and the dotted circle the size of the detector cell. The axes are labeled in arcsec (1 arcsec = 2995 km) and the scale bar is in units of smoothed counts per image pixel (0.492 by 0.492 arcsec). [from Ref. 9].

Detailed models are required for verifying this picture and also for predicting the strengths of Kα lines for elements other than oxygen, especially heavier ones such as Na, Mg, Al, Si, and S. Within this framework, it is possible to constrain the surface composition of these moons from x-ray observations, but this requires a greater signal-to-noise ratio than provided by the Chandra observations. The detection of x-ray emission from the Galilean satellites thus provides a direct measure of the interactions of magnetosphere of Jupiter with the satellite surfaces.

8. IO PLASMA TORUS

The Io Plasma Torus (IPT) is known to emit at EUV energies and below [62,63,64], but it was a surprise when CXO discovered that it was also a soft x-ray source [9]. The Jovian system has so far been observed with Chandra using the ACIS-S high-spatial-resolution imaging camera, which also has modest energy resolution, for two Jovian rotations in November 1999, and using the HRC-I high-spatial-resolution camera, with essentially no energy resolution, for one rotation in December 2000. X-ray emission from the IPT is present in both observations. The ACIS-S spectrum was consistent with a steep power-law continuum (photon index 6.8) plus a gaussian line (complex) centered at ~569 eV, consistent with Kα emission from various charge species of oxygen. Essentially no x-rays were observed above this spectral feature, consistent with the steepness of the power-law continuum. The 250-1000 eV energy flux at the telescope aperture was 2.4×10^{-14} erg cm^{-2} s^{-1}, corresponding to a luminosity of 0.12 GW, and was approximately evenly divided between the dawn and dusk side of Jupiter.

However, the line emission originated predominantly on the dawn side. During the ACIS-S observation (cf. Fig. 9), Io, Europa, and Ganymede were on the dawn side, while Callisto was on the dusk side. For the HRC-I observation, the x-ray emission was stronger on the dusk side, approximately twice that observed on the dawn side. During the HRC-I observation, Io, Europa, and Ganymede were on the dusk side, while Callisto was on the dawn side.

![Image of Io plasma torus](image2)

Fig. 9. Chandra ACIS-I image of Io plasma torus obtained on November 25-26, 1999. The axes are labelled in units of Jupiter’s radius, R_J and the scale bar is in units of smoothed counts per image pixel (7.38 by 7.38 arcsec). For this observation, Jupiter’s radius corresponds to 23.8 arcsec. The paths traced by Io (solid line to the east), Europa (dashed line), Ganymede (dotted line), and Callisto (solid line to the west) are marked on the image. The regions bounded by rectangles were used to determine background. The regions bounded by ellipses were defined as x-ray source regions. [from Ref. 9].

The physical origin of the x-ray emission from the IPT is not yet fully understood. According to the estimates given in [9], fluorescent x-ray emission excited by solar x-rays, even during flares from the active Sun, charge-exchange processes, previously invoked to explain Jupiter’s x-ray aurora [e.g., Ref. 2,38] and cometary x-ray emission [e.g., Ref. 6,57], and ion stripping by dust grains fail to account for the observed emission. Assuming bremsstrahlung emission of soft x-rays by non-thermal electrons in the few hundred to few thousand eV range, with a kappa, or generalized Lorentzian, distribution with a temperature of 10 eV and an index of $\kappa = 2.4$ [65], which is consistent with the in-situ Ulysses observations, [9] estimated an IPT soft x-ray luminosity of 0.03 GW. This falls short of but is a significant fraction of the observed luminosity of 0.12 GW.

9. MARS

There seems to be a tentative detection of Martian x-rays in the ROSAT data [66]. However, so far no x-ray emission has been unambiguously detected from Mars,
but this situation is expected to change soon. As at Venus, absorption of solar x-rays in either the carbon or oxygen K-shells followed by fluorescent emission of x-rays is suggested as the dominant process of x-rays production on Mars [61]. The predicted total soft x-ray intensity is 0.15 R, corresponding to an x-ray luminosity of about 2.5 MW [61], which exceeds the x-ray luminosity expected from the solar wind charge exchange mechanism on Mars [67]. A simulated image of SWCX-produced x-ray emission at Mars [67] indicates that this emission has a very different spatial morphology than the fluoresced x-ray emission [61], and future observations must be able to distinguish between the two processes.

10. SATURN

The existence of a magnetosphere at Saturn and the presence within it of energetic electrons and ions, both first observed by instruments on Pioneer 11 and Voyagers 1 and 2, provide the conditions under which auroral emission can be expected. The first indication of an aurora at Saturn came from measurements in the ultraviolet by Pioneer 11, followed up by observations with the IUE satellite, and confirmed by the Voyager 1 and 2 flyby encounters, whose results included localizing the sources of emission to regions near the pole in both hemispheres [e.g., Ref. 68; cf. Ref. 2 for review]. These particle and field properties make it probable that auroral x-ray emission occurs at Saturn. The UV emission can be accounted for by electrons precipitating along high latitude field lines into Saturn's atmosphere [2]. X-ray emission could be generated by bremsstrahlung involving the high-energy portion of the precipitating electron distribution. Alternatively, energetic (~1 MeV C, N, and O) ions have been observed in Saturn's magnetosphere [69]. A fraction of these could precipitate into Saturn's atmosphere at high latitudes, generating x-rays through charge exchange reactions as has been postulated to occur at Jupiter [2]. It has also been suggested that a potential source of heavy ions, additional to solar wind injected plasma, is nitrogen escaping from the atmosphere of Saturn's satellite Titan [70]. Other possible contributory sources, particularly in view of the recent observation of x-ray emission from three of the four Galilean satellites of Jupiter [9], are Saturn's rings and satellite surfaces, to the degree these are exposed to bombardment by energetic particle fluxes.

Following the discovery of x-ray emission from Jupiter with the Einstein Observatory [32], an observation of Saturn was undertaken [71]. No x-ray emission was seen. With the conversion of count rate into flux dependent on spectral shape, and spectral shape a consequence of the production mechanism, the 3σ upper limit at Earth from this observation was calculated to be 5 × 10^{13} erg cm^{-2} s^{-1} if the mechanism is dominated by ion-produced characteristic lines, and 2 × 10^{12} erg cm^{-2} s^{-1} if the mechanism is electron bremsstrahlung. Assuming the latter, and basing the expected x-ray flux on the intensity of the UV aurora observed by Voyager 2, a model calculation of the expected energy flux in the 0.2-3.0 keV energy range yielded a value of 8×10^{15} erg cm^{-2} s^{-1}, more than two orders of magnitude below the observed upper limit.

In 1992, [72] observed Saturn with the ROSAT position sensitive proportional counter, an instrument with superior sensitivity to soft x-rays relative to the corresponding instrument on the Einstein Observatory. The observed events were grouped into two energy bands. The harder band saw nothing significant, but the soft band, ranging from 0.1 to 0.55 keV, recorded almost three times as many counts as expected from background. The corresponding net energy flux is 1.9 × 10^{14} erg cm^{-2} s^{-1}. This is more than an order of magnitude greater than the model estimate for x-ray production based on electron bremsstrahlung [71], an estimate more likely to be on the high than the low side [73]. Furthermore, the observed flux amounts to 24% of the flux observed from Jupiter under identical instrument conditions after removing the effect due to their different distances from Earth. Even if the predominant mechanism at both planets is some form of ion precipitation, a mechanism which produces x-rays more efficiently than electron bremsstrahlung, but which recent Chandra Observatory results have called into question at Jupiter [10] and which seems less likely at Saturn, emission at Saturn would be expected to be less than 10% of that from Jupiter, a ratio more consistent with that of the observed auroral intensities [74]. The non-auroral x-rays from Saturn is expected to have a predominantly solar-driven elastic scattering, much like Jupiter (predominantly H\textsubscript{2} scattering with a small CH\textsubscript{4} component) with perhaps about a third as much integrated non-auroral power as Jupiter. Additional observations of Saturn with the Chandra or XMM-Newton Observatories are called for to follow up on the intriguing result from ROSAT.

11. SUMMARY

Table 1 summarizes our current knowledge of the x-ray emissions from the planetary bodies that have been observed to produce soft x-rays. Several other solar system bodies, which include Titan, Uranus, Neptune, and inner-icy satellites of Saturn, are also expected to be x-ray sources, but are yet to be detected. The x-rays from solar system bodies are relatively weak (< few GW) and are from a much colder (T <10\textsuperscript{5} K) environment than x-rays from stars and other cosmic bodies (T >10\textsuperscript{6} K). Nonetheless, the studies of planetary x-rays help advance our understanding of basic plasma-neutral (in both gas and solid phase) interactions that are important within our solar system, and plausibly in the extra-solar planetary systems as well.
Table 1. Summary of the characteristics of soft x-rays from solar system bodies

<table>
<thead>
<tr>
<th>Object</th>
<th>Emitting Region</th>
<th>Power Emitted$^a$</th>
<th>Special Characteristics</th>
<th>Possible Production Mechanism</th>
<th>References$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth</td>
<td>Auroral atmosphere</td>
<td>10-30 MW</td>
<td>Correlated with magnetic storm and substorm activity</td>
<td>Bremsstrahlung from precipitating electrons</td>
<td>[29]</td>
</tr>
<tr>
<td>Earth</td>
<td>Non-auroral atmosphere</td>
<td>40 MW</td>
<td>Correlated with solar x-ray flux</td>
<td>Scattering of solar x-rays by atmosphere</td>
<td>[31]</td>
</tr>
<tr>
<td>Jupiter</td>
<td>Auroral atmosphere</td>
<td>0.4-1 GW</td>
<td>Pulsating (~45 min) x-ray hot spot in north polar region</td>
<td>Energetic ion precipitation from magnetosphere and/or solar wind + electron bremsstrahlung +</td>
<td>[32,10]</td>
</tr>
<tr>
<td>Jupiter</td>
<td>Non-auroral atmosphere</td>
<td>0.5-2 GW</td>
<td>Quite uniform over disk</td>
<td>Resonant scattering of solar x-rays + Ion precipitation from ring current</td>
<td>[40,42,10]</td>
</tr>
<tr>
<td>Moon</td>
<td>Surface - Dayside</td>
<td>0.07 MW</td>
<td>Nightside emissions are ~1% of the dayside emissions</td>
<td>Scattering of solar x-rays by the surface elements on dayside. Electron bremsstrahlung + SWCX</td>
<td>[4,43]</td>
</tr>
<tr>
<td></td>
<td>- Nightside</td>
<td></td>
<td></td>
<td>of geocorona?</td>
<td></td>
</tr>
<tr>
<td>Comets</td>
<td>Sunward-side coma</td>
<td>0.2-1 GW</td>
<td>Intensity peaks in sunward direction $10^7$ km ahead of coma</td>
<td>SWCX with cometary neutrals + other mechanisms</td>
<td>[5,6,50]</td>
</tr>
<tr>
<td>Venus</td>
<td>Sunlit atmosphere</td>
<td>50 MW</td>
<td>Emissions come from ~120-140 km above the surface</td>
<td>Fluorescent scattering of solar x-rays by C and O atoms in the atmosphere</td>
<td>[7,8,61]</td>
</tr>
<tr>
<td>Io</td>
<td>Surface</td>
<td>2 MW</td>
<td>Emissions from upper microns of the surface</td>
<td>Energetic Jovian magnetospheric ions impact on the surface + ?</td>
<td>[9]</td>
</tr>
<tr>
<td>Europa</td>
<td>Surface</td>
<td>1.5 MW</td>
<td>Emissions from upper microns of the surface</td>
<td>Energetic Jovian magnetospheric ions impact on the surface + ?</td>
<td>[9]</td>
</tr>
<tr>
<td>Io Plasma Torus</td>
<td>Plasma torus</td>
<td>0.1 GW</td>
<td>Dawn-dusk asymmetry observed?</td>
<td>Electron bremsstrahlung + ?</td>
<td>[9]</td>
</tr>
<tr>
<td>Saturn</td>
<td>Auroral and non-auroral atmosphere</td>
<td>0.4 GW</td>
<td>Plausibly similar to Jovian x-rays</td>
<td>Electron bremsstrahlung + scattering of solar x-rays</td>
<td>[72]</td>
</tr>
<tr>
<td>Mars</td>
<td>Atmosphere</td>
<td>3 MW</td>
<td>Probable detection?</td>
<td>Solar fluorescence + SWCX</td>
<td>[66,61]</td>
</tr>
</tbody>
</table>

$^a$The values quoted are “typical” values at the time of observation. X-rays from all bodies are expected to vary with time. For comparison the total x-ray luminosity from the Sun is $10^{20}$ W and that from the heliosphere $10^{18}$ W [75].

$^b$Only representative references are given.

SWCX = Solar wind charge exchange = charge exchange of heavily ionized solar wind ions with neutrals.

12. REFERENCES


Session 7
Earth-like Planets and Moons in the Galaxy

Chair: S. Volonté
SEARCHING FOR AND CHARACTERISING EXTRASOLAR EARTH-LIKE PLANETS AND MOONS

Schneider Jean
CNRS - Observatoire de Paris, 92195 Meudon, France

ABSTRACT

The physical bases of the detection and characterisation of extrasolar Earth-like planets and Moons in the reflected light and thermal emission regimes are reviewed. They both have their advantages and disadvantages, including artefacts, in the determination of planet physical parameters (mass, size, albedo, surface and atmospheric conditions etc.). After a short panorama of detection methods and the first findings, new perspectives for these different aspects are also presented. Finally brief account of the ground based programmes and space-based projects and their potentialities for Earth-like planets is made and discussed.

Key words: extrasolar planets; terrestrial planets; space missions.

1. INTRODUCTION

The first discoveries of extrasolar planets by Latham et al. (1989), Wolszczan and Frail (1992) and Mayor and Queloz (1995) have triggered a renewal of the permanent question on the possible presence of life outside the Solar System. This question can now be addressed in scientific terms. Before detecting life on exoplanets, it is necessary to detect and characterize these planets. The table 1 presents on the left side the major characteristics of the extrasolar planets which we can reasonably hope to measure in a near future. The right side of the table summarizes the methods by which we can access these characteristics. The art is then to cross-correlate the questions on planet characteristics with the means of planet detections. The only successful detection method, radial velocity, has up to now (June 2002) provided 99 planets. A excellent review, with the first findings, is given by M. Perryman (1999). I will first remind the principle of the detection methods, he most promising being the direct imaging of planets. I will then discuss the reflected light and thermal emission approaches of the direct imaging of planets. Finally, the ground based programmes and space mission projects will be reviewed.

### Table 1

<table>
<thead>
<tr>
<th>Planet characteristics</th>
<th>Detection methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbit: $P, a, e, i$</td>
<td>Star's wobble:</td>
</tr>
<tr>
<td>Mass $M_{pl}$</td>
<td>- Rad. velocity</td>
</tr>
<tr>
<td>Radius $R_{pl}$</td>
<td>- Astrometry</td>
</tr>
<tr>
<td>Temperature $T_{pl}$</td>
<td>- Timing</td>
</tr>
<tr>
<td>Albedo $A_{pl}(\lambda)$</td>
<td>Transits</td>
</tr>
<tr>
<td>Environment:</td>
<td>Lensing</td>
</tr>
<tr>
<td>- Atmosph. gases</td>
<td>Imaging</td>
</tr>
<tr>
<td>- Clouds</td>
<td>Radio detect.</td>
</tr>
<tr>
<td>- Rings</td>
<td></td>
</tr>
<tr>
<td>- Moons</td>
<td></td>
</tr>
<tr>
<td>- Magnetospheres</td>
<td></td>
</tr>
<tr>
<td>Surf. structures:</td>
<td></td>
</tr>
<tr>
<td>- Continents, Oceans</td>
<td></td>
</tr>
<tr>
<td>Variations:</td>
<td></td>
</tr>
<tr>
<td>- Day, seasons</td>
<td></td>
</tr>
<tr>
<td>Life?</td>
<td></td>
</tr>
<tr>
<td>Biosignatures</td>
<td></td>
</tr>
</tbody>
</table>

2. DETECTION AND CHARACTERIZATION TECHNIQUES - FIRST RESULTS

Let us first remind how extrasolar planets can be detected and characterized.

1. Stellar wobble:
When the planet, with a mass $M_{pl}$ makes a revolution at a distance $a$ around its parent star located at a distance $D$ from the Sun, in fact both objects are in orbit around their common center of mass. Thus
the star makes a small circular orbit with a radius \(a = a \cdot M_\text{pl}/M_\star\) and a period naturally equal to the planet orbital period \(P = 2\pi \sqrt{a^3/(GM_\star)}\). This wobble affects, with a period \(P\), three star’s observables:

1. Its radial velocity, with an amplitude \(V_R = (M_\text{pl} \sin i/M_\star)(GM_\star/a)^{1/2}\)

2. Its position \((x, y)\) on the sky, with amplitudes \((\alpha_x = (M_\text{pl}/M_\star)(a/D), \alpha_y = (M_\text{pl} \sin i/M_\star)(a/D))\)

3. Its distance to the observer, with an amplitude \(D_\text{pl} = \alpha_x \sin i\). The latter can be measured by the perturbation of the time of arrival of periodic signals (such as pulsar pulses).

In the above formulae, \(i\) is the inclination on the sky plane of the planet orbit. The radial velocity and timing measurements do not provide the measurement of \(i\) and thus provide only the product \(M_\text{pl} \sin i\). On the contrary, the astrometric measurement providing two observables \(\alpha_x\) and \(\alpha_y\), this method gives both the mass of the planet and the inclination of the orbit.

With the exception of the two pulsars PSR 1257+12 (Wolszczan 1994) and PSR B1620-26 (Rasio 1994), the radial velocity method is the only one having made confirmed planet detections. A confirmation of the planet and a measurement of its mass by astrometry has been made with the Hubble Space Telescope Fine Guiding Sensors (Benedict et al. 2002). At the time of writing (28 June 2002), there are 99 planets in 86 planetary systems with 9 systems with 2 planets and 2 systems with 3 planets. An exhaustive catalog is maintained daily on the web (Schneider 2002a). The main lessons of these findings are (Perryman 1999, Schneider 1999):

1. At least 5% of stars have giant planets. This proportion constitutes an important element which has now to be explained by theories of planetary systems formation. It was a complete unknown when astronomers started to search for planets 15 years ago; it could have been close to 100% as well as less than say 0.1%.

2. Many orbits are surprisingly very close (up to 0.05 AU) from the star. As planets cannot form so closely to the parent star (Boss 1995), the standard explanation is that they start to form at about 5 AU and, along the formation process, migrate towards the star (e.g. Lin, Bodenheimer and Richardson 1996).

3. Another surprise is that more than half of the orbits have an eccentricity larger than 0.2. This is probably due to some gravitational perturbation, but its cause is presently not clear.

4. Statistically, stars harbouring planets are more metallic than the mean (Gonzalez 1996). This supermetallicity can be due to either an enrichment of the star by infalling planets, the preference for planets to form around metal-rich stars or both.

2. Transits:
The planet can produce a drop in the star light during transits of the star disk by the planet. The detection of a transit in the star lightcurve requires two conditions:

1. The orbital plane of the planet must be correctly oriented: for random orientations, the geometric probability is \(p = R_\star/a\). For a Jupiter (respectively an Earth) around a 1 \(R_\odot\) star, this probability is \(10^{-3}\) (respectively 0.5%).

2. The precision of the star’s photometry must be better than the depth \(\Delta F/F = (R_P/R_\star)^2\) of the transit. For a 1 \(R_\odot\) planet the drop is 1% (respectively \(10^{-4}\)).

In ground-based observations, the photometric precision is at best 0.1%. In space, it can reach a few \(10^{-5}\). Thus, while Jupiter-size planets can be detected in this way from the ground, one can detect Earth-sized planets from space. In fact, it is presently the only method capable of detecting and investigating further Earth-like extrasolar planets.

Late in 1999, the first planetary transit has been detected from the ground (around the star HD 209458), leading to the first determination of the radius of an extrasolar planet (Charbonneau et al. 1999). Also, the OGLE team has reported in early 2002 several tens of planetary transits candidates (Udalski et al. 2002); a few of them may indeed be confirmed as planets from radial velocity measurements. For more details, see the paper by H. Deeg at this Conference.

3. Direct imaging:
Although the most difficult, it is the most promising method. I will therefore describe it in more details. From the physical point of view, there are two kinds of emissions by a planet:

1. Reflected light:
The planet reflects the stellar light with a flux ratio given by
\[
\frac{F_{\text{refl}}(t)}{F_\star} = \frac{A_\text{pl}}{4} \times \left(\frac{R_\text{pl}}{a}\right)^2 \times \phi(t)
\]
where \(\phi(P,i,e,\omega,t)\) is an orbital phase factor (sinusoidal in case of a circular orbit) and \(A_\text{pl}\) the planet albedo. This ratio peaks at the same visible spectral range than the star itself and is typically \(10^{-5} - 10^{-6}\).

2. Thermal emission:
The planet, heated by the star at a temperature \(T_\text{pl} = T_\star \times (R_\star/2a)^{1/4}(1 - A_\text{pl})^{1/4}\), emits a thermal flux given by
\[
\frac{F_{\text{th}}}{F_\star} = \left(\frac{R_\text{pl}}{2a}\right)^2
\]
This ratio peaks at the mid-infrared and is typically \(10^{-7} - 10^{-7}\), i.e. about \(10^2\) larger than
in the visible range. There is no orbital phase factor here.

Note that the above formulae do not hold for non-spherical objects, i.e., planets with large Moons and planets with rings.

From an instrumental point of view, the planet is very difficult to detect by imaging because it is embedded in the diffraction halo of its parent star. The latter has an angular extension given, for a wavelength \( \lambda \), by

\[ \theta \approx \frac{\lambda}{L} \]

where \( L \) is telescope diameter. There are a number of “counter-measures” to circumvent this difficulty:

(a) \textit{Reduce} \( \lambda \).

But this approach is not compatible with the detection of the thermal flux \( F_\lambda \), requiring large \( \lambda \)’s.

(b) \textit{Increase} \( L \).

For very large \( L \), an interferometric architecture with several subapertures is required. A variant of the latter (named HyperTelescope), introduced Labeyrie (1996), makes use of “pupil densification” which brings the images of the separated entrance pupils co-adjacent in the focal plane without losing the initial angular resolution. It enables full snapshot imaging as for a monopupil telescope.

(c) \textit{“Turn off” the star (but not the planet):}

This can be done by interferometric nulling where the stellar path from two (or more) branches of an interferometer interfer destructively (Bracewell 1978). Since then, a series of coronagraphic masks have been proposed (Roddier and Roddier 1997, Guy and Rabbia 1996, Rouan, Riaud, Boccaletti et al. 2000 and Abe, Vakili and Boccaletti 2001); they do not require an interferometric architecture.

(d) \textit{“Smart” (non circular) apertures:}

Two approaches have been proposed: a square (apodized) telescope (Nisenson and Papaliolos 2001) or a Jacquinit pupil (Spergel 2001). In both cases, the stellar diffraction halo is not circular, but cross-shaped and the planet is visible when it lies in one of the diagonals of the cross.

Some of these approaches can be combined together.

3. CHARACTERIZING EARTH-LIKE PLANETS AND MOONS: REFLECTED LIGHT VERSUS THERMAL EMISSION

Assuming that the instrumental problems have been resolved, let us look with more details into the planets characteristics accessible by imaging.

1. Orbit:

If an object is detected close to the star, one wants to determine its orbit. Two orbital positions, together with the observation epochs, i.e., 6 observables \( t_1, t_2, x(t_1), y(t_1), x(t_2), y(t_2) \), are in principle sufficient to determine the 6 orbital parameters \( a, i, \omega, \Omega, T_0 \) and \( e \). But one has to verify that the object is not a close background star. This requires a third position measurement \((x(t_3), y(t_3))\) at a third epoch \( t_3 \).

This minimum number at least holds for the thermal emission which is independent from the orbital phase. For the reflected flux, one can take advantage of the phase dependance given by (1) to reduce this number to 2 orbital positions to deduce the orbit parameters. In that case it would indeed be unlikely that a background star had, by coincidence, a flux variation precisely given by (1). This reduction from 3 to 2 of the number of observations of a star required to assess the existence of a planet is important for missions scenarii and represents a significant advantage of observations in the visible.

2. Mass:

In principle it can be determined only from the dynamical perturbation of the star’s motion by the planet. Nevertheless, from a low spectral resolution spectrum (R=5) in the visible, one can infer whether there is a high, medium or low density atmosphere. From the latter, one can deduce, up to a factor 2-5 (Brown et al 2002), the mass of the planet (low mass planets do not retain their atmosphere, while high mass planets retain thick atmospheres). The thermal infrared is not suited for this type of studies.

3. Radius:

From formula (1) and from the fact that the albedo has an upper limit of 1, the visible flux gives an lower limit for the planet radius; unless a giant planet would have an albedo of 1%, it cannot be confused with an Earth-sized planet. The thermal emission gives, thanks to the formula (2) a safer value for the radius (unless the planet is surrounded by Moons (DesMarais et al 2001) or by rings (Schneider 1999).

4. Temperature:

For reflected light, it can be inferred from the star-planet distance and from the albedo through the relation \( T_\text{pl} = T_\star \times (R_\star/2a)^{1/2}(1 - A_\text{pl})^{1/4} \). But then one has in principle to know the planet albedo. Nevertheless, the latter formula shows that the temperature is not very sensitive to the albedo: a variation of \( A_\text{pl} \) from 0.3 to 0.7 gives an decrease of 20% for \( T_\text{pl} \). Here again, the thermal infrared gives a direct (and independent) measurement of \( T_\text{pl} \), safer than from the reflected flux.

5. Albedo colour \( A(\lambda) \):

The albedo can only be given by the reflected flux. But, as seen on formula (1), only the product \( A_\text{pl} \times R_\star^2 \) can be directly measured. Nevertheless, the measurement at different wavelengths gives the albedo...
colour, regardless of its absolute value. By itself this already constitutes a precious indication on the nature of the planet surface, as shown by Brown et al. (2002).

6. Albedo variations:
The formula (1) only gives the product $A_{pl} \times R_{pl}^2$. It thus does not enable to give the absolute value of the planet albedo. But from the time variation of $F_{refl}(t)$ one can, after correction of orbital effects, deduce the time variation $A_{pl}(t)$ of the albedo (since the planet radius is constant). It has 3, easy to disentangle, components:

- **A short term** (hours to days) periodic component: the latter would reveal the presence of surface inhomogeneities of the albedo by the modulation of $F_{refl}(t)$ due to the planet rotation. The period of the modulation gives the duration of the planet day, its amplitude gives the albedo contrast between different parts of the planet surface ("continents") and the shape of the modulation gives the spatial extension of "continents" (Schneider 1999, Ford, Seager and Turner 2001).

- **An annual periodic variation** gives the "seasonal" effects on the albedo value, due to the inclination of the planet rotation axis (like for the Earth) or to the orbit eccentricity.

- **Chaotic short term** (hours to weeks) variations would most likely be due to variations of the cloud coverage.

Of course, these times variations can be combined with their spectral aspects to give more insight in the planetary processes. The albedo is observable only for reflected light and is inobservable in the infrared where thermal emission is dominant.

7. Environment:

- **Atmosphere:**
  As already mentioned, the albedo colour gives the amount of Rayleigh scattering, and thus the density of the atmosphere (Brown et al. 2002).

- **Clouds:**
  As already mentioned, the most natural explanation of chaotic variations of the albedo would be a variable cloud coverage. Let us note that a similar chaotic variation can also be due to dust storms, like on Mars. In this case, the confusion with clouds can be removed by the colour characteristics of the albedo fluctuations: clouds have a white albedo, while dust is red.

- **Rings:**
  Their existence would be inferred from a non-Keplerian variation of the phase factor $\phi(t)$. Indeed, its standard mathematical expression $\langle \phi(t) \rangle = (1 - \sin \Theta \sin(2\pi t/P))/2$ in case of circular orbits) holds only for spherical bodies. In presence of rings, for half of the orbit, the observer sees only their backside, which is black, giving to $\phi(t)$ a more complicated expression depending on the detailed configuration of the rings (Schneider 1999). This case is not an exception, as shown by the Solar System planets; it is quantitatively not negligible since for instance the reflected solar flux from Saturn rings is as large as the planet reflected flux itself.

- **Moons:**
  They will most likely be first detected by the transit method (Sartoretti and Schneider 1999). For the coming generation of imaging space missions (e.g. Darwin/TPF), the angular resolution will not be sufficient to separate them angularly from their parent planet. It will nevertheless be possible to detect them by a photometric monitoring of the planet:

  - **Planet-satellite mutual transits** (Schneider 2002 b).
    A planet brightness drop with an amplitude $(R_{sat}/R_{pl})^2$ should appear with a period half the satellite revolution period. The geometric probability of this event nevertheless does not exceed $\approx 10\%$. The event is detectable in both reflected and thermal emission regimes.

  - **Planet-satellite mutual shadows** (Schneider 2002 b).
    It is most likely that the satellite orbits lies close the planet orbital plane. In that case, the satellite throws, once per orbit, a shadow on the illuminated part of the planet and, once per orbit, disappears in the planet shadow. An interesting feature of this event is that the satellite+planet flux drop has a very characteristic shape. In case of a satellite orbit lying exactly in the planet orbital plane, this shape is, for $\phi = \pi/2$, $\Delta F_{pl}/F_{pl}(\phi_{sat}) = \tan \phi_{sat}$, varying from 0 (when the satellite orbital phase $\phi_{sat} = 0$) up to a maximum $(R_{sat}/R_{pl})^2\sqrt{2}$ which is larger by a factor $\sqrt{R_{pl}/(2R_{sat})}$ than the drop due to mutual transits. For satellite orbits not lying exactly in the planet orbital plane, the evolution of the function $\Delta F_{pl}(\phi_{sat})/F_{pl}$ along the planet orbital revolution gives the two angular parameters characterizing the relative inclination of the planet and satellite orbital planes. In addition to being larger than mutual transits, mutual shadows have a geometric probability close to 1. This event can be seen only for reflected light, i.e. in the visible.

- **Magnetospheres:**
  At least 1 planets in the Solar System have a radio decametric emission stronger than the (quiet) Sun. It is probably due to a strong quasipolar magnetic field of internal origin, producing an extended magnetosphere in which particle acceleration takes place, inducing strong
non-thermal radio emission. Present theories and scaling law suggest that the radio flux is inversely proportional to \(d^{1/3}\) and proportional to \(R_{\text{pl}}^{1/3}\), and the intensity of the stellar wind (the latter being the \textit{ab initio} energy source). For a magnetic field of about 10 Gauss, the radio emission peaks in the decametric range, and in the metric wave-length range for higher magnetic field values (the latter being in fact less likely). The most favourable combination of the 3 mentioned factors leads to hope for decametric emissions 1000 times that of Jupiter. These would be detectable at 15 pc with the best receptors (Jupiter itself would be detectable at 0.5 pc). Planetary magnetic fields are usually attributed to turbulent rotation/convection motion of conducting material inside the core. Scaling laws predict magnetic field of a few Gauss for the already discovered extrasolar planets (Farrell et al 1998).

8. Surface properties:

- **Structures:**
  The case of continents and oceans has already been mentioned above for reflected light. In principle the modulation of the thermal emission by oceans and continents could also be detected during the diurnal planet rotation. But, while the continent/ocean contrast is about a factor 5 in reflected light, it is only 4/(T_{\text{ocean}} - T_{\text{cont.}})/T_{\text{mean}} \approx 10\% for thermal emission.

- **Internal heat vs/ stellar heating:**
  Depending on the orbital phase, the observer sees the illuminated side or dark side of the planet. There may exist a temperature contrast \(\Delta T_{\text{pl}} = T_{\text{pl.day}} - T_{\text{pl.night}}\) between these two sides. Along the orbital revolution it will provide an annual modulation of the effective planet temperature \(T_{\text{pl,eff}}(t) \approx (T_{\text{pl.day}}(1 - \sin i \cos(\pi t/P))/2 + T_{\text{pl.night}}(1 - \sin i \sin(\pi t/P))/2)^{1/4}\) (for \(e = 0\)). A low \(\Delta T_{\text{pl}}\) would mean a high atmospheric or oceanic circulation, while a high \(\Delta T_{\text{pl}}\) would mean a low lithospheric heat conductivity. For instance, for the Earth the day/night temperature difference is about 10 K leading to a relative thermal flux variation 4/(T_{\text{pl.day}} - T_{\text{pl.night}})/T_{\text{mean}} \approx 10\%.
  For the Moon and Mars the temperature difference is \(\approx 100\) K, giving a relative flux variation of a factor 2. The contrast \(\Delta T_{\text{pl}}\) can be due to several factors: surface (lithospheric and oceanic) thermal conductivity, oceanic and atmospheric circulation, and depends on the planet rotation rate. An additional source of thermal emission can be purely internal, due to tectonic activity and to rocks radioactivity. It could in principle produce a temperature in excess of the equilibrium temperature \(T_{\text{pl}} = T_e \times (R_e/2a)^{1/2}(1 - A_{\text{pl}})^{1/4}\). But the example of the Earth, for which the the tectonic and the

radiogenic heat flow is only 100 mW/m², compared to the \(\approx 1\) kW/m² heat flow produced by stellar heating, shows that this effect can be appreciable only for planets far away from their star, where the stellar heating is small as in the case of Europa. Together with the planet radius, mass (and thus the density), age, albedo, the measurement of the effective temperature and its modulation will provide precious constraints on the planet atmospheric, surface and internal structure. Of course, this measurement is possible only in the infrared regime.

9. Life?
What life is constitutes a philosophical problem far beyond the present scope. Let us adopt the traditional definition “Life” = far from equilibrium organic systems, transforming stellar light into complex organics. A traditional prerequisite is the presence of liquid water, imposing a planet temperature of about 300 K. The planet must therefore lie in the “habitable zone”, i.e. at a distance of \(\approx (T_e/T_0)^2(R_e/R_0)\) AU from the star (\(\approx\) from 0.1 to 1.5 AU, for M to F stars).

The detection of signatures of Life (“biosignatures”) makes use of two approaches:

- **“Dejecta”:**
  These are by-products of biological activity on the planet. The latter are mainly atmospheric gases such as O₂ (and its by-product O₃), CH₄. The key argument here is that on Earth all the molecular oxygen content of the atmosphere (20 %) comes from the photosynthetic activity of vegetation and bacteria. This argument is enforced by the fact that on Mars and Venus there is no oxygen or ozone. Since the main source of carbon for organics is the atmospheric CO₂, the latter must also be present in the planet atmosphere. The detection of O₂ is a priority in the sense that it gives an access to the degree of biological evolution on the planet (DesMarais et al. 2001, 2002). O₂ is detectable only in the visible, all the other gases are detectable in both visible and infrared regimes (DesMarais et al. 2001, 2002). These O₂ + O₃ signatures raise some questions (Labeyrie et al. 1999):
  - Only atmospheric abiotic processes of production of O₂ have been explored; there remains yet unexplored processes such as surface chemistry at the water-rocks-atmosphere interface at the planet surface.
  - It misses anoxicogenic photosynthetic organisms (Blankenship et al 1995).
  - It does not provide informations on the physiological mechanisms taking place.
  There is therefore some danger of triumphalism if someday oxygen is found. Only extensive investigations, including the full characterization of the planet, will validate the biogenic origin of oxygen.

- **“Vegetation”:**
  Whatever the detailed photosynthetic mecha-
nisms are, they must subtract energy from some part of the stellar spectrum reflected by the planet, leading to absorption features in this spectrum. This mechanism is responsible for the “red edge” at 750 nm in the terrestrial vegetation spectrum. The latter has been observed globally, for the first time, for the whole Earth seen as an unresolved source in the Earthshine spectrum (Arnold et al. 2002). The shape of this spectral feature gives some indication on the energy conversion mechanism, but the possible confusion with mineral absorption features has to be investigated further. It cannot be a safe biosignature by itself, it is useful only in association with other ones.

The best approach would be to combine dejecta-like and vegetation-like biosignatures.

The Table 2 summarizes the best wavelength regime for different planet characteristics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Visible</th>
<th>Infrared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Mass</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Temperature</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Albedo</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Day</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Seasons</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Clouds</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Rings</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Moons</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>O$_2$</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>O$_3$, CH$_4$, CO$_2$, H$_2$O</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Vegetation</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Intern. heat</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

It seems that more science can be done with reflected light observations, but it cannot do all of it and thermal infrared regime will provide important complements.

4. IMPLEMENTATION AND FUTURE PERSPECTIVES

4.1. Specifications of imaging detection facilities

In order to detect and characterize Earth-like planets and Moons according to the above objectives, a number of specifications of optical architectures are required:

**Collecting area:**
An Earth-like exoplanet around a main sequence star at 10 pc has an apparent magnitude around 25. A collecting area of at least 15 m$^2$ is necessary to detect it in 1 hour. For a monopupil telescope this means a diameter of at least 4 m.

**Spectral resolution:**
To detect the spectral bands of O$_2$ and O$_3$, a resolution of 20 is necessary.

**Angular resolution:**
In order to separate the planet from its parent star, an angular resolution $\lambda/L$ better than the star-planet angular separation $a/D$ is required. For a planet at $a = 1$ AU from a star at $D = 10$ pc, this requirement leads to a telescope diameter (or interferometer baseline) $L$ of at least 2m at $\lambda = 1 \mu$m (visible reflected light) and of at least 20m at $\lambda = 10 \mu$m (infrared thermal emission).

**Photometric precision:**
Objectives like albedo variation and Moons detection require a photometric precision of $\approx 1\%$ or better.

**Pointing flexibility:**
The photometric monitoring required for rings and Moons detection and the monitoring of albedo variation require a flexibility of the observation strategy sufficient to enable a frequent revisit on the detected exoplanets.

4.2. Ground-based programmes and space-based projects

The different detection and characterization methods have led to several ground-based programmes and space-based projects. Instead of describing all of them, I will discuss a few aspects relevant to Earth-like planets and Moons.

1. **Radial velocity surveys:**
Several high radial velocity precision surveys are underway (UVES and soon HARPS at ESO, HIRES at Keck). The present best achieved sensitivity is 2m/sec with the UVES spectrograph at ESO (Kürster et al. 2002). It is sensitive to a 10 $M_{\oplus}$ planet in the habitable zone of an M star. The minimal mass yet detected (June 2002) is a 36 $M_{\oplus}$ planet (Butler et al. 2002) around the G5 star HD 49674. The detection significance level is 18 $\sigma$; at a 6 $\sigma$ significance level, a 10 $M_{\oplus}$ planet would have been detected in similar conditions in the HZ of a K star. It is thus quite possible that super-Earths will be detected from the ground by RV in the coming years (Schneider 1998).

2. Transits:
The COROT mission (launch 2005, results 2007 - 08) will provide the frequency $f_\oplus$ of telluric planets:

$$f_\oplus = \frac{\text{Nbr. of stars with telluric planets}}{\text{Total nbr. of stars}}$$

It will more precisely give the distribution $f_\oplus(a, R_\oplus)$ of $f_\oplus$ up to $a \approx 0.1$AU for $R_{\text{pl}} = R_{\oplus}$, and up to $a \approx 0.3$AU for $R_{\text{pl}} = 2R_{\oplus}$. Unless there is a sharp
discontinuity above $a = 0.3 AU$ in $f_{\beta}(\alpha, R_{pl})$ a reasonable extrapolation of $f_{\beta}(\alpha, R_{pl})$ will give better than an order of magnitude for $f_{\beta}$ at the habitable zone. This result will constitute a master piece for the specifications of Darwin/TPF.

In 2009 - 11, the ESA mission Eddington and the NASA mission Kepler will provide a confirmation of and refinements to the COROT results.

3. Imaging and Biosignatures:
The requirements listed in section 4.1 have led to different types of space projects. The ESA Darwin project is an IR interferometer aiming to search for O$_3$ around terrestrial planets. Until 1999, NASA had an identical project, Terrestrial Planet Finder (TPF). In 1999, NASA started a re-evaluation of:
- Biosignatures
- The $\lambda$ range
- The optical architecture leading to two options:
  1) An IR interferometer aiming to search for O$_3$ around terrestrial planets, similar to Darwin
  2) A visible monopupil telescope with a "special optics to reduce the starlight by a factor of one billion, enabling to detect the faint planets."

NASA intends to choose between the IR Interferometer and the Visible Coronagraph around 2006.

Different new specific projects have meanwhile emerged:

- A 4 m UV-Vis Telescope + Coronagraphic mask (Brown et al. 2002) with the following performances:
  - An Earth-like planet is detectable in 1 hour (at 5 pc)
  - O$_2$ and O$_3$ are detected in 30 hours
  - "Vegetation" is detectable in 40 hours

- A 8m x 8m Apodized Square Aperture

- An Hypertelescope in a non-redundant linear array

- 3 "TPF-lite" precursors to TPF in the Visible:
  - Jovian Planet Finder (Clampin et al. 2002), a 1.5 m telescope with a coronagraph
  - Eclipse (Trauger et al. 2002), a 1.5 m telescope with a coronagraph and a wavefront correction system.
  - A 1.5 m class Apodized Square Aperture telescope

The situation of all these projects and perspectives is presently in permanent evolution. These different projects raise the question of an harmonization of the Darwin and TPF designs in case of a joint mission.

REFERENCES


Boss A., 1995, Science, 267, 360


Clampin M., Ford H., Illingworth G., et al., 2002, BAAS, 33, no. 4, 33.02


Labeyrie A., Schneider J., Boccaletti A. et al., 1999, ESA SP-451, 21


Mayor M. and Queloz D., 1995, Nature 378, 355


Roddier F. and Roddier C., 1997 PASP 109, 815


Schneider J., 2002b A&A, to be submitted
Spergel D., 2001 Applied Optics, submitted
Trauger J., Hull A. and Redding D., 2002, BAAS, 33, no. 4, 86.04
Udalski A., Paczynski B., Zebrun K. et al, 2002
Acta Astronomica, submitted
Wolszczan A., 1994 Science 264, 538
DETECTION OF TERRESTRIAL PLANETS AND MOONS WITH THE PHOTOMETRIC TRANSIT METHOD

Hans J. Deeg

Instituto de Astrofísica de Canarias, C. Via Lactea s/n, E-38200 La Laguna, Tenerife, Spain,
Email: hdeeg@iac.es

ABSTRACT

An overview is given over the photometric transit method for the detection of extraterrestrial planets. Discussed are: its basic principals, its derivable parameters, requirements for its implementation, and sources of confusion. Special emphasis is given space based experiments that may lead towards the detection of Earth-like planets. The detection of large moons may also be possible, either by their direct detection in the shape of a transit lightcurve, or by their influences on the mid-time of transit. For the latter point, estimations are given on the potential that will offer the upcoming three space missions, COROT by CNES, Kepler by NASA and Eddington by ESA for the detection of moons around extrasolar planets. By a similar principle, measurements of variations in the exact minimum time of eclipsing binaries are also expected to lead to a large number of detections of circumbinary planets.

1. INTRODUCTION

The possible existence of other worlds – that is, Earth-like planets - has been a source of inspiration for astronomers and lay-persons alike for centuries. While the large number of discoveries of extrasolar planets in the last few years [1] has been an enormous step in that direction, the ‘radial velocity’ method that is employed for these discoveries has fundamental limits that will not allow the discovery of planets smaller than about 10 Earth masses.

An alternative that has received increasing attention in recent years is the detection of planetary transits, where a planet crosses in front of its central star. By an external observer this is perceived as a slight dimming of the star that is being crossed. This ‘transit method’ was mentioned first already 50 years ago [2], though the first serious outline of the method did not appear until 1971 [3]. Further developments during the 1980’s and 90’s [4,5,6] culminated in the first proposal for a dedicated space mission[7].

* The same paper by O. Struve probably constituted also the probably first proposal to use radial velocity surveys for planetary detections, and even hypothesized on the presence of inner-orbit giant planets.

Especially since the discovery of the first transiting planet in 1999 [8,9], several ground based transit projects have been initiated, which are however mainly dedicated to the detection of short periodic ‘Hot Jupiters’ (for an overview, see [10]). Of more interest for the purpose of this article are however the experiments dedicated towards the detection of Earth-like planets. Incidentally, the first experiment employing the transit method was dedicated to just this. This was the TEP (Transits of Extrasolar Planets) project [11,12], in which the eclipsing binary star CM Draconis was surveyed during the years 1994-2000 for the presence of large terrestrial circumbinary planets, reaching a detection limit of 2.5 Earth diameters. Further ground based experiments towards Earth-like planet detection around M stars exist currently only as proposals, but several space missions are now approved for launch in the coming years. These should make the discovery of true Earth equivalents (in terms of size and distance from the central star, which also implies a temperature similar to Earth) around Solar-like stars possible until the end of the decade with the method of transits.

2. BASIC PRINCIPLES

![Fig. 1. Basic principle of the transit method. A planet transiting in front of its own central star will cause a small dip in the stellar brightness L*. δ is the transit's latitude on the stellar disk.
](image)

The basic idea of the transit method is shown in Fig. 1. The brightness variation during a transit, ΔL, and the star’s off-transit brightness L*, are approximately related to the radii of the planet and star by:
\[
\frac{\Delta L}{L} = \left( \frac{R_p}{R} \right)^2
\]

Equation (1) assumes a uniform surface brightness of the transited star. The exact shape of the brightness drop depends also on the latitude of the transit across the central star, and on the stellar limb-darkening. For a star with a limb darkening coefficient of 0.5, the maximum of the brightness drop is about 20% larger. Since the stellar limb-darkening is strongly dependent on the wavelength, a planetary transit event will also cause a characteristic, albeit small, color-change. If we assume the Earth-Sun system as typical, the observed brightness variation of an Earth-like planet would be about \( \Delta L/L = 8.4 \times 10^{-5} \). This brightness variation is much too weak to be detectable from ground based observatories, where limits are about one part in a thousand. This is the principal requirement that causes the need for space missions. Ground-based transit detections of terrestrial planets will therefore only be possible around M stars, which are smaller, and consequently, transits are deeper by an order of magnitude.

The duration of a transit, \( t_p \), is given by:

\[
t_p = \frac{T_p}{\pi} \left( \frac{R_c \cos(\delta) + R_e}{a_{pl}} \right),
\]

where \( \delta \) is the latitude of the transit on the stellar disk, \( a_{pl} \) is the major halfaxis of the planetary orbit, and \( T_p \) is the orbital period. Again for the Earth-Sun system, \( T_p \) for an external observer would be about 13 hours. The period \( T_p \) can trivially be determined if several transits at time intervals \( T_p \) are observed. In actual observing programs, the observation of several - at least three - transits is desirable, as the repeatability is a key feature to verify observed brightness drops as planetary transits. This requirement also makes space mission very attractive, since only they can ascertain observations over several planetary orbital periods without significant interruptions. A detailed example on the derivation of the various planetary and stellar parameters from ground based observations of a transit of HD209458 is given in ref. [13].

In order to produce transits, it is of course necessary that a planet-star system orbits in a plane that is within a small angle to the line of sight. Or, as seen from the perspective of the central star, from anywhere within the band in the sky that is subtended by its planet during one orbit, a transit can be observed. The probability for a randomly oriented planet-star system to produce transits is then given by:

\[
p_t = R_c/a_{pl}
\]

For the Earth-Sun system, this probability is 0.47%. Considering that Venus is another Earth-like planet with similar size, and its probability for alignment is 0.65%, the total probability for an external observer to detect any Earth-like planet around the Sun is about 1%. The probability to detect short periodic planets is of course larger, and in the case of Hot Giant planets it reaches over 5%. Another selection effect is given by the stellar populations that are being observed. Only main sequence stars are of interest in transit surveys, but they account only for about half of the stars within the brightness limits of most transit searches. Giant stars are too big and photometrically unstable to produced observable planetary transits. Thus, not more than one in 200 field stars can be expected to show transits of Earth-like planets - this if Solar Systems equivalents are very common. Consequently, thousands of stars need to be observed, causing the need for wide-field photometric cameras transit surveys.

Lastly, the problem of unequivocal detections of transiting planets should be mentioned. This becomes especially concerning in situations were the transit signal is relatively weak. The two principal issues here are: stellar micro variability causing transit like signatures, and eclipsing binaries masking as planet-star systems.

Regarding the first point, it has been demonstrated [14] that the two phenomena's different frequency domains allow in their separation, and hence variability is not expected to be a major obstacle in Earth-like planet detection. Also, the difference in colors between microvariability and transits will allow their discrimination where multi-color data are available.

Another source of confusion will be eclipsing binary (EB) stars. A grazing EB has a low-amplitude eclipse that may appear like a transit of a planet. Or, an EB with a large amplitude in the background of a brighter star could give a low amplitude signal in the summed light. These cases may be discriminated against transits by considering their period and duration, their shape, and their color, if such data are taken. Follow-up radial velocity observations would also reveal grazing EB's, whereas optical data with high spatial resolution may resolve background EB's. Some number of false positives in transits of weak signal may however be unavoidable.

* This scenario may be frequent in the planned space missions, as in all of these stellar light will be spread over several arcseconds in order to improve the dynamic range of the CCD cameras against saturation.
2. SPACE MISSIONS FOR THE DETECTION OF TERRESTRIAL PLANETS

Three space missions are currently in preparation, and all are approved for launches between 2005 and 2008 (Table 1). The first one is the French-led COROT mission[15,16], which contains a small 20 cm telescope dedicated to asteroseismology (AS) and extrasolar planet (EP) detection. COROT will take AS and EP observations simultaneously in stellar fields that are side-by-side in the sky, with two sets of CCDs, each one optimized to its specific task. With COROT’s maximum observing duration of 5 months on any single field, only short periodic planets will cause 3 observable transit events. Therefore, a prism that gives 3-color photometry has been added, with the intention to recognize planets from observations of single transit events, due to the unique color signature of transits. COROT has therefore some sensitivity to longer periodic planets as well, albeit less information will be gained on these. Two larger space missions are then foreseen for the years 2006-8: Kepler by NASA and Eddington by ESA. Kepler[1718] is the only mission dedicated exclusively to the detection of Earth-like planets. It will observe 100 000 K dwarf stars in a single large field in Cygnus over 4 years. Lastly, ESA approved recently (May 2002) the mission Eddington[1920] for a launch not later than 2008. Current studies for Eddington center on a design with 4 Schmidt telescopes of about 60 cm diameter. Observations in AS and EP are to be taken sequentially, with 2 years dedicated to AS and 3 yrs to EP, though of course observations for one topic may benefit the other. Like COROT, Eddington will probably employ some form of multi-color photometry. In Eddington’s case this is however motivated by the improved discrimination against false positives, as mentioned in the previous section. The number and wavelengths of colors bands in Eddington is still to be determined. In Fig. 2 an overview is given on the expected planet detection capabilities of these space missions.

---

Table 1: Overview on transit space projects

<table>
<thead>
<tr>
<th>Name</th>
<th>COROT</th>
<th>Kepler</th>
<th>Eddington</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>CNES (F) + partners</td>
<td>NASA</td>
<td>ESA</td>
</tr>
<tr>
<td>Launch date</td>
<td>2005</td>
<td>2007</td>
<td>2007-8</td>
</tr>
<tr>
<td>Mission duration</td>
<td>3 yrs</td>
<td>4 yrs</td>
<td>5 yrs (3 yrs for EP)</td>
</tr>
<tr>
<td>Telescope</td>
<td>20 cm</td>
<td>95 cm</td>
<td>4 x 60 cm*</td>
</tr>
<tr>
<td>Field of view</td>
<td>4 deg² (EP)</td>
<td>105 deg²</td>
<td>36 deg²*</td>
</tr>
<tr>
<td>Color-Bands</td>
<td>3</td>
<td>1</td>
<td>1-4*</td>
</tr>
<tr>
<td>Nr. of stars obs’d for EP</td>
<td>12 k</td>
<td>100 k</td>
<td>50 – 100 k</td>
</tr>
</tbody>
</table>

* Preliminary or still under discussion

---

Fig. 2. Masses and orbital distances of planets that are accessible for detection by the transit space mission, in comparison to ground based radial velocity measurements and the SIM proposal for astrometric planet detection. The habitable zone in terms of planet mass and distance is given for a solar-like star.

3. ANALYSIS OF PLANETARY ATMOSPHERES

During a transit, a planet’s atmosphere will appear as a partially opaque ring for an external observer. The scale height of this ring is of course wavelength dependent, giving therefore the planet an effective radius that is a function of wavelength. Precise transit observations with spectroscopes - comparing transit depth across different wavelengths- or by photometers with filters may therefore give an indication about a planet’s atmospheric composition. Studies on the effects of Hot Giant planet atmospheres onto transits have been published by [21,22,23], and a possible absorption by sodium from the planet around HD209458 has been reported [24] from observations with the Hubble Space Telescope’s STIS spectrograph. In that case, the transit at the sodium feature’s wavelength is reported to be about 2 x 10⁻⁷ times deeper than in adjacent bands. This small value makes it clear that detection of atmospheric features around terrestrial planets is going to become a very difficult task. Additionally, the detection of wide-band features -or colors- is complicated by the color dependency of transits depths and shapes due to the wavelength dependent stellar limb darkening [11,25].

An inquiry has been made if O₂ - an indicator for potential life- could be detected in transiting Earth like planets [26]. The result is that 8m class telescopes could detect the O₂ A-band absorption feature for such planets around main-sequence M stars of 10th magnitude. There are only few M stars with that brightness, but with significantly larger telescopes, or future improved instrumentation, O₂ detections may be extended to M stars of 13th magnitude.
4 TIMING OF TRANSIT OBSERVATIONS - POSSIBILITIES FOR FURTHER DETECTIONS

Special consideration we give here to the results that may be achievable from the precise timing of eclipse-like features in a lightcurve - be it an 'eclipse' of a binary star system, or a 'transit' from a star-planet system. In both cases, an unseen third body will cause the eclipsing system to be offset from the 3-body barycenter. The precision $\delta t_0$ of the measurement of the time $t_0$ of minimum light in a lightcurve dominated by white Gaussian noise can be estimated as follows:

$$\delta t_0 = \delta_L \left[ \sum_i \left( \frac{\partial L(t_i, t_0)}{\delta t_0} \right)^2 \right]^{-1/2}$$

(4)

where $L$ is the flux from the object at discrete times $t_i$ (the lightcurve) and $\delta L$ the relative error in the measurement of $L$. For lightcurves with equidistant points $t_i$, where $\Delta t = t_{i+1} - t_0$, and with the relation against a variation in the minimum time $t_0$,

$$L(t_i, t_0 + \Delta t) = L(t_{i+1}, t_0),$$

a simple calculation of the derivative of Eq.(4) is possible from a single lightcurve, with

$$\frac{\partial L(t_i, t_0)}{\delta t_0} = \frac{L(t_i, t_0) - L(t_{i+1}, t_0)}{2\Delta t}.$$  

(5)

As an example from ref. [27] (note that Eq. 3 in that reference contains an error and is corrected by above Eq. 4), the precision in minimum timing measurements of the eclipsing CM Draconis is about 1.5 seconds, assuming a relative photometric error of $\delta t_0 = 0.01$ and a sample time of $\Delta t = 5$ sec.

4.1 Planets around eclipsing binaries

Any wide field photometric survey for the detection of transits will unavoidably detect a substantial number of eclipsing binaries (EBs). Previously, EBs were mentioned as an undesirable source of confusion against real planetary transits - they do however also offer a further way to detect planets. Around close binaries there are stable orbits for planets if their half axis exceeds about 3 times the binary components' separation. Orbiting planets are then offsetting the binary star around the barycenter common to all 3 bodies. Consequently, the variations in light travel time from the binary will then make its eclipses appear earlier or later, with a periodicity given by the third body's orbital period (Fig. 3). The amplitude of the timing offset $\Delta t_0$ is given by equation (6):

$$\Delta t_0 = M_p a_p / M_* c,$$

(6)

were $M_p$ and $M_*$ are the masses of planet and star, and $c$ is the speed of light. It is noteworthy, that these planet detections do not require the presence of planetary transits. However, similarly to the radial velocity method, the planet's mass can only be given as $m \sin i$, where $i$ is the planet's orbital inclination. For planets around an EB it can however be expected that they orbit close to the EB's orbital plane, which is always close to $i = 90^\circ$, and hence $\sin i$ is also close to 1.

Observations of eclipse minima with a timing precision of about 6 seconds have already led to the establishment of a lower limit of 1-3 Jupiter masses for the presence of long periodic planets (500-2000 days; the mass limit is independent of the orbital period) around the EB system CM Draconis [28]. The upcoming space missions will of course provide much lower detection limits, and due to their photometric precision, sub-second minimum timing precision can be expected from them (Fig. 4). The lower limits for planet detection with these missions will be given by the brightest EB's they can observe without saturation (11-12 mag for Eddington).
Fig. 4. For the three upcoming space missions, the eclipse timing error of a typical eclipsing binary system (assuming an eclipse depth of 45% of off-eclipse brightness) is given on the left, in dependence of the stellar brightness. For an example system with a mass of 2 $M_{\odot}$ and for planets with orbital periods of 150 days, the detectable planetary mass is indicated on the right. The calculated timing error is based on photon noise dominated data.

Fig. 5. The timing precision that can be achieved from the observations of giant planet transits ($\Delta B/B_0 \approx 1\%$) by the space missions. To detect a moon in the example of the Saturn-Titan system as a 3 sigma detection, a timing precision of 10 seconds needs to be obtained (bold horizontal line and arrow).

4.2 Detection of Moons

With precise timing, planetary transits may also indicate the presence of moons around their planets. For one, a large moon may be visible directly in the shape of a transit [29]. This can be expected, if the ratio in sizes between moon and planet is not a great one. Another effect is the deviation of the center time of a transit from strict periodicity, as the planet -which will cause the principal transit- will be offset from the planet-moon barycenter, and may lead, or trail the barycenter in
individual transits. The observation of repeated transits with high timing precision may therefore indicate the presence of moons. As an example, the Earth leads or trails the Earth-Moon barycenter on its orbit around the Sun by up to 2.5 minutes. Similarly, Saturn is offset by its largest moon, Titan, by up to 30 seconds. Whereas the ratio in cross sections between Titan and Saturn is about 1:550 and hence a direct transit signal from a Titan-like moon cannot be expected to be observable, future space mission may well be able to detect offsets in transit times by 30 seconds (Fig. 5), hence leading to the expectation that some Galilean-like moons may be detected. Such a detection will of course need at least 3, but preferably more minimum-timing measurements to establish a periodicity in the timing offsets. With the space missions Kepler and Eddington, and their multiyear observing runs, moons around planets with orbital periods of up to 200 days may however well be detectable.

5. SUMMARY
Three upcoming space missions, all of them approved for launch before 2008, lead to the expectation, that a significant number of Earth-like, and more general, terrestrial planets will be known within a decade. The possibility to obtain precise timings of eclipses from planetary transits and eclipsing binaries alike gives the opportunity to detect large moons around giant planets in the first case, and to detect non-transiting planets around close binary stars for the second case. Both of these possibilities may significantly enhance the scientific return of these space missions.

REFERENCES
2 Struve, O., Proposal for a project of high-precision stellar radial velocity work, The Observatory, 72, 199-200, 1952
15 http://www.astrosat-nrs.fr/projects/corot/
18 http://www.kepler.arc.nasa.gov
19 F. Favata et al. (Eds.), Proc. of the First Eddington Workshop, Cordoba 2001, ESA publ. SP-485, 2002
20 http://sci.esa.int/home/eddington/
25 Jha, S., Charbonneau, D., Garnavich, P.M. et al., Multicolor Observations of a Planetary Transit of


HYPERTELESCOPES AND EXO-EARTH CORONAGRAPHY

Antoine Labeyrie

Collège de France, Observatoire de Haute Provence & Laboratoire d'Interférométrie Stellaire et Exoplanétaire (LISE), Observatoire de Haute Provence, 04870 Saint Michel, France, labeyrie@obs-hp.fr

ABSTRACT/RESUME

Hypertelescopes, or "multi-aperture densified-pupil imaging optical arrays" can provide direct images with full luminosity at arbitrary aperture dilution. They can be equipped with coronagraphic masks for difficult imaging situations such as the search for exo-Earths at infra-red and visible wavelengths.

For detecting faint planets, down to $10^{-10}$ relative intensity, multi-stage coronagraphs now appear feasible, both for telescopes and hypertelescopes.

A 37-element hypertelescope has been proposed to NASA for TPF. Because the high-resolution image separates the planet's peak from most of the zodiacal and exozodiacal background, sensitivity is gained with respect to planet detection schemes using a beamsplitter for nulling.

For resolving continents, and other details such as the "green spots", of an exo-Earth at 3 parsecs, a second generation hypertelescope will require more than 100 apertures of 3 m, arrayed as a 100 km diluted aperture.

1. INTRODUCTION

Only a few years ago it was realized that large optical interferometers can produce direct high-resolution images with full luminosity if they have enough sub-apertures [1],[2]. It further appeared that the "densified-pupil multi-aperture imaging interferometers", or hypertelescopes, thus defined can support established coronagraphic techniques and their recent variants [3].

2. PRINCIPLE OF HYPERTELESCOPES

Highly diluted many-element apertures can produce direct images according to the classical and simple Fizeau scheme, but the image becomes difficult to exploit efficiently at high dilution since the central interference peak gathers only a very small fraction of the energy. Most of it is dispersed in numerous sidelobes. With additional optics producing a densified exit pupil (fig.1), where the pattern of centers is preserved, a full-luminosity image is obtainable, although in a small field.

Figure 1: Hypertelescope principle. The Fizeau image at the prime focus of a phased optical array (left, sketched as a segmented lens) has numerous sidelobes if the aperture is highly diluted. With a pupil densifier (right), having an array of Galilean telescopes, light is concentrated in the central interference peak. Off-axis stars provide a displaced peak, up to a limit, which defines the Zero Order Field.

Its dynamic range can be high if there are many aperture elements, accurately phased. If these are periodically arranged as a grid, with square or hexagonal pitch, the exit pupil can be completely densified. As seen from the image plane downstream, it does not differ from a monolithic pupil, and can similarly allow coronagraphic nulling by inserting suitable masks in the image and in a relayed pupil.

The elementary field or "Zero Order Field" (ZOF) has a sky extent of angular size $\lambda/s$ where $\lambda$ is the wavelength, and $s$ the center-to-center spacing of entrance sub-apertures.

At given collecting area and array size, the size and spacing of sub-apertures decreases indefinitely if their number is indefinitely increased. The ZOF extent then also increases indefinitely, and the hypertelescope becomes equivalent to a filled giant telescope of identical size, with transmission attenuated as the relative collecting area.

---


© European Space Agency • Provided by the NASA Astrophysics Data System
These properties of hypertelecope imaging have been largely verified with simulations and small-scale testing on the sky [6].

2.1 Possible designs on Earth

On Earth, flat sites such as some of the salt lakes in Chile favor telescope arrays as large as perhaps 10 kilometers, according to the Optical Very Large Array concept, now up-dated to incorporate a pupil-densifier in its beam combiner. The forthcoming interferometric availability of ESO's four 8m telescopes in Chile is an interesting opportunity to achieve hypertelecope imaging within a very narrow field, and coupled with coronagraphy.

The expensive delay lines or telescope motions required for large flat arrays can however be avoided with concave array geometries, using a crater or large sink hole where mirror elements are arranged as a giant but diluted mosaic mirror [3], [4]. To avoid suspension cables such as those of the Arecibo radio telescope, balloons can carry cameras at the focal sphere where Fizeau images are formed. This requires correction of the spherical aberration from the virtual primary mirror, using for example Mertz-type optics. A pupil densifier follows, unless the crater can be densely paved with mirrors, which is unlikely for cost reasons at the scale of 1 to 5 kilometers. This so-called "Carolina" design, borrowing its name from a composite alpine flower, is proposed for "exploded" versions of the Extremely Large Telescopes currently studied by different institutions. At equal collecting area, they provide the same limiting magnitude and higher resolution. The elementary field is much smaller, but a mosaic of pupil densifiers can be arrayed across the diffraction-limited field of the focal corrector. Exposures then directly provide a dilute mosaic of sky images. And several balloons, independantly tracked, can carry as many focal packages providing independent science.

Following early stability tests of balloons made at Haute Provence a few years ago, more detailed testing with tethered balloons of 6-12m size is scheduled in the coming months.

2.2 Possible designs in space

In space, the prospect of using multiple free-flyer elements [4] and accurately positioning them is of obvious interest for Carolina-like structures. "Virtual craters", as large as hundreds or thousands of kilometers, may be considered. The luminance of the very few known optically emitting neutron stars is so high that optical hyperteleopes with size approaching a million kilometers, which can in principle resolve them, will still detect a practicable photon rate per resel (resolved element). Current studies of huge, laser connected, space arrays such as LISA for detecting gravitational waves suggest that such extreme hyperteleopes are also feasible with available technology.

In space, radio interferometry can also exploit the hyperteleope principle. With the multi-pixel detectors now developed, the direct image formation improves the sensitivity since the interference peaks formed by many apertures are intensified as N^2, if N is the number of apertures.

Following the early proposals[4] for optical interferometers based upon a flotilla of free-flyers, the MOFFIT study commissioned by ESA explored their feasibility. This led to the DARWIN concept of Léger & Mariotti, and the variants considered by NASA for its Terrestrial Planet Finder.

When the principle of hypertelecope imaging became understood in 1996, such versions of DARWIN and TPF having more aperture elements of smaller size were also proposed [2], [3]. Among these possible hypertelecope versions, the Exo-Earth Discoverer (EED) [3] concept (fig.2) has a dilute primary mirror, which is a spherical mosaic of mirrors much smaller than their spacings. One or more focal combiners
move independently along the focal sphere to acquire and track one or more objects. The wide primary field, tens of degrees, relaxes the global pointing requirements for the flotilla. Thus, the mirror elements may be driven by low-thrust, low-power thrusters such as the small rigid solar sails proposed to ESA in 1985 [5], [12] and now studied in more detail by R. Angel and S. Errico (private communication) for NASA.

The spherical primary mosaic requires correctors for spherical aberration at the beam combiners. Their minimal size is about 1% of the primary at F/2 effective focal ratio, and increases very fast beyond. We have used Mertz's algorithm to find two-mirror solutions correcting also coma.

The corrected image is fed into a pupil densifier, typically a pair of micro-lens arrays (fig.1) or a reflective equivalent. Also, separate densifiers can be arrayed at intervals of lambda / d to enlarge the narrow field of direct imaging.

With an added coronagraph, as discussed in the following sections, the EED design allows exo-planet detection. Its detection sensitivity is improved with respect to the original DARWIN and TPF schemes since the direct high-resolution image obtained provides a better separation of the planet from the background nebulosity (zodiacal, exo-zodiacal, thermal emission from mirrors).

3. CORONAGRAPHIC USES, INFRA-RED AND VISIBLE

A coronagraph can be added to hypertelescopes. With a fully densified exit pupil, in particular, the coronagraphic optics behaves like with a conventional telescope.

3.1 Possible designs for coronagraphs

A many-element hypertelescope can provide a densified exit pupil resembling for example the pupil of one Keck telescope. A classical configuration for coronagraphy uses the Lyot scheme, with a field lens carrying a field mask, and a Lyot stop in the relayed pupil to remove residual starlight diffraction by the field mask outside of the geometric pupil. The opaque disc originally used by Lyot as a field mask for a highly resolved source, the Sun, cannot work properly if made smaller than three or four diffraction rings. The newer types of phase masks introduced in the last few years can in principle make planets detectable closer to the star, but require that it be unresolved.

3.2 Active multi-stage coronagraphy

Efficient coronographic nulling of a star's peak and its diffractive splash across the image requires highly accurate control of the wavefront bumpiness, at the nanometer scale in visible light. Such control is difficult to achieve, even with active optics, and coronagraphic gains have since Bernard Lyot been limited to about 10^6.

It now appears that multiple coronagraphic stages, each equipped with an active loop, can be conceived to reach much higher gain values, towards the 10^10 Sun/Earth luminosity ratio.

At the exit of a telescope or hyperteleoscope with coronagraph, the image's residual star light is speckled if the star is unresolved or nearly so, as assumed in the following. It can be shown that the phase variations from speckle to speckle are randomly distributed between 0 and 2\pi, in spite of the typically much smaller phase residues on the wavefront up-stream from the coronagraphic masks. This is hardly surprising since the classical Airy rings at the focus of a perfect wavefront themselves have alternate phases of 0 and \pi.

By actively correcting the speckle phases to make them uniform, a highly constructive interference peak can be generated in the far field. Masking this peak removes most starlight from the beam.

The principle can be implemented, as sketched in fig. 3, with one more stage of relaying optics. It has a deformable mirror or transparent plate, for phasing the image's stellar speckles and concentrating most star light in a single speckle, i.e. a peak appearing as a "ghost star", in the following pupil plane. An occulting mask OM, much smaller than the pupil, removes it. This mask may be opaque or be a Roddier & Roddier phase-shifting dot. Both tend to also diffract some starlight outside of the camera field, relayed from the field lens FL. This further attenuates the spurious stellar illumination in the relevant part of the camera image, as happens within the Lyot stop of an ordinary coronagraph. In this respect, the relay optics indeed behaves like a coronagraphic stage, although the usual pupil and image planes are here interchanged.

The planet's peak relayed on C is little affected since: 1- the phase correction does not affect much the local phase uniformity within the planet peak since the scale size of deformations on the active element matches the speckle size, equal to the planet peak's size; 2- the mask OM is small with respect to the pupil, uniformly illuminated by the planet's light.

© European Space Agency • Provided by the NASA Astrophysics Data System
Figure 3: Additional stage of coronagraphy for coronographic telescopes or hypertelescopes. The cleaned image CI, output of the first coronagraph (not shown), is relayed by field lens FL and lens L to the camera C. The active corrector AC equals the phases of the star's residual speckles, so as to generate an interference peak or "ghost star" on the occulting mask OM, which therefore removes most stellar energy from the camera image. A wave analyzer, not shown, measures the stellar phase pattern in CI. Several such stages can be cascaded for deeper nulling, as long as the last one receives enough stellar photons to activate the active correction.

In principle, further nulling is achievable by iterating the scheme, with similar additional stages. A practical limitation to the number of such nulling stages, which can be cascaded, is the need for enough residual stellar photons in the last wave analyzer. A few detected photons per speckle and spectral channel are needed in each exposure for the moderately accurate phase measurements needed. At such levels, even fewer stellar photons are left in the following image, and they do not affect much the planet detection.

The starlight concentration in the pupil peak would be slightly improved by uniformizing not only the speckle phases but also their intensities (this is feasible to some extent with reverse phase-contrast, transforming pure intensity patterns into pure phase patterns. A second phase corrector would then be required). But the limited improvement does perhaps not justify the added complexity.

As an example, \(14 \times 10^9\) photons/s are detected in the visible from a 40 square meters aperture, either conventional or in hypertelescope form, on a \(m_\star=3\) star, if the combined transmission and quantum efficiency is 20\% and the bandwidth 400 nm. A terrestrial planet \(10^{10}\) times, or 25 magnitudes, fainter thus provides 1.4 detected photon/s in its image peak. With cascaded coronagraphic stages, the coronagraphic gain can in principle be pushed to \(10^5\) thus leaving 14 detected stellar photons per second in an average image resel. This is enough to measure the phase in each residual stellar speckle, so that it is corrected by one more stage of relay optics.

The color-dependance of the star's speckle pattern, and its phase, limit the spectral bandwidth, unless separated channels are arranged. Within a spectral band, a linearly approximated correction to a speckle's phase variation can be achieved by setting the deformable plate locally to provide a suitable optical delay. With phases \(\phi_1\) and \(\phi_2\) measured at two wavelengths \(\lambda_1\) and \(\lambda_2\) the delay \(\delta\) needed is determined by the conditions:

\[
\phi_1 + 2 kn = 2 \pi b / \lambda_1, \quad \text{and} \quad \phi_2 + 2 kn = 2 \pi b / \lambda_2.
\]

A wavefront analyzer, which does not have to be highly accurate, is needed to measure the phase of each stellar speckle in the entrance plane. In the last stage, the imaging camera C can itself serve as a wave analyzer, especially if mask OM is a Roddier & Roddier phase dot. Relevant information can be extracted from the image and its peripheral pattern, diffraacted by this mask, using phase-diversity methods. Known deformations applied sequentially to the plate AC can be used in this context. Just measuring the integrated peripheral intensity allows phasing by trial-and-error, to maximize this intensity. Optimal algorithms have to explore. In previous coronagraph stages, diverting precious planet light from the main beam should preferably be avoided, but sensors can exploit the peripheral image field, diffracted beyond the geometric field.

In addition to the higher coronagraphic gain reachable in principle with multi-stage coronagraphy, the tight wavefront tolerances can be relaxed. Optimal trade-offs between accuracy and the number of stages will have to be found.

Such multi-stage coronagraphy is applicable to conventional monolithic telescopes as well as hypertelescopes. In the latter case, the size of the sub-apertures often suffices to resolve the planets from their parent star. If so, it is of interest to install one or more stages of coronagraphy separately in each sub-beam, upstream from the beam combiner. If the star is unresolved or moderately resolved by the global aperture, more stages can also be useful downstream, especially if the pupil is fully densified.

4. STEPS TOWARDS "EXO-EARTH IMAGER"

Following the construction of ground-based hypertelescopes, or at the same time, a critical step towards space versions will be the experimentation of formation flying. Once mastered, the technique will become quickly scalable for arrays of any size and with any number of free-flyers. In addition to the
space precursors considered for DARWIN/TPF by ESA and NASA, a precursor for a hypertelecope version has been formally proposed to ESA [7].

Figure 4: Simulated 30 mn exposure of an Earth at 3 parsecs, using a 150 km EEI hypertelecope. It has 150 apertures of 3m, arranged as 3 concentric rings of 50 apertures. Starlight (not included in the simulation) is assumed to be removed with two or three stages of pre-combiner coronagraphs. Spectro-imaging with such angular resolution is of interest to search photosynthetic patches, their seasonal variations, and correlations with cloudiness. Such variations can probably discriminate abiotic mineral colors from those induced by photosynthetic life.

In its studies of the Terrestrial Planet Finder, NASA has included hypertelecope versions, studied by Boeing/SVS jointly with several groups. Their sensitivity advantage in the infra-red [8],[9] persists with as few as 6 apertures, according to the densification / re-dilution scheme of Roddier and Guyon [10]. At equal collecting area, the 37-element EED version has smaller mirrors, which may be easier to build. Simultaneous visible and infrared observing of extra-solar Earths is in principle possible.

Following this generation of hyperteleopes, smaller than a kilometer, EEI versions larger than 100 km will be needed to provide resolved images of exo-Earths. Simulations (Fig. 4) of the hypertelecope image formation in such cases [1] have shown that about 150 apertures of 3m size, arranged in three concentric rings forming a 150 km hypertelecope, provide a usable image where continents, large cloud formations, and "green spots" such as the Amazon basin become visible in a 30 mn exposure. For an exo-Earth at 3 parsecs, $10^{10}$ times fainter than its m,=3 parent star, the image is resolved in 32x32 resels, with 17 photons detected in each, assuming 20\% efficiency and a 400 nm visible bandwidth. In those resels where there is a substantial photosynthetic activity, the photon count can suffice for detecting the absorption bands with low-resolution spectroscopy [11].

Since the 3m sub-apertures resolve 1 A.U. at 30 parsecs in the visible, pre-combiner coronagraphy with several stages is preferable, as discussed in section 3.2. The visible range is of obvious interest for exo-chlorophyll detection in green spots, although any color can in fact be expected. Two or three stages of pre-combiner coronagraphy can bring the residual starlight level close to the level of the planet image. Installing all coronagraphy stages before the combiner gives more freedom for optimizing the aperture pattern, and also avoids the problem of dealing with incoherent residues from the highly resolved star in the combined image.

5. CONCLUSIONS AND FUTURE WORK

Hypertelecope architectures have attractive characteristics for powerful imaging instruments in space. Equipped with coronagraphic stages, which can be cascaded for extreme nulling of the stellar contamination, they allow in principle the detection of Earth-like planets, even at visible wavelengths. In the infrared range considered for DARWIN and TPF, hypertelecope versions such as the EED have a theoretical sensitivity gain, decreasing 10 to 100 times the observing time needed for detection.

A later generation, larger than 100 km and combining more than 100 apertures of 3m, will in principle provide resolved images of exo-Earths, where "green spots" will be detectable.

Following much verification work, with numerical simulations, laborotory models, and miniature hyperteleopes [6] pointed at real stars, the time has come for building precursor hyperteleopes on Earth and in space. Several projects are considered, jointly with other groups.

6. REFERENCES


SEARCH FOR SIGNATURES OF LIFE ON EXOPLANETS

Franck Selsis
Centro de Astrobiología (CSIC-INFA), Ctra. de Ajalvir km 4, 28850, Torrejón de Ardoz, Madrid

ABSTRACT

The spectroscopic characterization of Earth-like exoplanets is one of the goals shared by two major projects: Darwin (ESA) and TPF-Terrestrial Planet Finder (NASA). By producing low resolution spectra of the planetary thermal emission, they will be able to detect the 9.6 μm band of O2 which may be the signature of an O2-rich atmosphere sustained by a biological activity. In this paper, we review several studies focusing on the possible use of this ozone signature as a biomarker for terrestrial exoplanets. First, we stress the risk of false positive detection produced by abiogenic photochemical production of O2 and O3 and we show how to filter these imposters out by selecting only the simultaneous detection of O2-CO2-H2O (triple signature). In a second part, we investigate the false negative cases: when O2-rich atmospheres do not exhibit the O3 feature. This happens when the partial pressure of CO2 is too high or, unexpectedly, when the O3 layer is too dense. Finally, we present other eventual biomarkers that could trace life where O3 fails.

1. INTRODUCTION: THE SEARCH FOR O2-PRODUCING ECOSYSTEMS WITH DARWIN/TPF

The existence of other planetary systems has only been proved during the last decade, first around pulsars (Wolszczan & Frail, 1992), then around solar type stars (Mayor & Queloz, 1995). In this latter case, the discovered planets are massive (of the order of a Jupiter mass) and likely to be giant gaseous planets, which at least proved to be the case for HD 209458b (Charbonneau et al., 2000). Up to now, 92 planets with masses above 0.2 Jupiter mass have been found around 79 main-sequence stars (1).

The search for smaller planets of terrestrial type rises considerable scientific and philosopical interest, but is technically much more difficult. The space observatories under study by ESA (Darwin; Léger et al. 1996) and by NASA (TPF; Beichman et al. 1999)


1


© European Space Agency • Provided by the NASA Astrophysics Data System
negative result (when a planet contains life forms but does not produce any characteristic signature).

2. ABIOTIC SYNTHESIS OF O₂ AND O₃ AND "FALSE POSITIVE" DETECTION

The problem of abiotic production of O₂ with regard to the relevance of O₂ as a tracer of life has been addressed by several previous authors. Léger et al. (1999) have rejected the hypothesis of an oxygen-rich atmosphere from infalling comets, answering questions stressed by Noll et al. (1997). Ollivier (1999) proposed to search simultaneously for H₂O and O₃. Schindler and Kasting (2000) have discussed the false positive problem in the frame of radiative transfer computations of IR methane and ozone lines, but without modelling the photochemical production of these species. The work of Kasting (1988) on the evolution of Venus atmosphere stresses a case where an atmosphere can transiently be rich in O₂ through abiotic phenomena whereas having both H₂O and CO₂ present in detectable quantities. Chassefière and Rosenqvist (1995) made a pioneer study of O₂ production rates in CO₂ dominated humid atmospheres, and applied it to early stages of Mars.

Apart from the Earth, O₂ and O₃ are present in the atmosphere of other bodies of the Solar System (see table 1). In all cases its presence can be attributed to abiotic processes, mainly CO₂ and H₂O photodissociation.

2.1. Modeling

To simulate the atmospheres of possible terrestrial exoplanets, we used the model PHOEBE developed by Selsis (2000; 2002). It is a one dimensional code that computes from initial atmospheric conditions and for a given stellar spectrum the abundances for each chemical species as a function of altitude and time. It also gives the new temperature profile in radiative-convective equilibrium with a given chemical composition and the thermal emission spectrum at a 1 cm⁻¹ resolution. A full description of the model and of its parameters, like the complete list of the 150 chemical reactions included, as well as the validation of the code through modeling of Earth’s and martian atmospheres, can be found in Selsis (2000).

To investigate the abiotic synthesis of O₂ and O₃ in planetary atmospheres we simulated three general types of atmospheres: humid CO₂-dominated, dry CO₂-dominated and water-rich atmospheres. The results of these simulations are presented in the table 2.

2.2. Atmospheres containing both CO₂ and H₂O

In these simulations, the water vapor abundance at the surface was taken close to its saturation value. The most important reactions for abiotic production of O₂ in those dense and humid CO₂ atmospheres are:

Table 1. Observations of O₂ and O₃ in the solar system. For the martian atmosphere, equivalent mixing ratios are calculated with a mean total column density of 2.08 x 10^{23} cm⁻² corresponding to a mean surface pressure of 5.6 mbar. (*) The surface mixing ratio was measured by Viking at a 7.5 mbar pressure.

<table>
<thead>
<tr>
<th>species and references</th>
<th>column density cm⁻²</th>
<th>equivalent</th>
<th>mixing ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mars</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O₂</td>
<td>(x10^20)</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>Barlow (1972)</td>
<td>(2.8 ± 0.2)</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>Carleton et al. (1972)</td>
<td>(2.8 ± 0.3)</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>Trauger and Lunine (1983)</td>
<td>(2.3 ± 0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owen et al.</td>
<td></td>
<td>0.0013 (*)</td>
<td></td>
</tr>
<tr>
<td>(Viking - in situ) (1977)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O₃</td>
<td>(x10^-6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Espevak et al. (1991)</td>
<td>(4.0 ± 1.3)</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Clancy et al. (1999)</td>
<td>5.4 - 10.8</td>
<td>2.6 - 5.2</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O₂</td>
<td>Mills (1999)</td>
<td>&lt; 3 x 10^-5</td>
<td></td>
</tr>
<tr>
<td>Icy satellites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ganymede (G), Europa (E), Rhea (R), Dione (D)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atmospheric O₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hall et al. (1998)</td>
<td>(x10^14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O₂ trapped in ice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noll et al. (1996)</td>
<td>(x10^16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noll et al. (1997)</td>
<td>4.5 (G)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-6 (R &amp; D)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© European Space Agency • Provided by the NASA Astrophysics Data System
These processes are strongly coupled. First, the efficiency of O₂ production from H₂O photodissociation decreases with increasing CO₂ abundances, as CO₂ absorbs UV photons in the same wavelength range as H₂O. This mechanism for photochemical O₂ production is to some extent self-regulating, as the photons responsible for H₂O and CO₂ photodissociation can also dissociate O₂.

Second, in the presence of HO₂ radicals (H, OH, HO₂) produced by H₂O photolysis, catalytic cycles recombine O and CO into CO₂ (see "Oxygen loss" reactions above). This recycling into CO₂ requires very low abundances of HO₂ radicals since they are not consumed in the chemical schemes.

We made different runs with CO₂ partial pressure (P_CO₂) from 6 mbar (Mars) to 3.2 bar. Due to the chemical mechanism presented above, the O₂ build-up is limited for P_CO₂ < 1 bar with abundances of O₂ and O₃ that remains much too low to produce a detectable IR feature of O₃. With P_CO₂ > 1 bar, it is different: first, photons available for the photolysis of H₂O become rarer and, due to the radiating properties of CO₂, the middle atmosphere is extremely cold, preventing H₂O from reaching the altitudes where it could be photolysed. In this case, the production of O₂ and O₃ becomes much more efficient: F_O₂ can be as high as 10 mbar and, in the absence of hydrogenous radicals destroying O₃, an high altitude ozone layer even denser than Earth’s one can develop. However, the infrared ozone signature can only be detected with P_CO₂ lower than about 50 mbar. Indeed, high CO₂ pressure bands mask the 9.6 μm O₃ band and would not permit its detection (see Fig. 2).

2.3. Dry CO₂ atmospheres

In the humid CO₂-rich atmospheres studied above, HO₂ radicals produced by H₂O photolysis limit the production of O₂. What happens in a dry atmosphere? Atreya and Gu (1994) estimated that, in the case of Mars, CO₂ should be fully converted to CO and O₂ in less than 6000 years due to the higher efficiency of the O + O + M → O₂ + M compared to the CO + O + M → CO₂ + M. Nair et al. (1994) and Selvis et al. (2002) showed that it is not true. Indeed, as it was said before, when O₂ starts to build up in the atmosphere, it also absorbs the photons that previously led to its own formation by dissociating CO₂. O₂ production hence diminishes when its abundance increases. However, CO₂ photolysis in a dry atmosphere can lead to high levels of O₂ (≈ 3% in a 4 bar CO₂ atmosphere) even if the conversion of CO₂ into CO and O₂ is not complete. As the amount of O₂ increases in this case with P_CO₂, we investigated the case of a 1 bar atmosphere (mainly N₂) containing 50 mbars of CO₂, which is the limit above which O₃ signature is masked by CO₂ absorption. We found an O₂ abundance of about 0.4% and a signature of O₃ close to the detection limit by Darwin. Here, considering only the O₃ signature alone would lead a “false positive” case.

2.4. H₂O photolysis and H escape

We also investigated an eventual false positive detection due to the enrichment in O₂ that follows the photolysis of water vapor associated to hydrogen escape. In this case, the production of O₂ is maintained by:
- a constant delivery of water in the upper atmosphere via hydrated particles,
- a high rate of H₂O photolysis by strong UV influx,
- a high rate of H escape,
- a low surface sink of O₂.

It is possible to “tune” these parameters in such a way that a significant build-up of O₂ occurs. However, besides the fact that the required conditions are unrealistic or result in a non-observable case (see discussion in Selvis et al., 2002), O₃ is destroyed by the products of H₂O photodissociations through the following catalytic cycles:
OH + O → H + O₂
H + O₂ → OH + O₂
O + O₂ → 2O₂

OH + O₃ → HO₂ + O₂
HO₂ + O → OH + O₂
O + O₃ → 2O₂

Therefore, while O₂ may be produced (with an extreme upper limit of a 1% abundance), no detectable ozone layer can form this way.

2.5. The triple signature: O₃ + H₂O + CO₂

Relying on our simulations, it turns out that the simultaneous signature of H₂O, CO₂, and O₂ within Darwin’s spectral window cannot be due to abiotic photochemistry. Searching for the triple IR signature of O₂, CO₂, and H₂O with a Darwin-like instrument appears more robust than a direct but unique detection of O₂. Indeed, while O₂ can become abiotically a major atmospheric component (up to few percents) O₃ cannot be detected in such cases at the same time as H₂O and CO₂ due to the masking effect of CO₂, and/or to the catalytic cycles destroying O₃ and following H₂O photolysis. Thus, through the triple signature, Darwin effectively avoids false positive detections in all the cases. Moreover,
- as oxygenic photosynthesis extracts O₂ from H₂O and fix carbon from CO₂,
- as life appears to be, as far as we know it, indissociable from H₂O,
- as CO₂ is a constituent of all the known terrestrial planet atmospheres and an expected constituent of habitable extrasolar planets Kasting et al. (1993),
- as O₃ is a powerful, logarithmic, tracer of O₂ Léger et al. (1993),

searching for this triple signature in quest of photosynthetic sources of O₂ is not a restrictive strategy when compared to search of O₂, or O₃, alone.

3. NON DETECTABLE O₂-RICH ATMOSPHERES THROUGH THE O₃ SIGNATURE

If one considers it more important to avoid false positive detections than to miss inhabited planets, the triple signature O₂-CO₂-H₂O proves to be a robust biosignature by eliminating several potentially ambiguous cases. However, this signature is not a universal tracer of life: the inhabited early Earth did not exhibit such a signature until about 2 Gyrs ago. Moreover, as we are going to see, even an O₂-rich atmosphere sustained by an ecosystem does not imply a detectable O₃ feature in the thermal spectrum.

Table 2. Results of numerical simulations with PHOEBE. For different types of atmospheres (dry and humid) and for different levels of CO₂, this table indicates if the partial pressure of O₂ produced photochemically exceeds 1 mbar and if thermal spectrum exhibit a detectable signature of O₃ at 9.6 μm. None of the atmospheres modelled exhibit the triple O₂-CO₂-H₂O signature.

<table>
<thead>
<tr>
<th>atmosphere</th>
<th>P_{CO₂}</th>
<th>P_{O₂}&gt;1 mbar</th>
<th>O₃ band</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂ + H₂O</td>
<td>&lt;1 bar</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>&gt;1 bar</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>CO₂ without H₂O</td>
<td>&gt;50 mbar</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>&lt;50 mbar</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>≤50 mbar</td>
<td>yes</td>
<td>yes (weak)</td>
</tr>
<tr>
<td>H₂O + H escape</td>
<td>0</td>
<td>yes</td>
<td>yes (weak)</td>
</tr>
</tbody>
</table>

3.1. The formation of the 9.6 μm band of O₃

Ozone is produced by the unique atmospheric reaction:

O₂ + O + M → O₃ + M

where M is any molecule. In the atmosphere of the Earth, oxygen atoms involved in the production of ozone come almost exclusively from the photolysis of O₂ at UV wavelengths shorter than 242 nm. Ozone in turn is photodissociated by UV and visible radiation above the troposphere. The ozone abundance is also controlled by other trace gas: HO₂, ClO₂, and NO₂ respectively in the upper, middle and lower atmosphere. The absorption of near UV radiation by the ozone layer causes the stratospheric warming. Figure 3 presents the temperature profile of the atmosphere. The increase of temperature with altitude in the stratosphere is a direct consequence of the absorption by ozone of UV photons. The feature of the ozone band at 9.6 μm depends on both the O₃ profile and the thermal profile. As is shown by the contribution function plotted on figure 3, most of the IR photons of this band are emitted in the stratosphere. Indeed, O₃ is a strong absorber and the opacity is high in the lower atmosphere. Hence, the signal at 9.6 μm is seen as an absorption feature: its brightness temperature corresponds to an average temperature of the stratosphere (weighted by the contribution function) which is cooler than the surface. This leads to a very important fact: the detectability of the ozone signature in the thermal emission is determined by the temperature contrast between the surface and the stratospheric ozone layer. The satellites observing the infrared emission of the Earth can see an emission feature at 9.6 μm when they watch the Earth over the Antartic continent as in this case, the frozen surface is cooler than the stratosphere. Therefore, if the average temperature of the radiating ozone layer is close to the surface temperature, no signature would be visible with low resolution and signal-to-noise ratio. The atmospheric processes involved in the ozone signature are therefore coupled in a very complex way:
- the abundance of O₃ depends on the chemistry and, hence, on the temperature profile,
- the thermal structure depends on the O₃ profile,
Figure 3. Temperature profile of Earth’s atmosphere and contribution function of the 9.6 µm O₃ band as a function of the altitude. The dashed curve shows the proportion of infrared photons coming from a given altitude level when looking at the Earth from space. The integrated area is normalized to 1. Most of the infrared emission at 9.6 µm comes from stratospheric ozone. Hence, the brightness temperature in the ozone band corresponds to a mean temperature of the stratosphere (≈230 K). This produces an absorption feature in the global thermal spectrum which envelope corresponds to a black-body emission at surface temperature (≈285 K).

and on the incoming radiation, the IR signature of O₃ depends on both the temperature and the thermal profiles. These statements raise an important question: is the case of the Earth particularly favourable for the detection of ozone? Does any O₂-rich atmosphere show an ozone signature?

3.2. Variability of the O₃ signature with the spectral type of the star

The Sun is a G2 spectral type star. G stars represent about 12% of nearby stars (ESA, 1997). Hot stars (F, A) are rarer. Cooler stars, K and mostly M, dominate the distribution in the vicinity of the Sun which counts about 300 stars within 10 pc. As target stars observed by DARWIN will not be only solar type stars, it is necessary to investigate what could be a terrestrial atmosphere submitted to a non solar radiation spectrum.

To this end, we used our 1-D atmospheric model (PHOEBE Selsis et al., 2002) to simulate the photochemistry of an Earth around a F9 and a K2 star. The evolution of the chemical composition and thermal structure was computed from an initial state set as the current atmosphere of the Earth. The UV-Visible-NIR spectrum of the star is determined from both UV observations (Heck et al., 1984) and a synthetic spectrum (Hauschildt et al., 1999). The semi major axis of the orbit is not taken equal to 1 AU but is chosen in order for the planet to receive the same energetic flux as the present Earth and then to have the Earth effective temperature. The results obtained show that the amount of ozone increases with the UV/visible ratio which is high for the F9 star and low for the K2 star. The higher the UV flux is the more opaque the ozone shield is. This property of oxygen-rich atmospheres is very favourable to the development of life: photochemistry builds up an ozone shield “adapted” to the intensity of the lethal UV flux. In the K2 case, the low abundance of stratospheric ozone and the stellar flux depleted in energetic photons produce nearly no warming. Most of the ozone absorption takes place in the lower atmosphere where the heat capacity by volume unit is high and therefore has no effect on the temperature. On the contrary, in the F9 case, the middle atmosphere is warm, with a temperature maximum higher than the surface temperature.

Figure 4. Synthetic spectra of Earth-like atmospheres around K and F stars. The resolution power (λ/Δλ) is 25. The atmospheric thermal structure of the planet orbiting around a F star does not produce a visible O₃ signature despite a high ozone content. On the contrary, for the planet around a K star, the low abundance of O₃ is visible because of the high temperature contrast between the surface and the stratosphere.
Figure 5. Contamination of the 9.6 μm O₃ band by the CO₂ "hot bands". This is the spectrum of an Earth-like planet at 1.2 AU. Its surface temperature is the same as on Earth but with 50 mbar of CO₂ in the atmosphere. Although the O₃ signature is still present, it can hardly be distinguished from the CO₂ bands. For CO₂ pressures higher than 50 mbar, a Darwin-like instrument would not permit to establish the presence of the O₃ feature.

K2 case, a detectable signature results from the strong O₃ absorption, enhanced by the high temperature contrast. On the contrary, in the F9 case, despite an O₃ layer thicker than on Earth, the high stratospheric temperature produces an ozone emission which can hardly be distinguished from the surface continuum in the low sensitivity spectrum of a Darwin-like instrument. Too much ozone may erase its own signature. Despite the last result, the consequences of this study are very encouraging in the perspective of a DARWIN/TPF mission. Indeed, it has been showed that cool stars which are the most numerous are also more favourable to the detection of an oxygen-rich atmosphere than the rare hot stars. Moreover, unlike the hot stars, cool stars spend a long time on the main sequence and then would allow a long development of life. Of course, in this study, only one parameter, the spectral type of the star, has been modified compared to the Sun-Earth system. Other variables concerning planet (gravity, clouds, chemical effect of trace compounds) could have a significant influence on the detection of ozone and should be investigated.

3.3. Disappearance of the O₃ signature with high CO₂ pressures

We have seen that when the CO₂ pressure is high enough to photochemically produce an IR-absorbing ozone layer (P_{CO₂} > 50 mbar for a dry atmosphere, P_{CO₂} > 1 bar for a humid atmosphere), high CO₂ pressure bands affect the thermal spectrum. With Darwin resolution (R~20) and signal-to-noise ratio (SNR< 10), the two CO₂ features arising on both side of the 9.6 μm band hide the O₃ signature. By masking the abiotic ozone it sustains, CO₂ masks also the potential “false positive imposters” that photochemistry could produce in the mid-infrared. On the other hand, this masking effect at high CO₂ pressure can produce “false negatives”, as it may reject truly photosynthetic ecosystems that would be detectable under a CO₂-poor atmosphere: an Earth-like planet orbiting at 1.2 AU from a Sun-like star would require about 50 mbars of CO₂ to maintain surface liquid water. In atmospheres with P_{CO₂} above this value, CO₂ absorption hides the O₃ band and would make the detection of ozone impossible with a Darwin-like telescope (see Fig. 5). With future telescopes with much higher resolution and signal-to-noise ratio, the detection of O₃ is possible in theory, however, the presence in the spectrum of strong CO₂ bands would show the possibility of an abiotic origin of d and O₃. Therefore, for P_{CO₂} > 50 mbars, O₃ no longer traces life. This maximum level of CO₂ allowing a detection of ozone can be used to define a new edge within the habitable zone: the edge of the “search zone”. Beyond this narrow fraction of the HZ, shown on figure 6, life cannot be traced by the triple infrared signature of O₃.
CO₂ is not the unique atmospheric compound that could sustain habitable conditions. Indeed, methane is an even more efficient greenhouse gas. On Earth, biogenic methane may have been the main warming gas before the rise of O₂ that occurred between 2.3 and 1.9 Gyrs ago. Therefore, the definition of the habitable zone should rely also on the potential warming by methane and the search for CH₄ signature would be a complementary way to trace life in the outer part of the habitable zone where high levels of CO₂ would hide the O₃ feature. The reliability of methane as a biomarker will be discussed in the next section.
4. OTHER BIOMARKERS?

Between 5 and 20 μm, some direct or indirect biogenic compounds have bands which are potentially detectable. The table 3 gives the approximate detectable levels of CH₄, NO, NO₂, N₂O and NH₃ compared to their present abundance in Earth's atmosphere. Like molecular oxygen or ozone, these compounds are produced directly or indirectly by the terrestrial biosphere but can also have abiotic origins.

N₂O in Earth's atmosphere is produced by nitrifying and denitrifying bacteria. As no other efficient source of this compound is known, it may be a reliable biomarker. N₂O is a very efficient greenhouse gas with several strong absorption bands in the thermal spectrum. A level of N₂O 10 times the terrestrial one would be detectable by Darwin but the abundance of this gas through Earth's history is not known.

The photo-oxidation of N₂O in our atmosphere leads to nitrogens oxides like NO and NO₂. These compounds have also detectable bands within the thermal spectrum but only at levels much higher than the present terrestrial one. Moreover, NO and NO₂ can be produced abiotically by impacts or lightnings occurring in volcanic plumes. These species may have reached high levels in the prebiotic atmosphere of the Earth (Comoneras et al., 2002)

Methane is produced on Earth by biological methanogenesis and by the human activity. Released in the atmosphere, it is rapidly photo-oxidised into CO₂ and H₂O. The present level of methane in the atmosphere produces a feature in the IR spectrum which is just below a level required for a reliable detection by Darwin. Before the rise of oxygen on Earth that happened between 2.3 and 1.9 Gyr ago, CH₄ was probably much more abundant: indeed, the present biological production of CH₄ released in an anoxic atmosphere would induce an atmospheric abundance 100-1000 times higher leading to a strong IR feature around 7.5 μm. As methanogenesis is among the oldest metabolisms known, Earth's atmosphere was probably CH₄-rich during a large part of the first half of its history.

CH₄ is found in the atmosphere of Titan and also in the atmosphere of the giant gaseous planets. Physical conditions there are very different from those on inner terrestrial planets: reducing composition, low temperatures, low UV irradiation. On a terrestrial planet, a strong and permanent production of methane is required to compensate its photodestruction and to lead to detectable atmospheric concentrations. The only known abiotic process releasing CH₄ on Earth occurs in hydrothermal vents were the oxidation of iron by water produces reducing conditions (Holm & Andersson, 1998). The fraction of CH₄ released this way to this atmosphere is extremely low: less than 0.01% of the total flux. However, this abiotic source could have been higher in the past and could be higher on other extrasolar terrestrial planets. It is hence not sure that methane is a reliable biomarker as an abiotic production may exist. However, as abiotic photochemical production of O₂ in reducing atmosphere is inefficient, the concomitant detection of O₂ or O₃ with CH₄ or NH₃ would strongly suggest biological activity.

<table>
<thead>
<tr>
<th>species</th>
<th>bands (μm)</th>
<th>minimum abundance</th>
<th>Earth abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₄</td>
<td>7.5</td>
<td>10 ppm</td>
<td>2 ppm</td>
</tr>
<tr>
<td>NO</td>
<td>5.4</td>
<td>1 ppm</td>
<td>&lt; 1 ppb (lo)</td>
</tr>
<tr>
<td>NO₂</td>
<td>6.2</td>
<td>10-100 ppb</td>
<td>1 ppb (lo)</td>
</tr>
<tr>
<td>N₂O</td>
<td>17, 8.5, 7.8</td>
<td>1-10 ppm</td>
<td>0.3 ppm</td>
</tr>
<tr>
<td>NH₃</td>
<td>11-9.6</td>
<td>1-10 ppm</td>
<td>0.01 ppm</td>
</tr>
</tbody>
</table>

5. CONCLUSION

The presence of molecular oxygen as a major constituent of a planetary atmosphere or the existence of a dense ozone layer do not imply biological activity as both can result from abiotic photochemistry. However, in the mid-infrared (5-20 μm) and with a resolution power of less than 25 (parameters presently estimated for ESA Darwin space observatory) the signature of O₃ at 9.6 μm associated with those of H₂O and CO₂ cannot be produced by abiotic photochemistry. Indeed, when oxygen build up abiotically in a humid atmosphere, it implies wither a high partial pressure of CO₂ (>1 bar) masking the O₃ band or a high H₂O photolysis rate which produces HO₂ radicals preventing the formation of a detectable ozone layer. This triple signature proves to be a robust way to distinguish inhabited planets. Any other mean able to detect only O₂ or O₃ without giving information about the presence of water vapor and the amount of atmospheric CO₂ would incur a risk of false positive detection.

Though the detection of this triple signature implies an oxygen producing ecosystem, the contrary is not true. First, the ozone layer signature may be hidden by high CO₂ pressures (>50 mbar). A consequence of this CO₂ masking, relying on the definition of the habitable based on the greenhouse effect of CO₂ (Kasting et al., 1993), is that O₃ can only trace life in the inner circumstellar habitable zone. Second, the O₃ absorption or emission feature can be too weak to be detected. This happens when the temperature contrast between the O₃ radiative layer and the surface is low. This situation can be found in the habitable zone of stars hotter than the Sun, like F stars emitting a lot of UV and inducing the forma-
tion of a dense and hot O$_3$ layer. On the other hand, the good visibility of the 9.6 μm band in the habitable zone of cool K stars, much more numerous than G, and F types, is promising for Darwin. Finally, the fact that oxygenic photosynthesis appeared on Earth much earlier than the rise of oxygen is a proof that the existence of oxygen producers is not always detectable. Methane may be a complementary biomarker as high abundances of this compound, produced by biological methanogenesis, were likely on the early Earth before the rise of O$_2$. Such levels would produce a detectable feature at 7.5 μm that could trace anaerobic ecosystems comparable to the primitive biosphere of our planet. However, abiogenic production of methane is possible even if it may not lead to detectable levels on terrestrial planets. Abiotic synthesis of methane still have to be studied before giving to this molecule the status of biomarker. The purpose of the preliminary simulations presented in this paper is not to select definitively reliable or efficient biomarkers for the search for life on exoplanets but to understand the possible processes occurring in planetary atmospheres in order to analyse the future data provided by the next generation of space observatories. Detecting signs of life is only conceivable as the result of a detailed analysis of all the physical and chemical parameters deduced from observations.

REFERENCES

Barker E.S., 1972, Nature, 238, 447
Bracewell R.N., 1978, Nat, 274, 780
Carleton N.P., Traub W.A., 1972, Science, 177, 988
ESA, 1997, In: The Hipparcos and Tycho Catalogues (vol. 1-17), 0+
Kasting J.F., 1988, Icarus, 74, 472
Mayor M., Queloz D., 1995, Nat, 378, 355
Rouan D., Baglin A., Copet E., et al., 2000, Earth, Moon, Planets, 81, 79
Selsis F., 2000, Ph.D. thesis, Université de Bordeaux I
Trauger J.T., Lunine J.I., 1983, Icarus, 55, 272
A TEST FOR LIFE ON EXOPLANETS: THE TERRESTRIAL VEGETATION DETECTION IN THE EARTHSHINE SPECTRUM

L. Arnold(1), S. Gillet(2), O. Lardière(2), P. Riaud(3), J. Schneider(4)

(1)OHP CNRS 04870 Saint-Michel-l'Observatoire, France, Email: arnold@obs-hp.fr
(2)LISE CNRS 04870 Saint-Michel-l'Observatoire, France, Email: sgohp@obs-hp.fr, lardiere@obs-hp.fr
(3)LISE CNRS 04870 Saint-Michel-l'Observatoire, France, and LESIA CNRS Observatoire de Paris-Meudon, Place Jules Janssen, 92195 Meudon Cedex, France, Email: Pierre.Riaud@obspm.fr
(4)LUTH CNRS Observatoire de Paris-Meudon, Place Jules Janssen, 92195 Meudon Cedex, France, Email: Jean.Schneider@obspm.fr

ABSTRACT/RESUME

Spectroscopic observations of the Earthshine allowed us to make a relative measurement of the integrated Earth reflectance spectrum in which the terrestrial vegetation signature around λ=700nm has been detected. Therefore we conclude that the terrestrial vegetation, and thus terrestrial life, can be detected remotely when the Earth is seen as a single dot. We also conclude that vegetation can be detected on an extrasolar Earth-like planet, if a spectral resolution around 50 is available.

1. INTRODUCTION

When future space missions like ESA Darwin [1] or NASA TPF [2,3] will deliver their first low resolution spectrum of an Earth-like extrasolar planet, it is possible that we will look for spectral signatures able to unveil the possible presence of life on this planet.

Spectral biosignatures can be of two types. A first type consists of biological activity by-products, such as oxygen and ozone, in association with water vapour, methane and carbon dioxide [4,5,6]. These biogenic molecules present attractive narrow molecular bands. But oxygen is not a universal by-product of biological activity as demonstrated by the existence of anoxygenic photosynthetic bacteria [7].

A second type of biosignature is provided by signs of stellar light transformation into biochemical energy, such as the planet surface colour from vegetation [8]. This spectral signature is in principle a more universal biomarker than any biogenic gas such as oxygen, since it is a general feature of any photosynthetic activity. Unfortunately, it is often not as sharp as single, molecular bands: although it is rather sharp for terrestrial vegetation at ~700nm (9,10), see Fig.1), its wavelength structure can vary significantly among bacteria species and plants [7].

Fig. 1. Reflectance spectra of photosynthetic (green) vegetation, non-photosynthetic (dry) and a soil (from [9]). The so-called vegetation red edge (VRE) is the green vegetation reflectance strong variation from ~5% at 670nm to ~70% at 800nm.

Before initiating a search for extrasolar vegetation, it is useful to test if terrestrial vegetation can be detected remotely. This seems possible as long as Earth is observed with a significant spatial resolution [11], but is it still the case if Earth is observed as a single dot? A way to observe an integrated Earth is to observe the Earthshine with the Moon acting like a remote diffuse reflector illuminated by our planet. We present in Section 2.1 normalized Earth albedo spectra derived from Earthshine, showing several atmospheric signatures. We show in Section 2.2 how the vegetation signature around 700nm can be extracted.
2. OBSERVATIONS AND RESULTS
The Earthshine has been observed in 2001 with a low resolution spectrograph installed on the 80cm telescope at Observatoire de Haute-Provence.

2.1 Earth albedo $EA(\lambda)$
It can be shown [12] that the normalized Earth albedo is simply given by the ratio

$$EA(\lambda) = ES(\lambda) / MS(\lambda)$$  \hspace{1cm} (1)

where $ES(\lambda)$ and $MS(\lambda)$ are the measured Earthshine and sunlit Moon spectra, respectively.

![Fig. 2. Examples of measured Earth albedo spectra. Both spectra are normalized to 1 at 600nm, but the July spectrum is shifted upwards by 0.5 for clarity. The spectral resolution was ~50 in July, and ~240 in October. The July spectrum has been binned to 10nm/px to mimic the low resolution that might be used for the first extrasolar planet spectrum.](image)

The result is shown in Fig.2. The higher reflectivity in the blue shows that the Earth should be seen as a blue object from space. We met Dr. H. Schmitt, Apollo 17 astronaut, during the ESLAB 36 conference, who confirmed this and also told us that during the Apollo 17 flight to the Moon, the lunar crescent was very small and the Earthshine looked very bluish. This blue colour is due to the Rayleigh scattering in Earth's atmosphere and not only to the intrinsic blue colour of the ocean (discussed later in Fig.3). The $H_2O$ bands around 690 and 720nm, and $O_2$ narrower band at 760nm are clearly visible with a resolution of R-50. The slope variation occurring at ~600nm is partially the signature of the deepest zone of the broad ozone absorption band (Chappuis band).

2.2 Earth surface reflectance $SR(\lambda)$
Although the vegetation is partially responsible for the higher level of the spectrum above $\lambda=730nm$ in Fig.1, doing a quantitative measurement of the vegetation signal requires to remove the atmospheric absorption bands in this spectral region. Earth surface reflectance $SR(\lambda)$ can be written by the simple scalar definition

$$EA(\lambda) \sim SR(\lambda) \cdot AT^{\alpha_{E}}(\lambda)$$  \hspace{1cm} (2)

where $AT(\lambda)$ is the mean Earth atmospheric transmittance. The exponent $\alpha_{E}=2$ represents the typical airmass crossed by solar photons before going to the Moon.

![Fig. 3. An example of Rayleigh correction: The graph shows the 24-26 June spectrum $SR(\lambda)$ (above) after atmospheric absorption correction (Eq.2), but still containing the Rayleigh scattering signature. The spectrum is fitted with a Rayleigh law adjusted over the [500;670nm] window. The fit is then translated (dash) and adjusted to the [740;800nm] region of $SR(\lambda)$ to show the VRE (here VRE=7%). $SR(\lambda)$ is normalized to 0.3 at 550nm [13] to be compared to the ocean albedo [14]. The $SR(\lambda)$ higher slope in the blue is the signature of Rayleigh diffusion in Earth's atmosphere rather than simple ocean relectivity [12].](image)
The spectrum AT(λ) is measured and α is adjusted in order to remove all the atmospheric bands [12] as shown in Fig.3. SR(λ) does not represent the pure surface reflectance, but includes uncorrected atmospheric scattering. The Fig.3 shows SR(λ) fitted with the Rayleigh law A+Bλ^4 adjusted over the [500;670mm] window. The slope towards the blue does not hide the relatively sharp vegetation signature, which appears around 700nm. SR(λ) is then normalized to the Rayleigh fit (Fig.4).

**Fig. 4.** An example of data reduction sequence: The graph shows the June albedo spectrum EA(λ) (bottom). All atmospheric absorption features are then corrected according to Eq.2 and the spectrum is flattened with a Rayleigh law adjusted in the [500;670mm] window (Fig.3). The result is shown above with 1 and 10nm/px resolution. The measured red edge around 700nm is VRE=7% (Eq.3).

To quantify the vegetation signature, we define the Vegetation Red Edge (VRE) as

\[ VRE = \frac{(r_T-r_K)}{r_K} \]  

(3)

where \( r_K \) and \( r_T \) are the mean reflectances in the [600;670mm] and [740;800nm] windows in the spectrum after it has been flattened with a Rayleigh law as explained above. Flattened SR(λ) spectra are shown in Fig.5.

3. DISCUSSION AND CONCLUSION

![Image](image-url)

**Fig. 5.** Collection of SR(λ) spectra normalized to 1 at 600nm, but shifted upwards for clarity by 0.2, 0.4, 0.6 and 0.8, respectively. Note that only the [600;670] and [740;800nm] windows are used to estimate the VRE. The spectra are binned to 10nm/px.

We have found a VRE ranging from 4 to 10% with an accuracy estimated to be \( \sigma \approx 3\% \) [12]. These results are in agreement with estimations from models predicting 2 to 10% [12,15,16,17] and other measurements done in 2001 [18].

Although it seems that the Earth’s vegetation signature might be visible as a red edge at 700nm, it is difficult to measure in the Earthshine for two reasons. The first reason is related to its variable amplitude, induced by a variable cloud cover and Earth phase. The second reason is because it is hidden below strong atmospheric bands which need to be removed to access the surface reflectance including the vegetation signature.

For the Earth, our knowledge of different surface reflectivities (deserts, ocean, ice etc) help us to assign the VRE of the Earthshine spectrum to terrestrial vegetation. For an exoplanet, a VRE-like index might be as difficult to measure as for the Earth due to variable cloud cover of the planet. Even if an extrasolar planet would give a clear VRE-like spectral signal, its use as a biosignature would raise some questions because: i) for several organisms, such as *Rhodospseudomonas* [7], the red edge is not at 700nm, but at 1100nm; ii) some rocks, like schists, may have a similar spectral feature. For instance, spectra of Mars show a similar spectral feature at 3.5 μm, which were erroneously interpreted as vegetation due to their similarity with lichen spectra [19].

© European Space Agency • Provided by the NASA Astrophysics Data System
We nevertheless believe that, associated with the presence of water (and secondarily oxygen) and correlated with seasonal variations, a vegetation-like spectral feature would provide more insight than simply oxygen on the bio-processes possibly taking place on the planet. But since water, and thus clouds and rain, are essential for the growth of vegetation, extrasolar planets with a very low cloud cover and a corresponding high vegetation index are unlikely, more especially if the planet is seen pole-on, with a bright white polar cover. On the other hand, an extrasolar planet vegetation surface could be greater than on Earth (like during periods in the paleozoic and mesozoic eras on Earth for example).

One must also note that the measurement of an extrasolar planet VRE will not suffer from the intrinsic difficulty of the same measurement for the Earth through the Earthshine spectrum: The extrasolar planet albedo will simply be given by the ratio of spectra spacecraft/mother star. But a model of the exoplanet atmosphere is necessary to be able to remove the absorption bands that may partially hide the vegetation. Although the probability is weak, the planet may occult a background star, thus providing us a direct measurement of the planet atmosphere absorption.

The detection of a VRE index between 0 and 10% requires a photometric precision better than 3%. Exposure time to achieve this precision with Darwin/TPF on an Earth-like planet at 10pc with a spectral resolution of 25 is of the order of 100h based on recent simulations [20].

Finally, the Earth albedo spectral variations study is of interest for global Earth observation. It might provide data on climate change, as broad-band measurements recently showed [13]. We also think that the spectrum of Earthshine might be used for example to monitor the global ozone (with the Chappuis or Huggins bands).

4. REFERENCES

GROUND DEMONSTRATION OF HYPERTELESCOPES

S. Gillet, P. Riaud, O. Lardière, J. Dejonghe, A. Labeyrie, V. Borkowski, and L. Arnold

1L.I.S.E. Observatoire de Haute Provence 04870 Saint Michel l’Observatoire
sgohp@obs-hp.fr, lardiere@obs-hp.fr, labeyrie@obs-hp.fr, borkowski@obs-hp.fr

2L.E.S.I.A. Observatoire de Meudon Place Jules Janssen, F-92195 Meudon Cedex
pierre.riaud@obspm.fr

3Collège de France, 11 place Marcellin Berthelot, 75005 Paris
dejonghe@obs-hp.fr

4O.H.P. Observatoire de Haute Provence 04870 Saint Michel l’Observatoire
arnold@obs-hp.fr

ABSTRACT

We have built a 10 cm diameter interferometer having 78 apertures of 1 mm diameter and using a direct snapshot imaging mode: the pupil densification also called hypertelecope. We have tested the direct snapshot performance of this hypertelecope with laboratory simulated multiple stars and on the sky. Characterizations of our hypertelecope limited science results but this test offers interesting perspectives for the future on the ground and in space. Densification may have applications in various fields such as providing a similar limiting magnitude but with a significantly increased angular resolution compared to a monolithic telescope with a same collecting surface.

Key words: direct imaging interferometers; hypertelecope; high angular resolution; binary star.

1. INTRODUCTION

We present results obtained with a multi-aperture densified-pupil imaging interferometer or hypertelecope. Interferometry can provide various trade-offs between angular resolution and collecting area, but snapshot imaging may often improve observations and data collecting. In 1996, Labeyrie (1996) described an optical architecture allowing direct high-resolution imaging with high contrast. Pedretti et al. (2000) tested a first method of densification, using a diffractive approach. We describe here an improved optical scheme using a pair of micro-lens arrays for achieving the pupil densification. This miniature hypertelecope was tested on bright binary stars. We present here the results obtained on the sky.

Figure 1. A giant segmented lens L1 simulates the entrance optics. The incoming flat wavefront from an off-axis star focuses at the Fizeau focus (FF), forming the multi-peaked diffraction pattern sketched above. Once densified, the wavefront average shape remains flat and parallel but becomes stair-shaped since the slope of each wavefront segment is reduced by the pair of micro-lenses introducing identical propagation delays. The resulting wavefront is finally a stair-shaped wavefront with an unaffected average slope and focuses to a single narrow interference peak at the center of the broader diffraction pattern from each sub-pupil.

2. HYPERTELESCOPE PRINCIPLE

A classical multi-aperture interferometer with a periodic array of N small mirrors separated by the distance s can provide direct images. But if the spacing of the sub-apertures becomes much larger than their size, the image quality, in terms of its luminosity and the data analysis is considerably affected. The resulting image at the Fizeau focus (FF) has a low contrast central peak surrounded by a large number of secondary dispersed peaks (see Fig.1). The degradation is caused by the diffractive spreading of light from each small sub-aperture, generating a halo much broader than the central interference
peak and taking most light away from it. Several attempts to improve snapshot images by changing the entrance sub-apertures position have come up against the golden rule (Traub W.A. (1986)): any recombination of the entrance pupil destroys the image's properties. This latter rule was in fact much restrictive as demonstrated by Labeyrie (1996). Recombination of the entrance sub pupils is possible if the geometrical pattern formed by the center of each mirror is kept. If care is taken to configure the exit pupil so that sub-pupil centers are arranged like in the entrance aperture, a direct image is obtainable with full luminosity. The resulting densified image is a "windowed convolution" of the object: the normal convolution of the object with a spread-function, the "interference" function is followed by a multiplication by a "window" function which is the diffraction pattern from a single exit sub-pupil (Fig.2). This window function defines a small region of the sky, where the densified image of a star appears as a white central peak. Outside this region, the image of the star appears dispersed. Following Gillet et al (2001), we call ZOF (Zero Order Field) this narrow usable field and HOF (High Order Field) the peripheral sky field of size $\lambda/d$ where $d$ is the size of one sub-aperture:

$$ZOF(\text{sky}) = \frac{\lambda}{d \cdot \gamma_D}$$  \hspace{1cm} (1)

where $\gamma_D$ is the densification ratio defined by:

$$\gamma_D \equiv \left(\frac{D_i}{D_o}\right) = \frac{d_e}{d_i} \text{ if } D_i = D_o$$ \hspace{1cm} (2)

$D_i$ and $D_o$ the entrance and exit baselines, $d_i$ and $d_o$ are the entrance and exit mirror diameters.

When a star moves off axis, its corresponding dispersed peak begins to appear at the opposite edge of the sub-aperture's Airy disk. Thus, any point source outside the ZOF is still imaged through its dispersed peak. For a square periodic array, the number of resolved elements in the ZOF is $4N$, where $N$ is the number of sub-apertures.

3. EXPERIMENTAL SETUP AND RESULTS

Following numerical simulations and a first experimented hypertelecope using a diffraction based densification method (Pedretti et al. (2000)), we tried to improve the image and densification quality. We made a miniature hypertelecope with 10 cm baseline in order to have nearly diffraction-limited image quality without adaptive optics. A lens L1 (Fig.3) produces a pupil image, 10 times smaller than the entrance aperture, which is masked by a grid of 78 holes of 0.1 mm size, centered 1 mm apart. The virtual grid thus defined in the entrance aperture has 1 mm holes spaced 10 mm apart on the sky. Two arrays of convergent and confocal micro-lenses (ML1 and ML2) achieve the pupil densification. Collimated beams from each sub-pupil become recollimated and widened when transmitted through the facing pair of micro-lenses. The densification factor, ratio of ML2 and ML1's focal lengths, provides 80% filling (diameter) in the exit pupil ($\gamma_D = 6$). The micro-lens arrays utilized were fabricated with enough lens-to-lens uniformity of thickness to keep piston errors within Rayleigh's tolerance, as required for a highly constructive interference, providing a high Strehl ratio, in the star's "high-resolution" image.

The main hypertelecope characterizations are described on table 1. With its aperture size of 10 cm and equivalent mirrors of 1 mm diameter, the ZOF extent is $1.22(\lambda/d)/\gamma_D = 28''$ at $\lambda = 650$ nm and the angular resolution is $\lambda/B = 1.34''$.

![Figure 3. Hypertelescope experimental setup. The incoming light beam from a Newtonian telescope is collimated by lens L1. A Fizeau mask installed for convenience in the pupil plane following L1, rather than at the primary mirror, has $N = 78$ holes of 100 $\mu$m size each. It defines in the entrance aperture a virtual "dilated giant mirror" of 10 cm size with $s = 1$ mm sub-apertures.](image-url)
The experimental setup required some preliminary laboratory testing. The micro-lens arrays were aligned with a He-Ne laser. Unlike the first hypertelecope which used diffraction (Pedretti et al. (2000)), we experienced a critical point with the rotational alignment of the micro-lens arrays which requires a precision of less than 1°. The final image has then a central white peak surrounded by several secondary dispersed peaks if densification is not completely achieved. Once the alignment was completed, we have tested the imaging capabilities of the hypertelecope with laboratory simulated multiple stars. Figure 5 shows the laboratory images obtained with equivalent R, G and B filters. These preliminary results are in good agreement with theory in terms of flux concentration and field of view for the ZOF. In a second step, we assessed the hypertelecope’s capabilities on bright stars. On Fig 4 we show Castor (α Gem), of magnitude $m_ν = 1.98$. The magnitude difference with its brightest companion is $Δm = 0.9$. Exposures were taken with and without the Wratten filters WR(25), WV(58) and WB(80A). The four residual peaks surrounding the central one result from the incomplete densification utilized, 80%. The exposure times varied from 10 to 30 min. On Fig (5) the companion of α Gem appears clearly. We measured for JD=2452193.67 a separation $ρ = 3.8$ and an angle $α = 68.2°$. The literature gives (Heintz (1988)), for this date $ρ = 4.0$ arcsec and $α = 63.8°$. 

<table>
<thead>
<tr>
<th>Table 1. Hypertelecope characteristics.</th>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertelescope size B</td>
<td>10 cm disk</td>
<td></td>
</tr>
<tr>
<td>Number of sub-apertures</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>Mirror size (entrance/exit)</td>
<td>1 mm/100 μm</td>
<td></td>
</tr>
<tr>
<td>Spacing s (sky)</td>
<td>10 mm</td>
<td></td>
</tr>
<tr>
<td>Collecting surface</td>
<td>78.5 mm²</td>
<td></td>
</tr>
<tr>
<td>$ML_1$ focal length</td>
<td>20 mm</td>
<td></td>
</tr>
<tr>
<td>$ML_2$ focal length</td>
<td>120 mm</td>
<td></td>
</tr>
<tr>
<td>Densification ratio</td>
<td>6 (80 %)</td>
<td></td>
</tr>
<tr>
<td>Field of view (ZOF)</td>
<td>28 arcsec</td>
<td></td>
</tr>
<tr>
<td>Angular resolution</td>
<td>1.34 arcsec @ 0.65 μm</td>
<td></td>
</tr>
<tr>
<td>Spectral range</td>
<td>450 to 750 nm</td>
<td></td>
</tr>
<tr>
<td>Image sampling</td>
<td>0.62 pixel/arcsec</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Hypertelecope measurements on Castor</th>
<th>measured</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separation r (arcsec)</td>
<td>3.8 ± 0.3</td>
<td>4</td>
</tr>
<tr>
<td>Position angle a (°)</td>
<td>68.2 ± 5</td>
<td>63.81</td>
</tr>
<tr>
<td>$Δm$ (Castor A-Caster B)</td>
<td>0.8 ± 0.15</td>
<td>0.9</td>
</tr>
</tbody>
</table>

4. EXAMPLE OF A GROUND BASED HYPERTELESCOPE

As sketched in Fig.6, the Carlina approach is inspired form Arecibo’s radio-telescope. It uses the
natural curved shape of the ground as a stable substrate carrying elements of a spherical mirror, in the form of an exploded or sparse mosaic. Spherical mirror elements of 20 to 80 cm size, arrayed at 1 meter interval and carried by stable fixed tripods can be efficiently utilized according to the hypertelescope principle. This design requires an aberration corrector such as a Mertz corrector fixed on a 20 to 30 m high pylon. A 5-15 meters baseline is interesting for a first hypertelescope instrument. This Carina-like design (Fig. 7), beginning with 7 telescopes (the first ring and central obstruction), may be extended to 19 and 37 telescopes for a better (u,v) coverage. In a first step, the observations would be only available in a speckle mode if the sub-apertures diameters are smaller than 25 cm in visible or 60 cm in the K band. Addition of low order adaptive optics (only piston and tip-tilt corrections) would allow observation with a coronagraphic mode. The densification may be achieved with a pair of micro-lens arrays. This Carina precursor is under study at the Haute Provence Observatory.

5. CONCLUSION

The pupil densifier is an optimization of (u,v) coverage for all interferometers. The laboratory and sky results confirmed the snapshot imaging possibilities with hypertelescopes. This densification method using two micro-lens arrays, increases noticeably the image quality and is easily controllable. The field of view is limited to the ZOF, the non-aliased field of view of the interferometer but may be adapted to the required observation with a modular densifier. Densified pupil imaging can be more efficient in terms of magnitude limitation than a classical interferometer. Coronagraphic and spectroscopic implementations are possible. Numerical simulations show that the resulting densified images may be used for spectroscopic and coronagraphic implementations with a focal mask such as the FQPM and such methods may be appropriate for the research and detection of exoplanets. This test is only a miniature version of hypertelescopes but announces interesting observation methods: various densified pupil implementations are possible such as a ground based Carina design. Space version with less than forty small free-flyers may have powerful observing characteristics if coupled to a coronagraph. This last concept is being further explored in our group.

REFERENCES

Gilmozzi R. et al., 1998, SPIE, 3352,778
Heintz W.D., 1988, PASP, 100, 834
CORONOGRAPHY STUDY FOR HYPETELESCOPE

S. Gillet\textsuperscript{1}, P. Riaud\textsuperscript{2}, A. Boccaletti\textsuperscript{2}, L. Arnold\textsuperscript{4}, O. Lardi\`ere\textsuperscript{3}, V. Borkowski\textsuperscript{1}, and A. Labeyrie\textsuperscript{1,3}

\textsuperscript{1}L.I.S.E. Observatoire de Haute Provence 04870 Saint Michel l'Observatoire
sgohp@obs-hp.fr, labeyrie@obs-hp.fr, borkowski@obs-hp.fr
\textsuperscript{2}L.E.S.I.A. Observatoire de Meudon Place Jules Janssen, F-92195 Meudon Cedex
pierre.riaud@obspm.fr, anthony.boccaletti@obspm.fr
\textsuperscript{3}Collège de France, 11 place Marcelin Berthelot, 75005 Paris
dejonghe@obs-hp.fr, lardiere@obs-hp.fr
\textsuperscript{4}O.H.P. Observatoire de Haute Provence 04870 Saint Michel l'Observatoire
arnold@obs-hp.fr

ABSTRACT

Following the idea developed in (Boccaletti et al. 2000), a snapshot imaging interferometer is proposed as an alternative to the NASA Origin project: "Terrestrial Planet Finder". This concept is based on densified pupil imaging (Labeyrie 1996) and phase-mask coronagraphy by (Rouan et al. 2000). The so-called Four-Quadrant Phase-Mask features a large starlight attenuation ($10^{-9}$). Thorough calculations indicate that the detection of Earth-like planets amidst zodiacal and exozodiacal clouds is faster with an imaging system than with a nulling interferometer as originally proposed for the TPF design. Detailed numerical simulations have been carried out taking into account several sources of noise and as a result we found that Earth-like planets can be imaged up to about 25 pc with a large interferometer in the thermal infrared. Finally this concept seems promising for discovering life hints with medium resolution spectra in both the visible (O\textsubscript{2}, CH\textsubscript{4}, H\textsubscript{2}O and Chlorophylla signature) and the thermal infrared (CH\textsubscript{4}, H\textsubscript{2}O, O\textsubscript{3}, CO\textsubscript{2}, NH\textsubscript{3}).

Key words: hypescope, coronagaph, exoplanets, TPF, nulling interferometer.

1. INTRODUCTION

This article describes our recent progress in the study of the hypescope-coronograph concept for the detection and observation of Earth-like planets, the challenging goal of NASA's Terrestrial Planet Finder (TPF hereafter) space mission. Hypescope provides the possibility of observing direct high resolution images with an interferometer (Labeyrie 1996). Previous calculations (Boccaletti et al. 2000) have shown that this concept improves noticeably the detection capabilities of exo-planets, compared to the initial instrument concept, based upon the idea of a nulling interferometer according to Bracewell's option (Bracewell 1978; Bracewell 1979). We present here more realistic numerical simulations, incorporating several sources of noise that were not accounted for in our earlier simulations (Boccaletti et al. 2000), i.e. mainly the zodiacal and exozodiacal contaminations but also co-phasing errors among the sub-apertures and the thermal emission of mirrors. These simulations are performed with a new version of the phase-mask coronagraph (Rouan et al. 2000), the Four-Quadrant Phase Mask, which improves significantly the detection of faint circumstellar sources (Riaud et al. 2001) with respect to the former Roddier's phase-mask (Roddier et Roddier 1997). We present briefly the concept of hypescope in section 2. Section 3 presents the different sources of noise taken into account and we finally discuss the theoretical results of exo-planets detectability.

2. HYPETELESCOPE PRINCIPLE

A hypescope or also pupil densified interferometer is a imaging interferometer which performs snapshot images. The principle was previously described (Labeyrie 1996; Boccaletti et al. 2000; Riaud et al. 2002). Unless the use of noiseless detectors, densifying the pupil allows to take high resolution snapshot images. The dilated mirrors forming a so-called Fizeau interferometer, are optically put together and form a densified pupil, provided that the pattern formed by the center of the sub-pupils is kept with no changing. The relationship between the object $O$ on the sky and its image $I$ is given by:

$$I = [O \otimes Im] \cdot W \quad (1)$$
Figure 1. Attenuation of a planet by the coronagraph (solid) and the diffractive envelope ZOF (dotted) as a function of the angular separation ($\lambda/B$ unit). The maximal transmission is achieved at $1.4 \lambda/B$. However, the planet light lost from the zero-order image due to the envelope's attenuation finds its way into the first-order images appearing at known relative positions within the envelope.

$Int$ is the interference function, Fourier transform of the apertures. $W$, the windowing function, defines the hypertelescope’s field of view, also called the ZOF (Zero Order Field) and is the Fourier transform of an exit sub-pupil. In case of circular sub-apertures, the windowing function $W(\theta)$ is an Airy pattern and the ZOF in the high-resolution image corresponds to the first dark ring. Unlike the Fizeau interferometer and its infinite field of view, the pupil densification shrinks the field according to $W$ while intensifying the part of the interference function appearing within the diffractive envelope. However, a star located outside the ZOF but inside a region called the HOF (High Order Field) provides an interference peak inside the ZOF, but with a radial dispersion in polychromatic light. Thus, the HOF, larger than the ZOF, and defined by $HOF = \lambda/d$ where $d$ is the entrance mirror diameter, is taken into account for the detection calculation of exoplanets hereafter presented in section 3.

We consider here a hypertelescope with 37 hexagonal sub-apertures, placed on free-flyers, and located on a spherical surface according to a periodic hexagonal pattern, suitable for full densification in the exit pupil. A coronagraphic device is attached to the hypertelescope imager for attenuating the central star and detecting the possible planet’s image. We considered, among a wide range of coronagraphs, the Four-Quadrant Phase-Mask (FQ-PM hereafter) (Rouan et al. 2000; Riaud et al. 2001) which appears to be the most efficient coronagraph concept in terms of star light rejection. The FQ-PM is assumed to be perfectly achromatic and the hypertelescope has no central obscuration, which is achievable in practice by slightly off-setting the combiner optics at the focus of a large diluted mosaic mirror, whether paraboloidal or spherical.

3. NUMERICAL SIMULATIONS

Several sources of noise have to be taken into account when observing an infra-red signal from an extra-terrestrial planet orbiting around a bright star. The major source of noise is the residual star’s light and, to a lesser extent, the speckle noise originating from phase aberrations in the optical train (cophasing defects or mirror roughness). We also take into account the attenuation by the coronagraph on both the star and the planet and the transmission by the ZOF envelope (Fig.1). Each flux is calculated in a resel $\lambda/B$. Table 1 list the different sources of noise which are contaminating the exo-planet peak in the densified-pupil image at a $\theta$ angular distance from the star, with $J_p$ photo-events detected per second and per resolution element (resel) for the planet.

Including the different sources of noise, the $S/N$ can be written as:

$$S/N = \frac{J_p \cdot t}{\sqrt{(\sum J_l \cdot t + N_{dark} + N_{resel})}}$$  \hspace{2cm} (2)

Eq. 2 is also valid for a Fizeau interferometer since it has the same planet/background contrast. However, the hypertelescope provides the same $S/N$ in a
shorter time since its sensitivity is largely improved by the pupil densification which intensifies the interference peak in the general case where the detector have readout noise and dark current.

One of the questions concerns $S/N$ ratio of a planet located in the "habitable zone" i.e. with an effective temperature $T_P \approx 300$K. This "habitable zone" depends on the spectral type of the star which has to be taken into account. In order to have the largest sample of target stars, since the ZOF is very small ($\sim 12\lambda/B$ in diameter), it is important to exploit the higher-order dispersed peaks of a planet lying outside the ZOF to allow the detection of terrestrial planets for any type of stars and at any distances. Table 2 gives the flux remaining in the dispersed peaks when the bandwidth is reduced to accommodate the peak dispersion (in one resel) from order 0 (inside the ZOF) to order 7. As the planet moves away from the ZOF, the flux per resel of a higher-order peak decreases linearly owing to the spectral dispersion, while the planet's intrinsic luminosity varies as the inverse square of its distance to the observer. As a consequence, for a given spectral type, the $S/N$ of a terrestrial planet located in the HOF is larger than in the ZOF (as seen on Fig. 2 for the F0V type).

Figure 2 shows the $S/N$ estimated from Eq. 2 for stars belonging to 9 spectral types (F0V, G2V and M5V) located at any distance up to 25pc. The planet peak (whatever the order of the peak) is attenuated by the windowing function towards the edge of the ZOF and by the coronagraph towards the center of the field, as apparent in the $S/N$ profiles presented in Fig. 2. Therefore, given the star's distance and spectral type the detection is optimal for a specific order. In the case of Fig. 2, Earth-like planets are detectable in order 0 for M5V stars, in order 0 and 1 for G2V stars, and from order 1 to order 5 for F0V stars.

The theoretical $S/N$ obtained at $\lambda = 10\mu$m, for a sample of 389 main-sequence stars (M5, M0, K5, K0, G5, G2, G0, F5 and F0) contained in the Hipparcos catalog (ESA, 1997) (see Fig. 3), is shown on Fig. 4a and 4b with respectively a zodiacal flux of $10.8\,\text{mag}/\text{arcsec}^2$ (median value) and $13.0\,\text{mag}/\text{arcsec}^2$ (median value at the ecliptic pole). We assume a baseline of 80m, 10h of integration time and an exo-zodiacal cloud 10 times brighter than the zodi. As expected, the number of potentially detectable planets is considerably affected by the intensity of the zodiacal cloud. The $S/N$ can be larger than 10 for stars closer than 15 pc and as large as $\sim 300$ for a nearby for

<table>
<thead>
<tr>
<th>Noise</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_{\nu}(\theta)$</td>
<td>background from the residual starlight</td>
</tr>
<tr>
<td>$J_{\nu}(T_\nu)$</td>
<td>background from the zodiacal cloud with a temperature of $T_\nu$</td>
</tr>
<tr>
<td>$m_{\nu}(\theta)$</td>
<td>background from the accretion disk</td>
</tr>
<tr>
<td>$m_{\nu}(T_\nu)$</td>
<td>background from interstellar dust clouds with a temperature of $T_\nu$</td>
</tr>
<tr>
<td>$m_{\nu}(T_\nu)$</td>
<td>background from the millimeter continuum wave</td>
</tr>
<tr>
<td>$m_{\nu}(\theta)$</td>
<td>background from the mirror optical emissivity with a temperature of $T_\nu$</td>
</tr>
<tr>
<td>$J_{\rm conf}$</td>
<td>confusion noise from background stars in the HOF</td>
</tr>
<tr>
<td>$A_C$</td>
<td>attenuation by the coronograph</td>
</tr>
<tr>
<td>$A_H$</td>
<td>attenuation by the windowing function</td>
</tr>
<tr>
<td>$N_{\text{dark}}$</td>
<td>thermal noise from the focal plane array (FPA)</td>
</tr>
<tr>
<td>$N_{\text{read}}$</td>
<td>readout noise from the FPA</td>
</tr>
<tr>
<td>$QE$</td>
<td>quantum efficiency of 45% (Rockwell Si:As detector)</td>
</tr>
<tr>
<td>$T_O$</td>
<td>optical transmission of 46%</td>
</tr>
<tr>
<td>$T_F$</td>
<td>filter transmission of 45% (N band)</td>
</tr>
</tbody>
</table>

Table 1. Sources of noise budget

<table>
<thead>
<tr>
<th>order</th>
<th>Flux in %</th>
<th>Flux (spectrotrum)</th>
<th>$\lambda_{\min}$ in $\mu$m</th>
<th>$\lambda_{\max}$ in $\mu$m</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>86</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>98.5</td>
<td>84.3</td>
<td>7.65</td>
<td>12.75</td>
</tr>
<tr>
<td>2</td>
<td>85.5</td>
<td>74</td>
<td>8.5</td>
<td>11.9</td>
</tr>
<tr>
<td>3</td>
<td>66.4</td>
<td>59.4</td>
<td>8.925</td>
<td>11.475</td>
</tr>
<tr>
<td>4</td>
<td>53.2</td>
<td>38.1</td>
<td>9.18</td>
<td>11.22</td>
</tr>
<tr>
<td>5</td>
<td>44.3</td>
<td>30.8</td>
<td>9.35</td>
<td>11.06</td>
</tr>
<tr>
<td>6</td>
<td>38.1</td>
<td>26.6</td>
<td>9.47</td>
<td>10.93</td>
</tr>
<tr>
<td>7</td>
<td>32.9</td>
<td>26.6</td>
<td>9.56</td>
<td>10.08</td>
</tr>
</tbody>
</table>

Table 2. Flux of interference peaks at $\lambda = 10\mu$m and the corresponding wavelength range ($\lambda_{\max} - \lambda_{\min}$) integrated in one resel.
nearby G2V and K0V stars (α Centauri). In the favorable case (Fig.4b), terrestrial planets could be detected around 67% of stars in our sample. This first sample of stars is of course biased by the selection of only 9 spectral types but was useful to investigate the optimal baseline and the effect of the zodiacal light. Then to derive realistic performance, we also carried out the same calculation for any F, G, K and M main-sequence stars (667 targets) within 25pc with an optimal baseline of 80 m. In table 3, we show the full sample completeness in the N band (assuming each star has an Earth-like planet in orbit). Earth-like planets are potentially detectable around 73% of nearby stars in that case.

The interferometer baseline has also an important

![Figure 4](image)

**Figure 4.** a : Signal to noise ratio obtained for the sample of Fig. 3 with a baseline of 80 m and a zodiacal flux of 10.8mag/^2. Earth-like planets are potentially detectable around 12 % of the stars (47 stars). The zodiacal flux is 10 times brighter than the zodi. b : Signal to noise ratio obtained for the sample of Fig. 3 with a baseline of 80 m and a zodiacal flux of 13.0mag/^2. Earth-like planets are potentially detectable around 67 % of the stars (262 stars).

![Figure 5](image)

**Figure 5.** Coronographic images obtained with a 37 apertures hypertelescope. The simulation includes photon noise (T=10 hours, collecting area=10.6m^2, transmission=46%, star magnitude at 10µm=4.7), readout noise (5e^-/pixel/frame), differential piston and tip-tilt errors between sub-apertures (λ/170rms) and mirror roughness (λ/170rms). The first image (left) is the coronagraphic image before subtracting the quadrants and results from a combination of 3 narrow bands (8.4 ± 0.75µm, 10.2 ± 0.75µm and 12 ± 0.75µm). The Exo-Zodiacal light is dominant. The second image (right) shows Venus, the Earth and a secondary peak of Jupiter after subtracting opposite quadrants. The circle is the size of the Zero Order Field at 10.2µm.

<table>
<thead>
<tr>
<th>Spectral type</th>
<th>F</th>
<th>G</th>
<th>K</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of stars</td>
<td>62</td>
<td>217</td>
<td>193</td>
<td>195</td>
</tr>
<tr>
<td>detectivity</td>
<td>45%</td>
<td>98%</td>
<td>80%</td>
<td>46%</td>
</tr>
</tbody>
</table>

**Table 3.** Detectivity for a full sample of stars (667) with 37 telescopes and an 80m baseline. The threshold is fixed to a signal to noise ratio greater than 3 and the exposure time is 10 hours. The zodiacal flux is 13 mag/^2 and exo-zodiacal flux is equal to 10 Zodi.

impact on the detectivity. Figure 6 shows the completeness of the sample (fraction of detected planets) as a function of the baseline and for several intensi-
ties of the zodiacal cloud. As the flux of the zodiacal light decreases the hypertelescope becomes obviously more sensitive to fainter stars, thus improving the sample completeness. The dependence with the baseline is somewhat more complex and is correlated with the zodiacal flux. The number of positive detections reaches a maximum for an optimal baseline which depends on the zodiacal flux.

![Graph showing percentage of detected Earth-like planets as a function of baseline (from 40m to 100m) and for a zodiacal flux of 10.8, 12.0, 12.7, 13.0, and 14.74 mag/m² (Boulainger et Pérault 1988)).](image)

For a fainter zodiacal flux (Boulainger et Pérault 1988) \( (L_z = 14.74 \, \text{mag/m}^2) \), planets are becoming detectable around fainter stars, i.e. late-type stars or distant stars. However, late-type stars are most numerous and a larger baseline \( (90m) \) is then required to angularly separate the terrestrial planets. The maxima of the curves is then moving towards large baselines if the zodiacal flux decreases. For instance, with \( L_z = 13 \, \text{mag/m}^2 \), the maximum of detection is achieved for a baseline of \( \sim 80m \) and for G-type and K-type stars. With a much brighter zodi \( (L_z = 10.8 \, \text{mag/m}^2) \) the curve is almost flat because only planets around the brightest stars are detected, which means F and G stars for small baselines \( (\sim 40m) \) and K and M nearby stars for large baselines \( (\sim 100m) \).

Figure 5 presents the result of numerical simulations using a FQ-PM to attenuate the central star, obtained with 37 telescopes, and 80 meters of baseline. Zodiacal and exo-zodiacal backgrounds are included in the image together with co-phasing defects and mirror roughness. The central G2V star \( (m_v = 6.33, m_K = 4.70) \) provides a photon flux of \( 1.64 \times 10^6 \, \text{photons/s} \) in the N filter \( (\lambda = 10.2 \pm 2.6 \mu m) \) for a total transmission of 48%. Venus, the Earth, and Mars were added in the ZOF, Jupiter is added in the HOF. The luminosity of the zodiacal light is \( L_z = 7.7 \times 10^{-7} \, \text{W/(m}^2 \cdot \text{sr} \cdot \mu \text{m}) \) corresponding to \( 14.07 - 13 - 12.5 \, \text{mag/m}^2 \) for respectively 8.4-10.2-12.0 \mu m. As seen on Fig. 5, the FQ-PM provides a sufficient rejection rate, after subtracting opposite quadrant, for the snapshot detection of Venus, the Earth and the dispersed peak of Jupiter with 37 sub-apertures in a total integration time of 10 hours. Mars would require a longer integration time (12 hours).

4. CONCLUSION

The new numerical simulation performed here confirm the capabilities of the hypertelescope with coronagraph concept to achieve exo-planet detection in the mid-IR. We find here that the dominant sources of noise when searching Earth-like planets in the mid-IR are the exo-zodiacal and zodiacal clouds. The stellar residuals, with the speckles generated by phase errors are fainter if the starlight rejection is larger than \( 10^5 \).

The main drawback, the field of view limitation, can be overcome if exo-planets are detected using their secondary dispersed peaks. This makes it unnecessary to vary the array size for zooming the image, as suggested before. One can define an optimal fixed baseline according to the most frequent stellar spectral types to be observed. For instance, with a baseline of 80m an Earth-like planet is potentially detectable around 73% of target stars in less than 10 hours. A subsequent step which we now prepare is the testing of a hypertelescope with a phase-mask coronagraph. A study with a miniature version of it for laboratory testing is under study.

REFERENCES

Bracewell R.N. 1978, Nature 274, 780-781
The Hipparcos and Tycho Catalogues, 1997, ESA SP-1200
A WAVEFRONT ANALYSIS ALGORITHM FOR MULTI-APERTURE INTERFEROMETERS AND HYPERTEOSCOPES

Virginie Borkowski(1), Frantz Martinache(2), Deane Peterson(3), Antoine Labeyrie (4)

(1) Alcatel Space Industries, 100 boulevard du Midi, B.P.99, 06156 Cannes la Bocca Cedex, France, Collège de France & LISE, Observatoire de Haute Provence, 04870 Saint Michel, France, borkowski@obs-hp.fr
(2) Observatoire de Haute Provence, 04870 Saint Michel, France, martinache@obs-hp.fr
(3) State University of New York, Stony Brook, NY, USA, dpeterson@astro.sunysb.edu
(4) Collège de France & LISE, Observatoire de Haute Provence, 04870 Saint Michel, France, labeyrie@obs-hp.fr

ABSTRACT

We present here a new algorithm utilizing three-dimensional (3D) Fourier transforms for phasing interferometers and hypertelescopes. We develop a code in C which does 3D Fourier transforms of spectro-images, arranged as an (x, y, λ⁻¹) input cube [1] (λ⁻¹ is the inverse wavelength). Piston errors are determined by measuring the height of the dots in the output cube. The autocorrelation of the entrance pupil gives the positions of the columns containing these dots.

1. INTRODUCTION

Large multi-element optical arrays such as the diluted hypertelescopes studied in our group require efficient methods for measuring piston errors, as needed for adaptive correction. Unlike the case of monolithic telescopes, such diluted multi-aperture systems cannot use a Shack-Hartman analyser since there is no continuity of the wavefront across sub-pupils, even if there is a fully densified exit pupil like in the Exo-Earth Discoverer and the Exo-Earth Imager concepts. We therefore explored a possible extension of the "dispersed fringes" method already utilized visually by Michelson in the 1920's to adjust his 20-feet interferometer at Mt Wilson.

2. PRINCIPLE OF THE x, y, λ⁻¹ ALGORITHM

As an example, we may consider an entrance pupil with three apertures. The optical set-up produces a star image which is a 2D Fourier transform of the pupil. It resembles a honeycomb pattern in this case. We record it at several wavelengths to generate the input cube of the algorithm. The proportionality of fringe spacing to wavelength is corrected by enlarging the monochromatic images as λ⁻¹, like in the two-aperture fringe finder system of the GI2T. This makes the fringe stratifications parallel in the input cube thus distorted, and their slope is proportional to the piston error. With three apertures, the fringes in the distorted input cube thus appear as a 3D honeycomb, tilted in response to the piston errors (fig. 4). A 3D Fourier transform produces in the output cube a pair of symmetrical dots for each component fringe stratification. Their heights is proportional to the piston error δ. With 3 apertures, there are 6 dots arranged as a flat but tilted hexagon in the output cube. Changing the piston errors moves the dots along vertical columns.

In the output cube, the location of these active columns matches the autocorrelation peaks of the pupil (Fig. 1-6). Indeed, the pattern of columns at the output, i.e. a vertical projection of the output cube, is the 2D Fourier transform of a horizontal slice in the input cube, according to classical results [11]. Since this slice is itself the square modulus of the pupil's Fourier transform, the output projection is the autocorrelation of the pupil.

Fig. 1. Entrance pupil with three apertures

2. Autocorrelation of the entrance pupil

3. Central plane of the output cube. The coloured dots match the autocorrelation's ones.

4. Tilted honeycomb-like array of columns, obtained as the input cube for a triple aperture.

5. Output cube corresponding to fig. 5, with dots arranged as a flat hexagon, tilted like the input honeycomb.

For the numerical simulations, we first generate random piston errors $\delta(x, y)$ in the multiple aperture and calculate the input cube, i.e. the distorted stack of quasi monochromatic images, by 2D-Fourier transforming the entrance aperture at many different wavelengths $\lambda_n$. For simulating the distortion applied to the real data stack, needed as mentioned above for parallelizing the stratifications contributed by each pair of sub-apertures, we express the phase term in the 2D-Fourier code as $2\pi \{8x + 9y + 10 \delta(x, y)\}$ instead of the normal form $2\pi \lambda^{-1}\{a x + by\}$, where $x, y$ are the sub-pupil positions and $\alpha, \beta$ are the angular positions in the focal plane. The input cube thus simulated, in square modulus form, is then 3D-Fourier transformed into the output cube, also utilized in square modulus form.

3. THE PROBLEM OF REDUNDANCY

If the entrance pupil has redundant baselines (fig. 7), they provide several dots in some columns of the output cube, but at different heights corresponding to the different piston errors. One must do some work to figure out which dot corresponds to which pair of sub-apertures.

Fig. 7. Example of a redundant aperture with 4 elements. "o" indicates a sub-aperture.

Calculating the aperture's autocorrelation function tells how many dots are to be expected in each column (fig. 8).
A particular sub-aperture $d_0$ can serve as a reference. For the dots $d_{01}$, for example, the output cube provides directly the optical path difference between sub-apertures $d_0$ and $d_1$. Likewise we obtain the piston error on $d_0$. For $d_{21}$ ($d_{21} = d_1 - d_2$) we have the optical path difference between sub-apertures $d_1$ and $d_2$ thereby making the difference between the heights of $d_1$ and $d_{21}$ (known thanks to $d_{01}$) gives the piston error on $d_2$. Similarly, we recover all the piston errors (fig.11).

Thus, moderate levels of redundancy can still lead to unambiguous solutions for the piston errors with no additional effort. On the other hand, commonly used arrays with extreme levels of redundancy, such as the hexagonal arrays used by the Keck telescopes, make the direct solution for the piston errors impossible. One such example is shown in fig.9.

The redundancy problem can be solved by marking a subset of columns related to a reference aperture. The marking can be done in several ways: by displacing the reference aperture with respect to the periodic grid, or by modulating its intensity, its spectrum or its polarisation state, etc...

The use of an "out-rigger" reference aperture at a distance $a$ from the array which is larger than its size $b$ (fig.10) creates a subset of columns which are translated outside of the initial output pattern. This subset provides a direct map of all piston errors with respect to the reference aperture. The enlargement of the array, which can be a problem, is avoided by locating the reference aperture within it but at a non-redundant position. With a periodic array, having a square or hexagonal pitch, this is a matter of displacing the reference aperture with respect to the grid. In the output cube, each baseline connected to the reference aperture provides a dot in a column which is similarly displaced with respect to the main grid, thus solving unambiguously for piston errors. In the other columns, the several dots may then be discriminated and exploited to improve the signal/noise ratio.

Fig. 9. Hexagonal array (left) and its autocorrelation (right, shown at reduced scale).

Fig. 10. Use of an "out-rigger" sub-aperture as piston reference for a redundant array: one additional aperture located far from the array, at spacing larger than the array size (Fig.10: left) provides separated sub-sets in the autocorrelation (Fig.10: right shown with reduced scale). The outer subsets provide direct maps of piston errors.

In many case the addition of an "out-rigger" sub-aperture is undesirable. Alternately, a reference aperture can be marked by blinking with a shutter. This blinks the related dots in the output cube. The blinking dots again provide a direct map of piston errors.

Yet another way of marking columns uses a deliberate offset in the optical path of the reference aperture. If the offset exceeds the expected path differences among the others apertures, then the cloud of points splits, with an upper cloud containing the marked dots.

4. CONCLUSIONS

Our recent simulations show that the new algorithm can be utilized for phasing multi-aperture systems with very diluted pupils. With respect to methods previously considered [6] it may bring simplifications of the optics, especially if multi-spectral images are desired in the science channel.

We now have to compare the sensitivity of the method, and its signal/noise ratio in different detection regimes, with those attainable by other methods. We are also exploring possible solutions for reaching adequate processing speed, as needed in particular for ground-based observing with multi-element
interferometers and hypertelescopes. On Earth, the LIDAR method of atmospheric mapping proposed by Townes [10] may however prove superior in the infra-red, while also pushing the magnitude limit.

5. REFERENCES


6. ACKNOWLEDGMENT

The authors would like to thank Ettore Pedretti for his critical readership of this manuscript and for helpful discussions.
REVIEW ON HABITABILITY AND BIOMARKERS

L. Kaltenegger(1), M. Fridlund(1), J. Kasting(2)

(1)Scientific Project Department ESA/ESTEC, Postbus
299, 2200 AG Noordwijk The Netherlands, Email:
Lisa.Kaltenegger@esa.int
(2)Penn State University, 443 Deike Building
University Park PA 16802 USA, Email:
kasting@essc.psu.edu

1. ABSTRACT

The detection of O₂ or its product O₃ is our most
reliable biomarker so far. The existence of H₂O in
liquid state on the surface of a planet is considered
essential for the development of life; even so, it is not a
bioindicator. CO₂ indicates an atmosphere, and
abundant CH₄ can indicate biological sources, although
depending on the degree of oxidation of a planet’s crust
and upper mantle non-biological sources could also
produce large amounts of CH₄.

In the thermal part of the spectrum, the shape gives
a measure of the temperature of the object examined.
The mid-IR spectra can determine the planet’s albedo,
the temperature of the observable emitting regions and
thus the planet’s size.

Visible to near-IR spectra offer higher spatial
resolution for the same collecting area, are minimally
affected by temperature and therefore able to determine
the abundance of atmospheric species. However the
visible/near-IR continuum does not give direct
indication of the planet size because of the possible
albedo range.

2. DEVELOPMENT OF LIFE ON EARTH

The Earth formed about 4.5 billion years ago. The
release of gases from the interior formed the primitive
atmosphere, most likely dominated by carbon dioxide,
with nitrogen being the second most abundant gas and
trace amounts of methane, ammonia, sulfur dioxide
and hydrochloride acid, and oxygen [5]. Carbon
dioxide and/or methane played a crucial role for the
development of an early greenhouse effect that
counteracted the lower solar output.

Oxygen released into the atmosphere by photolysis of
H₂O would be removed by surface oxidation and
interaction with reactive gases. Two major processes
have changed the primitive atmosphere, the reduction
of CO₂ and the increase of O₂ [3]. A huge amount of
CO₂ must have been removed from the atmosphere,
most likely by the burial of carbon into carbonate
rocks, though the process is still debated. Primitive
cyanobacteria are believed to have produced oxygen.
After most of the reduced minerals were oxidized,
about two billion years ago, atmospheric oxygen could
accumulate.

3. INDICATIONS IN THE ATMOSPHERE FOR
HABITABILITY OF PLANETS

Biomarkers are features whose presence or abundance
requires a biological origin [15]. They are created
either during the acquisition of energy and/or the
chemical ingredients necessary for biosynthesis (e.g.
atmospheric oxygen and methane) or are products of
the biosynthesis (e.g. complex organic molecules and
cells).

In the IR it is essential to observe the 9.6-μm O₃ band,
the 15 μm CO₂ band, either the 6.3-μm H₂O band or its
rotational band that extends from 12 μm out into the
microwave region, and the 7.7-μm band of CH₄ as a
potential biomarker for early-Earth type planets. The
9.6-μm O₃ band is highly saturated and is thus a poor
quantitative indicator, but an excellent qualitative
indicator for the existence of even traces of O₂. Ozone
is a very nonlinear indicator of O₂, the ozone column
depth remains nearly constant as O₂ increases from
0.01 present atmosphere level (PAL) to 1 PAL.

Fig. 1. Earth spectra in the IR [1]

Observations from 8 μm to 12 μm of the H₂O
continuum allow estimations of the surface temperature
of Earth-like planets. However the atmosphere of
planets that are warmer than about 310 K will be
opaque in this region because of continuum absorption
by water vapour.

In IR and near IR, the emission of a planetary
atmosphere is very bright, but has to compete against
the planet’s blackbody radiation and the direct and
reflected starlight [3].
In the near-IR O$_2$ has a strong absorption feature at 0.76 µm, a broadband 0.45-µm to 0.75-µm O$_2$ absorption, and a strong H$_2$O band at 0.94 µm. The quantitative abundance of O$_2$ can be determined from the two features. If the planet is CO$_2$ rich, the weak 1.06-µm band could show its abundance. The best CH$_4$ band short of 1.0µm is at 0.88µm, but only shows absorption when the concentration is considerable greater than on modern Earth.

![Graph showing spectra of Sun and Earth](image)

Fig.2. Spectrum of a Black Body at 5777K, and 288K, closely resembling the Sun and Earth

Estimates of planet size and albedo, that will be crucial to determine the characteristics of a planet, can be determined from mid-IR observations, but not in the visible or near IR range [15]. We also need to explore the infrared spectrum in order to discover carbon dioxide at low spectral resolution [Owen 1980].

4. LIFE AND HABITABILITY

The search for life requires a general definition of life. Leger [12] defines a living being according to biochemistry and biophysics as a system that:

- contains information
- is able to replicate itself
- undergoes random changes in its information package that allow Darwinian evolution to proceed

This definition is currently used by the DARWIN team. It has to be refined further. If we look for life on this basis the closest life form would be a computer virus. This definition of life would give no starting point, how to search for life forms - as their structure might be unknown to us and the signals that show their present also.

"Our imagination in this domain is both too narrow and too wide. It is too narrow to make us sure that we do not miss interesting types of Livings. And it is too wide, so that we do not know in which direction we should start our search for other forms of life [19]."

To define search criteria, we presume that life has a carbon-based structure not too dissimilar from our own. Basing our assumptions on carbon organic compounds based life forms seems reasonable. Out of Earth’s $10^3$ known molecules, excluding DNA, there are only about 10% that do not contain carbon [16]. T. Owen [17] argues that carbon chemistry with water as a solvent is the most likely base of life. Ammonia is often suggested as a substitute for water and is nearly as good in many categories [17]. But it lacks the capability of protecting itself from destruction by ultraviolet light.

Carbon is a highly abundant element and exhibits a remarkable ability to form a host of highly complex molecules. The advantage of carbon over silicon is that carbon can move easily between a fully oxidized and a fully reduced state, whereas other elements like silicon tend to form large stable polymers that make it very difficult to recover the element. Thus rocks constitute a very significant sink of silicon.

Arguments in favour of water as a medium for life are that it is highly abundant, an excellent solvent, and remains liquid over a broad temperature range. The temperature range is high enough for chemical reactions to occur rapidly but also to allow large molecules to form. It has an extremely high dielectric constant, ε = 80, that allows salt ionization and the capability of building H-bonds with dissolved molecules. Specific conformation of macromolecules can form when in solution in water due to the solvent's attraction of hydrophilic groups (OH, CO, COOH,...) and repulsion of their hydrophobic ones (CH, CH$_2$,...). These conformations allow specific chemical reactions to build reproducible complex structures.

Water and ammonia behave quite differently from a photochemical standpoint. When H$_2$O is photodissociated to form H and OH, these byproducts usually react to reform H$_2$O. (This is not necessarily true for stratospheric water, which is why photodissociation at high altitudes can lead to hydrogen escape and water loss.) NH$_3$, by contrast, photodissociates to form NH$_2$ and H, and the NH$_2$ radicals react to form hydrazine, N$_2$H$_4$, and ultimately N$_2$. N$_2$ is stable against photolysis because of its strong triple bond. Thus, over time, an ammonia-rich atmosphere will be converted to N$_2$.

Basing the search for life on the carbon chemistry assumption it is then possible to establish criteria for habitable planets in terms of their size and distance from their stars. Gases in their atmosphere such as oxygen, methane and water vapour in combination would indicate the presence of life. Atmospheric features can provide clues of possible life forms for at least 2 billion years [17], more than $10^7$ times longer.
than radio signals reveal the presence of an advanced civilization on our planet.

5. HABITABLE ZONE

The Habitable Zone HZ around a star is defined as the zone around a star within which starlight is sufficiently intense to maintain liquid water at the surface of the planet, without initiating runaway greenhouse conditions that dissociate water and sustain the loss of hydrogen to space, see J. Kasting et al. [7]. The planet's effective temperature $T_p$ depends on the temperature $T_*$ and thus brightness of the star, the planet's albedo $A$ and its distance to the star $a_p$.

$$T_p = \left( \frac{1-A}{1+4A} \right)^{1/4} T_* \left( \frac{R_\star}{a_p} \right)^{1/2}$$ (1)

Observations in the mid-infrared can determine a colour temperature of the surface, the low clouds or high clouds from spectral shape [15]. Liquid water would be stable on Earth's surface for temperatures up to 647 K, the critical point for water, because the atmosphere acts like a pressure cooker. Using the estimates on size and temperature of the planet, we can calculate the total surface emission and the unknown albedo (2) and (3) [15].

$$\left( \frac{r_{pl}}{D} \right)^2 = \frac{L(\lambda)}{\pi} B(\lambda, T)$$ (2)

$$1-A = \left( \frac{D}{r_{pl}} \right)^2 \theta^2 \Delta \lambda \cdot L(\lambda)$$ (3)

$r_{pl}$ = planet radius (unknown)
$D$ = distance to planetary system
$L(\lambda)$ = star integral flux received
$\theta$ = angle of planet max elongation
$\Delta \lambda$ = observed colour temperature
$F(\lambda)$ = observed thermal planet flux at wavelength $\lambda$
$B(\lambda, T) = \text{Planck function at wavelength } \lambda \text{ for temperature } T.$

While the effective temperature of Venus and Earth are nearly identical, 220K and 255K respectively, the planets surface temperatures are 730K and $\sim$288K respectively, due to the different greenhouse gas abundance [4][11][18][20]. Observations should be able to discriminate between the different cases, but will be unable to penetrate clouds. Planets with small fractions of habitable surfaces or habitable surfaces that are hidden by deep, totally opaque material cannot be detected. For small and intermediate cloud coverage, we can determine the surface characteristics fairly well, while for extensive cloud cover we can only determine the characteristics of the cloud layer and above.

Providing that habitability requires liquid water on the planets' surfaces, for Earth-like planets with CO$_2$-H$_2$O-N$_2$ atmospheres, criteria of the HZ are defined by Kasting et al [9]. The inner edge of the HZ is determined by water loss via hydrogen escape and photolysis. The outer edge is determined by the formation of CO$_2$ clouds, which increase a planet's albedo, leading to runaway glaciation. In these limits climate stability is provided by a feedback mechanism in which atmospheric CO$_2$ concentrations vary inversely with planetary surface temperature. Other authors like Forget et al [5] have shown that CO$_2$ clouds may warm a planet's surface; thus, the HZ could be wider than found in the Kasting et al. model.

The TPF Science Working Group (TPF-SWG) [15] identified the waveband between 8.5 µm to 20 µm and preferably 7 µm to 25 µm, respectively, for the search for biomarkers in the mid-IR region and 0.7 µm to 1.0 µm and preferably 0.5 µm to 1.1µm for the visible to near IR region. The mid-IR spectra can give the planet's albedo, the temperature of the observable emitting regions and thus the planet's size.

![Fig.3. Normalized Infrared Thermal Spectral Models of the Earth [15]](image-url)
It is also essential to observe the 9.6-μm O$_3$ band, the 15μm CO$_2$ band, either the 6.3μm H$_2$O band or its rotational band that extends from 12μm out into the microwave region, and the 7.7μm band of CH$_4$ as a potential biomarker for early-Earth type planets. Observations from 8 μm to 12 μm of the H$_2$O continuum allow estimates of the surface temperature of Earth-like planets. However the atmosphere of planets that are warmer than about 310K will be opaque in this region because of continuum absorption by water vapour.

In the near-IR, O$_3$ has a strong absorption feature at 0.76 μm, a broadband 0.45-μm to 0.75-μm O$_3$ absorption, and a strong H$_2$O band at 0.94 μm. If the planet is CO$_2$-rich, the weak 1.06-μm band could show its abundance. The best CH$_4$ band short of 1.0 μm is at 0.88 μm, but only shows abundance when the concentration is considerable greater than on Earth.

![Fig.4. Normalized Visible Reflection Spectral Models of the Earth [15]](image)

We need to explore the infrared spectrum in order to discover carbon dioxide at low spectral resolution [17]. Freons would even be better indicators of life, particularly advanced life forms, as they have no known abiotic or biotic sources; the gases are all manufactured [8].

The advantage of operating in the thermal IR instead of the visible is the 10$^7$ times larger ratio of the planet/star emission. The much longer wavelength compared to the visible allows for greater tolerance. The broad absorption band would allow for a lower spectral resolution. Due to instrumental limitations like low spectral resolution for space missions - the gases that could be identified in the IR are H$_2$O, CO$_2$ and O$_3$.

The disadvantage of operating in the thermal IR is that cooling of the telescope is required, along with much larger apertures (or arrays of smaller apertures configured as an interferometer). By contrast, a single 8-10 m diameter telescope could be used for planet characterization in the visible.

![Fig.5. Calculated IR spectrum detectable by the DARWIN mission [2]](image)

CO$_2$ plays a key role in modulating Earth's climate, but it is also observed in the atmosphere of uninhabitable planets in our solar system, Mars and Venus, as shown in fig.6. The presence of gaseous H$_2$O in a planet's atmosphere and the knowledge of the planet's distance from its parent star and its radiating temperature should determine if the planet has liquid water on its surface. That could be measured by a space-based interferometer.

6. SURFACE BIOMARKERS

An interesting example for surface biomarkers is the red edge signature from photosynthetic plants at about 750 nm. The reflectivity changes by almost an order of magnitude, much larger than the change due to chlorophyll absorption. Photosynthetic plants have evidently developed that strong infrared reflection as cooling mechanism to prevent overheating and chlorophyll degradation [15]. Phytoplankton can cause temporal change in large areas of the ocean, but most of the reflected flux from Earth does not come from the ocean. The ocean is very dark in the optical and has strong water absorption bands in the IR.

Keeping in mind that the chances are very small that another planet has developed the exact same vegetation than Earth, finding the same signatures would be thrilling, but seems unlikely. Lovelock [13] concluded that the presence of large amounts of gases in the atmosphere out of
thermodynamic equilibrium is a criterion for the presence of biological activity. There is no definite physical basis for this criterion; however, on our planet the only processes that actually produce large quantities of gas out of equilibrium in the atmosphere are biological ones. All planetary atmospheres are out of thermodynamic equilibrium because their photochemistry is driven by UV photons from their parent star. Thus, Lovelock’s criterion is useful, but it must be applied with caution.

The simultaneous presence of some reduced gas such as CH₄ would even be a better indicator, as it is unstable around oxygen. Oxygen is a chemically reactive gas. Reduced gases and oxygen have to be produced concurrently to be detectable in the atmosphere, as they react rapidly with each other.

On Earth, the rate of biological O₂ production is much larger than the abiotic one. A planet that has a much smaller source of reduced volcanic gases would not remain habitable over a long period because of the missing mechanism for recycling CO₂. Could it have an O₂-rich atmosphere?

A planet, bigger than Mars, but not big enough to maintain active volcanism might be able to retain its O₂ [7]. If this abiotic planet had liquid water at its surface, it should have a vanishing small atmospheric O₂ concentration, as H₂ should consume O₂. If not, O₂ might be detectable.

A planet that has lost a lot of water could also accumulate a large amount of O₂. If H₂O is transported to the top of the atmosphere, where it is dissociated, hydrogen would be lost to space in a rapid rate while oxygen was retained. That effect might allow O₂ to surpass the rate of outgassing of reduced gases and possibly create a transient O₂-rich atmosphere[6]. This signature would eventually disappear on a volcanically active planet, but might generate a substantial ozone layer during the time it exists.

If oxygen sinks on a planet are inactive, abiotic O₂ would accumulate due to UV photolysis and hydrogen escape. The lack of liquid water slows down the rates of both physical and chemical weathering.

If a planet is small enough, somewhere between the size of Earth and Mars, it could retain a thick atmosphere but not its volcanism [7][9]. The sinks due to water weathering would also be missing in the case of a low surface temperature. Consequently a planet outside of the HZ could have an atmosphere rich in abiotic O₂. Such a planet would be small and its surface temperature would be below the freezing point of water; thus, it should be identifiable because by its location outside of the HZ and the fact that its water vapour bands would be weak.

A star hotter than the Sun emits more UV radiation and could therefore induce high photodissociation rates. On Earth-like planets with surface temperatures around 300K the first process is strongly reduced by the existence of cold traps at the tropopause.

7. OXYGEN AND DIAGNOSTIC OF LIFE?

It is thought that carbon in the primitive Earth atmosphere was mostly fully oxidized. Building of organic material required the reduction of CO₂ [16].

\[ \text{C}_2\text{O}_3 + \text{H}_2\text{O} + \text{energy} \rightarrow (\text{CH}_2\text{O}) + \text{O}_2 \]

This reaction replenished free oxygen, which is a highly reactive gas. If not continuously supplied, it would disappear in only about 4 million years. The presence of free O₂ and H₂O in a planet's atmosphere seems to indicate the presence of carbon-based life.

8. CONCLUSION: INDICATORS OF HABITABILITY

The presence of O₂ is an indication of life, but only for planets well within the HZ. Mars-like planets have small O₂-sinks and Venus-like planets large abiotic O₂ sources. The presence of a large amount of water vapour would imply habitability but not necessarily life. An independent temperature measurement will be needed to make sure that the planet is not in the middle of a runaway greenhouse stage, with water rapidly dissociating to produce the oxygen we observe. To constrain the alternative scenarios, the determination of
the albedo and size of the planet and its distance from the parent star is very important. A planet with O\textsubscript{2} and H\textsubscript{2}O absorption bands in its spectrum that lies within the HZ is the main target of a mission that searches for habitable planets like the DARWIN mission. There is the possibility that life exists on planets that do not show H\textsubscript{2}O and O\textsubscript{2}, if the production of O\textsubscript{2} by photosynthesis is not able to overcome the oxygen sinks. On Earth that describes the situation up to 2 Gyr ago [12]. Estimates of planet size and albedo, that will be crucial to determine the characteristics of a planet, can be determined from mid-IR observations. The visible/near-IR continuum does not give direct indication of the planet size because of the possible albedo range. We also need to explore the infrared spectrum in order to discover carbon dioxide at low spectral resolution [17].

Fig.7. Concept of the ESA DARWIN mission [1]

9. REFERENCES

1. DARWIN: http://sci.esa.int/home/darwin/index.cfm (07.05.2002)
14. Marcy, G.W. et al 1998, Giant planets orbiting faraway stars, Scientific American, Magnificent Cosmos, Volume 9, Number 1
15. David J. Des Marais, Martin Harwit, Kenneth Jucks, James Kasting, Jonathan Lunine, Douglas Lin, Sara Seager, Jean Schneider, Wesley Traub and Neville Woolf, Biosignatures and Planetary Properties to be Investigated by the TPF Mission, JPL 01-008, 2001

© European Space Agency • Provided by the NASA Astrophysics Data System
Scenario description of the construction of a Lunar South Pole Infrared Telescope (LSPiRT)

Paul van Susante

(1) Colorado School of Mines, 815A 14th street, Golden, CO 80401, USA, Email: paulvans@mines.edu

Abstract

In this paper a description will be given of the design of the telescope and the scenario that was chosen to construct this telescope. The Location and type of mission requires a design that is scenario and environment driven. The Scenario determines way of construction (on earth, in orbit, on moon), timeline and maximum masses, sizes and volume of the payloads to be delivered to the lunar surface.

The location on the Moon means that no infrastructure or resources are available without processing except for the lunar regolith itself. The environment determines the choices of materials and details (dust, ground situation) The local terrain determines communication, energy-supply, light, resources, location, foundation and the way of construction.

It appears that it is possible to built a telescope with the same capabilities as the Next Generation Space Telescope on the Lunar South Pole except for the sky coverage that will be limited by the location and orientation. The telescope has a diameter of 8 meters and is an altitude-azimuth design. The bearings will be made of super-conducting magnets that use flux-pinning to stabilize themselves while at the same time they are very energy-efficient. The foundation will be put together and dug in-situ using robots and telepresence in a virtual reality environment and using local laser rangefinders. If all goes well the telescope is expected to have settlements no larger than 0.03 mm during operation.

When the telescope is built, an infrastructure has been created for energy supply and relaying communications. The total mission is achieved by launching 3 Ariane 5 rockets in the 2006 configuration that can launch 20,000 kg in GTO.

Introduction

- On the Moon, unique situations exist for observing the infrared parts of the spectrum because of the lack of an atmosphere. In the Polar areas this situation is extremely good because of the Permanently shadowed areas which belong to the coldest places (50K) in our solar system. This means that the surrounding infrared background radiation which disturbs the measurements is very low. The South Pole offers the best location to build such an observatory. From research done a few years ago by several sources [1,2] with data of the Clementine mission it appears that a unique combination exists at the South Pole. Permanent Shadowed areas located within a few kilometers of a small area that is almost permanently lit by the sun, often referred to as "peak of eternal light" (PEL). By placing a communication relay on one of the Lunar Mountains it is also possible to have direct communications with this PEL which can not be seen directly from Earth. This combination results in the location choice for the placement and construction of the Lunar South Pole Infrared Telescope in Shackleton Crater together with a communication relay station in the form of a lander at Malapert Mountain and another communication relay and energy supply station in the form of a lander at the Peak of Eternal Light.

In Figure 1, an outline is given of the locations where equipment will have to be landed and placed according to the chosen scenario.
Fig. 1: Top View of the First part of the Scenario

Fig. 2: The Euromoond lander as it would land on Malapert Mountain as a communication relay station. Because the part of the Moon where the rest of the activity will take place is not in direct view of Earth it would mean a very difficult way of operating and it would require a great deal of autonomy of the landers which would make everything very complex. Because of this lander, a direct communications link can be maintained and the other parts of the scenario can be remotely controlled.

Fig. 3: An impression of how RoCaDi would come rolling down from the crater rim to deploy the power/data cable. Because the crater is in permanent shadow, the power for operating RoCaDi would come from lander 2 on the rim. Because it can not be guaranteed that the lander can be in a direct line of sight for communicating with RoCaDi, communication will also be done by way of the cable. It is possible to guide RoCaDi remotely that way to find a spot which is in direct line of sight and with a special camera it is possible to investigate the dark crater to find a suitable spot for construction of the Telescope. The level of light reflected of the top of the rim is comparable with Full Earth light.

Fig. 4: The Rolling Cable Distributor (RoCaDi) in a possible Configuration.
As Mentioned before, 3 launches are needed to transport all equipment. A description of the 3 launches will follow.

**LAUNCH 1**

(total 3,500 kg on lunar surface with Ariane ESC-B after 2006)
- This launch will consist of 2 landers.
- Landing at mountain Malapert with lander 1
To establish the communication relay to Earth for landing on the PEL, during construction and operation. It will be done by a precision landing on one of the highest points on the mountain, target size 100x100 m²

1.2 Landing at rim of crater Shackleton, peak of eternal light with lander 2

To establish the second leg of the communication system with Earth for the telescope construction and operation and also to generate power for RoCaDi, construction and operation of the telescope. This lander will also transport and deploy the Rolling Cable Distributor (RoCaDi). Again this requires a precision landing on the Peak of Eternal Light (PEL) at the Rim of Shackleton Crater with a size of the target area of 200x200 m². Once landed they will deploy the solar cells for power generation. Once online and operating RoCaDi will be deployed.

1.3 RoCaDi deployment
RoCaDi will lay the power/data cable from PEL to the bottom of Shackleton Crater. After deployment from lander 2 rolling remotely controlled down to the crater while unrolling the power/data cable. Down at the bottom of the crater it will search for a suitable area of 250x250 m² where the telescope can be constructed and where the landers 3 and 4 can touch down without problems and where the crater floor has not too much boulders to remove and craters to fill. Once the landing site is found, it will function as landing beacon in the eternally dark crater

**LAUNCH 2**

(total 3,500 kg on lunar surface with Ariane ESC-B after 2006)
- First landing in crater Shackleton with lander 3, near RoCaDi, near (<100 m) identified suitable telescope construction site
This lander will transport both construction robots and the first telescope parts, the foundation poles and the lower azimuth ring segments. It will land near RoCaDi using it as a beacon to guide the lander to a safe spot down in Shackleton Crater. Then both construction robots will be deployed and the power/data cable will be connected via RoCaDi to docking station for the construction robots. After power connection, the first action will be to place the laser range finders to be able to determine the relative position of all elements and the robots. Now the construction can begin with the following steps:
  - Start of preparing construction site, removing boulders
  - Dig holes for foundation poles in local grid on the exact right positions
  - Place foundation poles (robot 1 holds in exact vertical position while robot 2 fills hole till ground-level)
  - Apply pre-load (1.5 m regolith above ground-level)

**LAUNCH 3**

(total 3,500 kg on lunar surface with Ariane ESC-B after 2006)
- Second landing in crater Shackleton with lander 4, near RoCaDi, near (<50 m) previous lander but ~ 100 m from the telescope construction site.
This final lander will transport all other parts, the upper azimuth ring segments, the main support struts plus the entire upper part which consists of the altitude axis, the instruments, cooling and radiators, the mirrors and the mirror holding structure plus the command module and communication dish. It will land near RoCaDi using it and lander 3 as a beacon.

After landing, the robots remove the pre-loads from the foundation poles and construction can continue. Then the support arms are deployed in preliminary positions. The lower azimuth ring segments will be placed preliminary on the foundation poles and the arms. The lower azimuth ring segments will be lined out exactly and interconnected.
All preliminary connections will be fixed in place
- foundation pole - support struts
- foundation pole - lower azimuth ring
- support struts - lower azimuth ring
- lower azimuth ring segments itself
Connect power/data line to complete lower azimuth ring (make sure it can’t be damaged by further construction) Then assemble upper azimuth ring segments on top of lower azimuth ring. Place main struts on the upper azimuth ring. Align and fixate all segments of the upper azimuth ring and main struts. Connect power/data line to the complete upper azimuth ring. Install altitude axis in altitude bearings in the main struts. Connect the instruments, cooling section and radiators to the altitude axis. Connect the mirror support frame to the axis. Install the secondary support struts to the mirror support frame. Align and fixate all elements. Install the secondary mirror and then install the primary mirror. Set up the command module and communication dish.
Connect power/data line to the complete mirror section and command module and communication dish.

Now the telescope is ready for testing and operation. In the next part, an overview with images will be given of the construction phases.

Fig. 7: The lower azimuth-ring (part 2) is located on top of the foundation poles. It has also a diameter of 8 meters and is also a super-conducting magnetic bearing

Fig. 8: The foundation pole (one of six) when just placed in its correct position. The height of the pole is 3 meter and it is 1 meter below the surface. The radius of the foot is 0.4 m which results in a surface area of 0.5 m² per pole.

Fig. 9: The foundation pole when set in the ground and the installed pre-load up to 1.5 m above the original surface level.

Fig. 10: The foundation pole after removal of the pre-load and opening of the support arms in preliminary positions just before placement of the lower azimuth ring segments.

Fig. 11: The foundation poles in preliminary opened position
Fig. 12: The foundation poles in opened preliminary position with ground level indicated.

Fig. 13: Installation of the last lower azimuth ring segment before fixation.

Fig. 14: Part 1 and 2 complete, the foundation and the lower azimuth ring.

Fig. 15: Part 1, 2 and 3 complete, the foundation, the lower and upper azimuth ring are complete together with the struts which hold the outer altitude rings.

Fig. 16: The telescope when it is fully mounted and ready to enter the test and operational phase. Settlements during operations will not exceed 0.03 mm. During construction settlements will be about 13 mm.

Conclusion

It appears to be a feasible scenario that uses mostly proven technology on earth, though some parts should be researched in more detail and qualified for space. This scenario is explained in much more extensive detail in [3]. This study was done partly at ESTEC, the European space research and technology centre in Noordwijk, The Netherlands and at the Delft University of Technology. Contact me for more information.

References


THE NUMBER OF HABITABLE PLANETS IN THE MILKY WAY OVER
COSMOLOGICAL TIME SCALES

Werner von Bloh, Siegfried Franck, Christine Bouhama, Hans-Joachim Schellnhuber
Potsdam Institute for Climate Impact Research, P.O. Box 60 12 03, 14412 Potsdam, Germany,
Email: bloh@pik-potsdam.de

ABSTRACT

A general modelling scheme for assessing the suitability for life on any Earth-like extrasolar planet is presented. This approach is based on an integrated Earth system analysis in order to calculate the habitable zone in main-sequence-star planetary systems. A new attempt by Lineweaver [1] to estimate the formation rate of Earth-like planets over cosmological time scales is applied to calculate the average number of habitable planets in the Milky Way as a function of time. The combination of this results with our estimations of extrasolar habitable zones yields the average number of habitable planets over cosmological time scales. We find that there was a maximum number of habitable planets at the time of Earth’s origin.

Key words: habitable zone, Earth-like planets, panspermia.

1. INTRODUCTION

One prominent and still open question is whether life exist on other planets. Recent progress in astronomical measurement techniques has confirmed the existence of a multitude of extrasolar planets. Up to nearly 100 planets have been observed around solar-type stars, where most of them are giant planets. The ultimate quest of extrasolar planet research is to identify Earth-like planets located in the habitable zone of their host stars. The habitable zone (HZ) is defined as the region around a star within which a planet might enjoy moderate surface temperatures required for higher life forms. At present time, however, the detection of such Earth-like planets is still beyond technical feasibility. Therefore we are restricted to theoretical considerations in order to estimate the number of habitable Earth-like planets which could principally harbour life. The knowledge of two factors is crucial for this attempt. On the one hand it is necessary to know the planet formation rate (PFR) of Earth-like planets, on the other hand the probability that the planet formed is habitable has to be assessed. Recent results of Lineweaver [1] can be used to calculate PFR as a function of cosmological time. The extent of HZs for different types of main-sequence stars has been determined by different authors [2, 3, 4]. In the following we apply our definition of habitability [5, 6], that does not just depend on the parameters of the central star, but also on the properties of the planetary geodynamics. The calculation of the HZ for different central star masses allows us to determine the probability that an Earth-like planet is habitable. Finally it is possible to estimate the number of habitable planets in the Milky Way.

2. METHODOLOGY

To calculate the PFR it is necessary to estimate the star formation rate (SFR). Based on the most recent observational data, Lineweaver [1] fits the SFR for the universe to an exponentially increasing function for the first 2.6 Gyr after Big Bang followed by an exponential decline. He uses this fit to quantify star metallicity as an ingredient for the formation of Earth-like planets. The metallicity \( \mu \) is built up during cosmological evolution through stars, i.e.,

\[
\mu \sim \int_0^t \text{SFR}(t') \, dt'.
\]

Then the PFR can be parameterised in the following way:

\[
PFR = 0.05 \cdot \text{SFR} \cdot p_E(\mu) \cdot [1 - p_J(\mu)],
\]

where \( p_E \) is the probability that Earth-like planets are formed and \( p_J \) the probability for hot Jupiter formation with orbits at which they would destroy Earth-like planets. The prefactor 0.05 reflects the assumption that 5% of the stars are in the range of 0.8–1.2 solar masses \( (M_\odot) \). The relation between metallicity and the probability \( p_E(1 - p_J) \) is a so-called Goldilocks problem: if the metallicity is too low, there is not enough material to build Earth-like planets; if the metallicity is too high, there is a high probability of forming hot Jupiters. Taking all these effects into account, one can derive the time-dependent PFR. In Fig. 4a we show the PFR recalculted.
from Lineweaver [1] and rescaled to the present star formation rate in the Milky Way of about one solar mass per year.

The number of habitable planets in the Milky Way, \( P(t) \), can be calculated with the help of a convolution integral:

\[
P(t) = \int_0^t PFR(t') \times p_{\text{hab}}(t - t') dt'.
\]  (3)

The probability, \( p_{\text{hab}} \), that an Earth-like planet at time \( \Delta t \) after its formation is within the habitable zone (HZ) can be expressed as follows [7]:

\[
p_{\text{hab}}(\Delta t) = \frac{1}{C} \int_{0.8 M_*}^{1.2 M_*} M^{-2.5} \int_{R_{\text{inner}}(M, \Delta t)}^{R_{\text{outer}}(M, \Delta t)} R^{-1} dR dM,
\]  (4)

where \( C = 1.57 M_\odot^{-1.5} \) is a normalisation factor resulting from solving Eq. 4 between the central-star-mass-dependent minimum and maximum HZ boundaries 0.1 \( M/M_\odot \), 0.8 \( M/M_\odot \), and 4 \( M/M_\odot \), respectively. In order to estimate \( p_{\text{hab}} \), the following assumptions are made:

1. The stellar masses \( M \) are distributed according to a power law [8] \( \propto M^{-2.5} \);
2. the distribution of planets can be parameterised by \( p(R) \propto R^{-1} \), i.e. their distribution is uniform on a logarithmic scale in the distance \( R \) from the central star [7, 9];
3. following Lineweaver [1] we restrict our attention to the set of Sun-like stars in the mass range from 0.8 to 1.2 solar masses \( (M_\odot) \); and
4. \( R_{\text{inner}} \) and \( R_{\text{outer}} \) are the inner and outer boundaries of the HZ, respectively. They are explicit functions of the central star mass and the age of the corresponding planetary system [5].

In previous studies climatic constraints, e.g. the presence of liquid water at the planetary surface, have been used to assess the habitability of terrestrial planets around different types of stars [3]. Our method [6] defines additional constraints: first, habitability is linked to photosynthetic activity and second, habitability is strongly influenced by the "geodynamics" of the Earth-like planet. To estimate these constraints for the determination of star and inner and outer boundaries of the HZ we use our Earth system model [5]. It couples the increasing central star luminosity, the silicate-rock weathering rate, and the global energy balance to estimate the partial pressure of atmospheric carbon dioxide, the mean global surface temperature, and the biological productivity as functions of time \( \Delta t \) (see Fig. 1). The main point is the long-scale balance between the \( CO_2 \) sink in the atmosphere-ocean system and the metamorphic (plate-tectonic) sources. This is expressed with the help of dimension-less quantities:

\[
f_{\text{wr}} \cdot f_A = f_{\text{tr}},
\]  (5)

where \( f_{\text{wr}} \equiv F_{\text{wr}}/F_{\text{wr},0} \) is the weathering rate normalised by the present value, \( f_A \equiv A_c/A_{c,0} \) is the

continental are a normalised by the present value, and \( f_{\text{tr}} \equiv S/S_0 \) is the spreading rate normalised by the present value. Models with fixed continental area and fixed tectonic activity \( (f_A \equiv f_{\text{tr}} \equiv 1) \) are called geostatic models (GSM). We favour so-called geodynamic models (GDM) that take into account both the growth in continental area and the decline in the spreading rate. With the help of Eq. 5 we can calculate the normalised weathering rate from geodynamics via continental growth model and time function of spreading rate. For the investigation of an Earth-like planet under the external forcing of any main-sequence star we apply the linear growth model [10].

The relationship between the stellar luminosity, \( L \), and the radiation temperature, \( T_{\text{rad}} \), for the pertinent central star mass range is given by the Hertzsprung-Russell diagram. The connection between the stellar parameters and the planetary climate can be formulated by using a radiation balance equation [11], i.e.,

\[
\frac{L(t)}{4\pi R^2} [1 - \alpha(T, P_{\text{atm}}, T_{\text{rad}})] = 4 I_R(T, P_{\text{atm}}).
\]  (6)

Here, \( \alpha \) denotes the planetary albedo and \( I_R \) the outgoing infrared flux.

We define the HZ as the region around a central star within which an Earth-like planet has a non-vanishing biologial productivity \( \Pi \). \( \Pi \) is defined as a function of surface temperature, \( T \), and \( CO_2 \) atmospheric partial pressure, \( P_{\text{atm}} \):

\[
\frac{\Pi}{\Pi_{\text{max}}} = \left( 1 - \left( \frac{T - 50^\circ C}{50^\circ C} \right)^2 \right) \times \left( \frac{P_{\text{atm}} - P_{\text{min}}}{P_{1/2} + (P_{\text{atm}} - P_{\text{min}})} \right).
\]  (7)

\( \Pi_{\text{max}} \) is the maximum productivity and is assumed to be twice the present value of \( \Pi_0 \) [12]. \( P_{1/2} \) is the value at which pressure-dependent factor is equal to 1/2 and \( P_{\text{min}} = 10^{-5} \) bar (10 ppm) is the minimum value for \( C_4 \)-photosynthesis [13, 14]. There is
no photosynthesis-based life possible if $T$ is outside the temperature-tolerance window $[0^\circ \mathrm{C} \ldots 100^\circ \mathrm{C}]$ and $P_{\text{atm}}$ is lower than $10^{-5}$ bar. We assume a maximum value of $P_{\text{atm}} = 10$ bar. In this way the HZ is the habitable $R$-corridor in time, $\Delta t$:

$$
\begin{align*}
\text{HZ} & := \{ R \mid \Pi(P_{\text{atm}}(R, \Delta t), T(R, \Delta t)) > 0 \} \\
& = [R_{\text{inner}}(\Delta t), R_{\text{outer}}(\Delta t)]
\end{align*}
$$

(8)

3. RESULTS AND DISCUSSION

Figure 2. Width and position of the HZ (grey shaded) as a function of time for three different central-star masses ($M = 0.8, 1.0, 1.2 M_\odot$) for an Earth-like planet. $t_{\text{max}}$ is the maximum life span of the biosphere limited by geodynamic effects. $\tau_H$ indicates the hydrogen burning time on the main sequence limiting the life span of more massive stars.

We calculate the behaviour of this virtual Earth system at various distances $R$ from the central star. In Fig. 2 we have plotted the width and position of the HZ for the GDM for three different central star masses, $M = 0.8, 1.0, 1.2 M_\odot$ over time. First we can find that the width and the position of the HZ depend strongly on the mass of the central star. Furthermore, up to about 3.5 Gyr of co-genetic stellar and planetary evolution the outer boundary of the HZ is steadily increasing as a result of increasing central-star luminosity. After this point, the continental area has grown to such a size that weathering is very effective in bringing $\text{CO}_2$ out of the atmosphere and decreasing the outer boundary of the HZ which finally joins the inner one. For $1.2 M_\odot$ central stars life would be limited to 4.9 Gyr after starting co-genetic evolution because the central star leaves the main sequence and becomes a red giant. For 0.8 and $1.0 M_\odot$ central stars this limitation appears up to 6.5 Gyr after starting co-genetic evolution because continental growth and decline in spreading rate force atmospheric $\text{CO}_2$ content below $10^{-5}$ bar. In Fig. 3 the width $\Delta R = R_{\text{outer}} - R_{\text{inner}}$ of the HZ as a function of time and central star mass is plotted to emphasize the above mentioned qualitative characteristics.

Now we can come back to the calculation of the number of habitable planets within the Milky Way, $P(t)$ by evaluating Eqs. 3 and 4. The results for the calculation of $P(t)$ are presented in Fig. 4b. The value $P(t = 13.4 \text{ Gyr})$ of about $10^7$ is of the same order of magnitude as produced by recent calculations [15]. Evidently, the $P(t)$ has a distinct maximum at 8.5 Gyr. This just before the
time of Earth's origin ($t = 8.8$ Gyr). This supports the idea that interstellar panspermia (see, e.g., [16, 17]) might have caused a kick start to the processes by which life originated on Earth: there is palaeogeochemical evidence of a very early appearance of life on Earth leaving not more than approximately 0.4 Gyr for the evolution of life from the simple precursor molecules to the level of the prokaryotic photoautotrophic cells [18, 19].

ACKNOWLEDGEMENTS

This work has been supported by the German Federal Government and the Länder agreement (HSPN, grant number 24-04/235;2000).

REFERENCES

LABORATORY STUDIES ON COMPLEX ORGANIC MOLECULES ON MARS
PART 1 - RATIONALE

I.L. ten Kate (1), R. Ruiterkamp (1,2), O. Botta (2), B. Lehmann (3), C. Gomez Hernandez (3), N. Boudin (4), B.H. Foing (4),
P. Ehrenfreund (1,2)

(1) Leiden Observatory, P O Box 9513, 2300 RA Leiden, The Netherlands; email: tenkate@strw.leidenuniv.nl.
(2) Soft Matter/Astrobiology Group, Leiden Institute of Chemistry, P. O. Box 9502, 2300 RA Leiden, The Netherlands
email: ruiterka@strw.leidenuniv.nl, o.botta@chem.leidenuniv.nl, pascale@strw.leidenuniv.nl
(3) ESA Thermal and Life Support division, Keplerlaan 1, P. O. Box 299, 2200 AG Noordwijk The Netherlands
email: Bernd.Lehmann@esa.int, Cesar.Gomez.Hernandez@esa.int
(4) ESA Research Support Division, Keplerlaan 1, P. O. Box 299, 2200 AG Noordwijk The Netherlands
email: Nathalie.Boudin@esa.int, Bernard.Foing@esa.int

ABSTRACT

The search for organic molecules and traces of life is the future perspective of several missions to Mars. In order
to know what those mission should be looking for, laboratory experiments under simulated Mars conditions
are necessary. Especially since the Viking mission did not find any traces of organic compounds in the Martian
soil. In this paper the history of the search for life on Mars and the context of our laboratory studies [1], are
described. Furthermore it gives a short description of the experiments. This paper is the first part of a series
of three papers. The second paper will describe the experiments and methods [2], the third paper will be a
status report [3]. Both the second and the third paper can be found in the proceedings of the Second European
Workshop on Exo/Astrobiology, ESA Special Publication SP-518.

1. THE HISTORICAL PERSPECTIVE

One of the fundamental questions in Solar System exploration has always been if there is life beyond the
Earth and especially on Mars.

In 1892 Camille Flammarion published his ideas on life on Mars in a book called, "La Planète Mars et ses
Conditions d'Habitabilité". It contained a compilation of all credible telescope observations of Mars carried out
until then, and Flammarion's main conclusion was that Mars is obviously habitable. A few years before in 1877
Giovanni Schiaparelli had already published his apparent "discovery" of canals. Infected by Schiaparelli's "discoveries" Percival Lowell stated that the canals could be nothing else than the result of the work of very intelligent beings. He confirmed this with observations showing that the canal network was too regular to be natural, concluding they should have been created by a species more advanced than humans. Furthermore he concluded that the polar caps seen on the planet could be nothing else than water ice and that the dark spots seen along the canals were growth of vegetation. In 1961 some of Lowell's ideas appeared to be still alive when Vauclouleurs published his 'The Physics of the Planet Mars'. He wrote that Mars had a
85 mbar nitrogen atmosphere, is cold, but with a tolerable surface temperature, has seasonal changes probably
due to vegetation, and that the polar ice caps are not composed of frozen CO2 but of water-ice.

In 1960 a new era in the exploration of Mars begun. The Soviet Union was the first country to send a spacecraft,
Mars 1960A, to Mars. Unfortunately this mission and seven of its successors failed, making US Mariner 4, in
1965, the first mission to reach Mars. Mariner 4 showed that Mars had a cratered, rust-coloured surface, with signs of ancient presence of liquid water on some parts. From its data a surface atmospheric pressure of 4.1 to
7.0 mbar could be estimated, no magnetic field was detected. Mariner 6 and 7, launched in 1969, showed that Mars' south polar cap is predominantly composed of CO2, and that the atmospheric surface pressure lays between 6 and 7 mbar. In 1971 the Soviet Union sent two spacecraft to Mars, Mars 2 and 3. Although the landers failed, the orbiters sent back new data that enabled creation of surface relief, temperature and pressure maps, and gave information on the Martian gravity and magnetic fields. Mariner 9, launched too in 1971, made the first detailed images of the volcanoes, Vales Marineris, the polar caps and the moons Phobos and Deimos. It also revealed new data on global dust storms, the tri-axial figure of Mars, the rugged gravity field as well as evidence for surface Aeolian activity. In 1973 Soviet Mars 5 and 6 revealed more data on the surface and the atmosphere, followed by the Viking missions in 1975 (see next chapter).

The first completely successful mission after the Viking Missions the Mars Pathfinder Mission, was launched in
1997. Apart from the collection of many atmospheric data, such as early morning water ice clouds, the focus
was on surface science by using a small rover to drive around the surface [4]. After Pathfinder only two orbiter
reached Mars - Mars Global Surveyor (MGS), launched in 1996, and Mars Odyssey in 2001. MGS discovered
possibly water on Mars, by imaging relatively young
landforms and gullies. These results were endorsed by the results of the Mars Odyssey that found large quantities of hydrogen and water-ice just underneath the surface of Mars [5] [6] [7]. The results of these missions have put the search for possible life in a completely new perspective.

2. THE VIKING MISSION

The Viking mission consisted of two spacecraft, Viking 1 and Viking 2, each composed of an orbiter and a lander [8]. The main goals of the mission were to obtain high resolution images of the Martian surface, characterise the structure and composition of the atmosphere and surface, and search for evidence of life. The Viking landers carried several experiments, among them a biological and a molecular analysis experiment, whose purpose was the search for life-related organic molecules and organisms.

- Biological experiments and results

The biology experiment searched for the presence of Martian organisms by looking for metabolic products. The experiment was equipped with three instruments that incubated samples of the Martian surface under varying environmental conditions, the gas exchange (GEx) experiment, the pyrolytic release (PR) or carbon assimilation experiment and the labelled release (LR) experiment [9] [10].

The GEx experiment measured in two modes the production and/or uptake of CO₂, N₂, CH₄, H₂, and O₂ [11]. In the "humid" mode, a nutrient medium, composed of a complex mixture of organic compounds and inorganic salts, was added without soil contact, and the soil was only exposed to the water vapour in the atmosphere. The results showed that some CO₂ and N₂ was desorbed from the soil and that the oxygen accumulated rapidly after humidification in the headspace above the sample. This rapid accumulation of oxygen, in combination with the facts that (a) adding water in a later stadium had not caused further release of oxygen, and (b) oxygen was released from a sterilised sample (145° C for 3.5 hours), clearly excludes a biological explanation of the results. The results from the "wet" mode confirmed this interpretation [7], because (a) the absorption of CO₂ also occurred in sterile samples, and (b) because the CO₂ production rate slowed down, when the used nutrient was replaced with fresh nutrient.

The PR experiment was designed to detect the photosynthetic or chemical fixation of ¹⁴C to CO₂ or ¹³C to both [12]. The results showed that heating the samples to 175°C strongly reduced the reaction of ¹⁴CO₂ and ¹⁴CO₂ in the sample, although heating to 90°C did not have any effect. The data also suggested that reaction proceeded better in light, but storage of the soil within the spacecraft (in the dark) for four months did not affect the reaction.

The LR experiment used radio-respirometry to detect metabolic processes [13] [14]. The LR showed the rapid release of the labelled gas when a aqueous solution of dilute radioactive organic compounds was added. After the initial reactions the evolution of labelled gas slowly continued and terminated when 90% of the nutrients were still left. When the samples were heated to low temperatures (40-50°C) the reaction slowed down, raising the temperature to 160°C caused the reaction to end. The reactions in the soil had stopped as well when the samples were stored at spacecraft temperature for four months.

- Molecular analysis experiments and results

The main purpose of the molecular analysis experiments of the Viking landers was to investigate whether or not there are organic compounds present at a significant concentration at the surface of Mars [6]. A gas chromatograph mass spectrometer (GCMS) was used to analyse the soil. In the four samples taken from surface and subsurface material from both landing sites, no organic compounds of Martian origin were present at levels in the parts per billion (ppb) and parts per million (ppm) range [6]. Furthermore no traces of organic substances expected from meteoritic delivery were found [15].

- Conclusions

The results of the molecular analysis experiments unanimously point towards the absence of organic compounds in the Martian soil. Several explanations have been proposed, which all point towards the suggestion that the production and inflow rate of organic material is much smaller than the destruction rate. The biology experiments initially could have pointed to life in the Martian soil, especially the data of the LR experiments. However, in combination, and considering the non-detection of any organic compound in the upper soil made people search for a non-biological explanation. Recent experiments indicate that the pyrolysis products generated from several million bacterial cells per gram of Martian soil would not have been detected at the ppb level, by the molecular analysis experiment [16].

3. POSSIBLE DESTRUCTION MECHANISMS

Since the science results of the Viking mission several scenarios have been proposed to explain the absence of organic material in the Martian regolith. A combination of (a) short wavelength UV, which destroys organic compounds, and (b) oxygen, H₂O₂, metaloxides, or other oxidising agents is able to remove organic
compounds from the surface [6]. Additional support for this theory comes from recent work that showed that \( \text{H}_2\text{O}_2 \) is a good candidate for the thermally labile oxidant that produced rapid evolution of \( ^{14}\text{CO}_2 \), during the Viking LR experiments [17]. Furthermore, new evidence for the reactivity of the Martian soil due to superoxide ions, such as \( \text{O}_2^- \), has been published [18].

Overall the hypothesis involving oxidising agents in the regolith is still the most widely accepted one, although this subject requires future investigations.

4. THE MARS SIMULATION CHAMBER

It is still unclear why no traces of organic compounds, which should have been accumulated from meteoritic delivery [19], have been found by Viking in the Martian surface. Likely organic compounds are destroyed on the exposed surface, but may survive when protected in greater depth of Martian dust or inside rocks. In order to determine the stability of specific organic compounds, laboratory simulations are essential to understand chemical pathways on the Martian surface.

In this context an experimental programme is developed at the European Space Research and Technology Centre of ESA (ESTEC) and at Leiden University. The experimental research will include the investigation of organic molecules subjected to simulated Martian atmospheres and soil analogues. An atmospheric simulation chamber in combination with a solar simulator will be used to collect data on the combined effects of UV photo processing, atmospheric conditions and the presence/absence of oxidizing agents on organic molecules. All those described effects will be studied independently and in combination in order to get insight in the individual processes and their interactions on organics in the Martian soil. The organic compounds represent analogues for abundant meteoritic and cometary molecules and entail aliphatic and aromatic hydrocarbons, fullerenes, amino acids and nucleo-bases, carbonaceous solids and terrestrial analogues (i.e. kerogens).

The Mars Simulation Chamber (MSC) (Fig. 1 and 2), used for the experiments, is a 80 cm long, 60 cm high vacuum chamber, which will be filled with a simulated Mars atmosphere. This atmosphere can be adjusted to present and hypothetical past Martian conditions. To allow sample handling without external contamination two differentially pumped rubber gloves, similar to a glove box, were mounted on the side of the chamber. The solar simulator is mounted on top of the chamber to provide past and current solar spectra on Mars.

**Figure 1. Mars Simulation Chamber**

**Figure 2. Schematic outline of simulation chamber**

Six, 20 cm long, 1,5 cm diameter, stainless steel, sample holders (Fig. 3 and 4) will be filled with a mixture of Martian soil analogues and organic samples and sealed with O-ring stoppers. The sample holders will be mounted onto a copper ground-plate that allows temperature control by means of slush solutions that are pumped through the plate and stored in the MSC.

**Figure 3. Sample holder**

Three sample sites allow in-situ measurements of the volatiles at three different heights within the containers. The temperature of the medium in the chamber can be cooled by means of a cold shroud positioned around the...
sample containers. The temperature of the sample holders can be set within the same range. Feed-through lines allow the cooling medium to enter the cold-shroud and the sample holders.

The experimental protocol comprises three parts: preparation of the chamber, setting of the experimental parameters, and \textit{in situ} and post-processing measurements of the samples.

**SUMMARY**

Experiments will be performed using a vacuum chamber. In this chamber the Martian atmosphere, atmospheric pressure and temperature can be simulated, as well as the solar spectrum. These parameters can be adapted to the present and the past situation on Mars.

The following effects will be studied by the experiments:

- The effects of the Martian atmosphere.
- The effect of UV irradiation on organic molecules embedded in the soil.
- The effect of oxidation on organic molecules in the soil.
- The effect of thermal cycling on the surface.

This research is developed in the frame of a Mars Express Recognised Cooperating Laboratories RCL, in the category Exobiology and is supported by the BioScience Initiative of Leiden University and the Vernieuwingsimpuls.

**REFERENCES**


Special Session
Young Planetary Explorers (YPE)
LUNAR EXPLORATION AND THE SOCIAL DIMENSION

Virgiliu Pop(1)

(1)Ph.D Candidate, Law School, University of Glasgow, Glasgow G12 8QQ, Scotland
lawyer@space.com

ABSTRACT/RESUME

The scientific results of the lunar exploration missions have been the subject of many articles and books – yet their spiritual implications need not be overlooked. Holy texts of different religions were carried to the Moon, astronauts delivered prayers from the lunar orbit, and some of them even found new spiritual vocations on the Moon. People of all faiths were united in prayer when the astronauts faced danger, and the lunar missions received the blessings of many religious leaders – although there were few voices that saw the lunar landings as spiritually wrong.

1. INTRODUCTION

Viewed as a pinnacle of the technological evolution of humankind, the Moon landings are not void of social implications. From the protests of some groups in the 60’s that opposed what they saw as an unjustified spending of public money, to Buzz Aldrin’s holy communion on the Moon, the societal aspects of Man’s greatest achievement deserve a thorough investigation. In what follows, I shall briefly deal with a particular facet of the social dimension of lunar exploration: religion and spirituality.

2. A SPIRITUAL JOURNEY

The technological ascent into the physical Universe was doubled by a spiritual ascent to the supernatural heavens, and the lunar landings equate in many aspects with a pilgrimage to a holy land. The Apollo programme offered the astronauts the opportunity to replicate the experience described by John Gillespie Magee, Jr. in his poem “High Flight” – “Oh! I have slipped the surly bonds of Earth .../ And, while with silent lifting mind I’ve trod / The high unsupassed sanctity of space / Put out my hand, and touched the face of God.” [1]

The Soviet cosmonauts may have failed to discover God in space – as they failed to discover God on Earth. The American astronauts, adversely, sent prayers from outer space.

On Christmas Eve 1968, the Apollo 8 crew - Frank Borman, Bill Anders and Jim Lovell – broadcasted from the lunar orbit the first ten verses of the Book of Genesis in the Bible. This was a religious message whose contents was meaningful to people of other faiths, not only Christians [2]. It is claimed that a quarter of world’s population - nearly 1 billion people in 64 countries - heard the Apollo 8 reading and greeting, either on radio or on TV; and delayed broadcasts that same day reached 30 additional countries. [3].

The Christmas message, appreciated by people of many faiths, was however not appreciated by the American Atheists, that brought a legal suit to halt prayers in space. This did not stop Buzz Aldrin to hold a private communion service on the Moon once the Eagle has landed there. He said later, - “I would like to have observed just how the wine poured in that environment, but it wasn’t pertinent at that particular time. ... I was thinking more about our particular task, and the challenge, and the opportunity that had been given me. It was my hope that people would keep this whole event in their minds and see, beyond minor details and technical achievements, a deeper meaning behind it all -- a challenge, a quest, the human need to do these things” [4].

The lunar journey had a tremendous spiritual impact on Apollo 15’s James Irwin. One of his books is called “To Rule the Night: The Discovery Voyage of Astronaut Jim Irwin”. He often describes the lunar mission as a revelation: “I felt the power of God as I’d never felt it before.” [5]. Later on, he became an evangelical minister, forming the evangelical High Flight Foundation that he started in order to share with others his faith and to testify to the love of God that he discovered on the Moon. In his own words - "God had a plan for me, to leave the Earth and to share the adventure with others, so that they can be lifted up." [6].

A conversion often mentioned that however never happened was Neil Armstrong’s. It has become an urban legend (almost as famous as the Mr. Gorsky one) that Armstrong heard the Muslim call to prayer while on the Moon, and having heard it again on Earth he converted to Islam. Armstrong has several times – respectfully but firmly - denied this; an explanation for the birth of the legend may be that Neil Armstrong’s address is in an Ohio municipality called Lebanon. [7].

It was indeed a spiritual journey; Buzz Aldrin recalls the audience with Pope Paul VI, that presented the Apollo 11 crew with a Lladro porcelain figurine representing the Three Kings. The Pope – Aldrin says - "spoke of a mission that had taken place two thousand years before. It was a mission undertaken by three men who were guided by the stars so that they might deliver a message to all humanity. Such a beautiful comparison with our mission moved us all very deeply indeed" [8].

3. BENEDICTION AND DISAPPROVAL

The lunar journeys of the astronauts carried the blessing of most religious people, from clergy to laymen, that prayed for the success of the missions. The Apollo 13 emergency prompted a worldwide surge of prayers from people of all religions; newspapers in many countries published cartoons representing the earth with a pair of hands clasped in prayer [9]. The other missions were as well viewed with good eyes by most religious leaders; Man's voyages to the Moon prompted Pope Paul VI to deliver, on his Angelus of 7 February 1971, a "Hymn to the Glory of Man", reading inter alia: "Honour to Man, king of the Earth, and today prince of the heavens!" [10].

In July 1969, YukiTaka Yamamoto, a Japanese guji (high priest) of the Shinto tradition, came to the United Nations Headquarters Interfaith Chapel where he performed kiganai – a prayer of purification for the success of the Apollo Mission to the Moon. Asked by a reporter what good an ancient ceremony would do for a rocket on its way to the Moon, guji Yamamoto pointed out that three men were flying to the Moon for the purposes of peace and science, and if their hearts and minds were not united and pure, they could not succeed in their mission; it was therefore necessary to offer a proper prayer by a priest so that the astronauts might leave the Earth in a purified state. [11]

Man's first lunar journey was a reason for celebration by the Hopi tribe of Native Americans. In their folklore, a legend speaks about the light of a new day beginning when the Eagle lands on the Moon. And the Eagle spaceship has indeed landed on the Moon in 1969. [12]

Other Native people were however unimpressed by the Apollo success; to the Eskimo, going to the Moon was a matter of routine, given that their flying shamans would - they say - often journey into the sky [13].

There were voices – albeit few – that regarded the lunar voyage as a sacrilege, as a trespass into God's territory or at least as something wrong. Apparently, the Indonesian island of Bali lodged a protest at the United Nations against the US for desecrating a sacred place. [14]; in India, astrologers wondered if the Moon was too tainted for use in soothsaying now that a man has walked on it. [15] In Korea, some of the older population believed that the Moon was teeming with life - especially a rabbit that granted wishes - and to them, being on the Moon and interacting with it was sacrilegious and would bring consequences [16]. The strongest protest was prompted, however, not by the Apollo missions, but by the later placement of an ounce of the ashes of planetary scientist Eugene Shoemaker aboard the Lunar Prospector in 1998. Navajo Nation President Albert Hale protested this, in his words the Moon being "a sacred place in the religious beliefs of many Native Americans" being "sacredly, a gross insensitivity to the beliefs of many Native Americans, to place human remains on the Moon". In reply, a NASA official promised the U.S. space agency will never be so insensitive again as to place cremated human remains on the Moon without wide consultation. [17]. This may have impact on the future plans of Celestis Inc, that already offers burials in space, and on the plans of Sunray Co. of Japan, that at the end of 1992 proposed building a cemetery on the Moon because - they say - "Now that many Japanese live faraway from their family graves, [...] it is better that graves be built on the Moon, because we can see the Moon almost every night." [18]

4. LUNAR CHARMS, OFFERINGS AND RELICS

It is very often that astronauts took with them religious items to the Moon. Some were carried like amulets that would protect them; other relics were carried to the Moon and brought back to Earth in order to attain a magical aura of having been to another world. And some items were left on the Moon as offerings - like a message from Pope Paul VI containing the text of Psalm 8, placed on the Moon by the crew of Apollo 11 [19].

At a point, the number of religious relics flown on the Moon reached vast proportions – due in part to the then rather loose rules of what could be taken into space. Astronaut Al Bean confesses that "as long as it didn't get too heavy, you could carry lots of stuff [...] I took things for so many different people [on Apollo 12], when I got home I mostly didn't know who they were for. [...] So I still have three tiny gold crosses, two small Pope Paul VI medals, and one Holy Bible on microfilm that belong to somebody who has never asked me for them." [20] The Bible he refers to is the "First Lunar Bible", a special microfilmed edition of the King James Bible that can easily be read under a microscope. The Lunar Bible, taken by the Apollo astronauts on several lunar missions, was produced by
the “Apollo Prayer League”, being “crafted by the hand of man to glorify God on another world”.

While most of the religious items were of Christian provenance, this was not an absolute rule. Farouk El-Baz, a leading figure of the Apollo programme, gave the Apollo 15 crew a copy of the first Sura of the Koran for their lunar mission, in order to protect them [21]. The passage reads in part “Praise belongs to God, the Lord of the worlds […] Wee serve and Wee ask for aid. Guide us in the right path […]”.

Unity – an interdenominational religious movement – brought as well its contribution to the spiritual journey of the Man to the Moon; the Prayer of Protection written by Unity’s Poet Laureate James Dillet Freeman was taken aboard Apollo 11, and it reads: “The Light of God surrounds us / The love of God enfolds us / The power of God protects us / The presence of God watches over us / Wherever we are, God is.” [22]

While humankind left religious items on the Moon, it also brought back religious relics from there. Embedded in the Space Window of the Washington National Cathedral is a piece of lunar basalt returned by the crew of Apollo 11. The Space Window is the result of the efforts of two NASA administrators - Thomas Paine and James Fletcher, and of Francis Sayre, then Dean of the Cathedral. On a solemn ceremony held in July 1974, Neil Armstrong presented to the Dean the lunar sample – a “fragment of creation from beyond the Earth to be embedded in the fabric of this house of prayer for all people.” [23]

5. CONCLUSION

Spirituality has always been a part of space and lunar exploration, making no exception from Einstein’s rule: “Science without religion is lame and religion without science is blind”.

6. REFERENCES

6. Planet Arrington <planetarrington.com> [defunct]
21. NOVA Online Television, *To the Moon*, Broadcast Transcript NOVA #2610, air date 13th of July 1999 <www.pbs.org/wgbh/nova/transcripts/2610totheMoom.html> [defunct]
23. Lindsay, H., *Tracking Apollo to the Moon*, *Springer Verlag, 2001* <psc.g.org.au/~jsaxon/space/book/Apollo11.html#Dedication>
Introduction to Astronomy - An Outreach Project
Douglas P.G. Smith
Department of Mechanical Engineering, University of Strathclyde
Manor House, Gosford Road, Longniddry, East Lothian, EH32 0LF, United Kingdom
douglaspogsmith@hotmail.com

ABSTRACT

This is the plan for an Outreach Project, to teach school children the joys of Astronomy and the importance of Space. It also includes my reasons as to why this is necessary, and the ways in which I believe the issue should be tackled.

INTRODUCTION

Most students wouldn’t think of working in the Space sector because it seems to detached from everyday life. They don’t know enough about it, because it doesn’t get enough publicity. There were some stories in the news last week when it was announced that Mars Odyssey had detected subterranean ice, but few people realise the implications of this discovery.

The average person thinks that space and science don’t have much to do with them, that it is dictated by a handful of top scientists and engineers. It seems very distant and, and the possibility of playing a role in it is not realistic. Most grown up people have at least a vague interest in space and science, but they are ambivalent about what is being done. They are glad to accept the results and findings and leave the thinking and research to others, because they are perhaps put off by technical jargon and equations.

When most people think of space, they think NASA, and don’t realise the amount of work being done throughout the rest of the world and closer to home by ESA. Therefore people are discouraged because that seems so unattainable, especially not being a U.S. citizen, and they are unaware of the possibilities and opportunities that exist in Europe.

Space seems to have lost its street cred, and it doesn’t seem to be in touch with the majority of the population. Ask someone to name a Formula One driver and they could probably name a few; ask them to name an astronaut from not more than 20 years ago and they would probably have no idea. Millions of people watch Formula One drivers doing laps round a track, and worship them like heroes; so why don’t astronauts who are flying into space get any recognition and publicity outside of the industry. This is vital in order to have more children saying when asked what they would like to do when they grow up – I would like to go into space! Space must be cool once again, like it was during the space race, like it is to be a footballer or music star nowadays.

Since the Space Industry is primarily funded by the Governments through taxes, it is important that more people have a say about what they want, because if people are uninterested by it then space budgets will be cut and redistributed to other sectors. Therefore it is necessary to educate the people about what is being done in space, what is planned for the future and why the advancement of science and technology to explore the universe and try to discover its origins are so important.

This is especially worrying to me as a UK citizen which has a very limited space industry. There is little awareness in the UK of space activities – for instance, the vast majority of the population won’t have heard of the British National Space Centre (BNSC), and even fewer know what it does. I also wonder what percentage of Edinburgh’s citizens knows that it has been awarded the title of ‘European Space City’.

Luckily we are starting to see improvements with projects like the Beagle 2 Mars lander, which will hopefully raise public enthusiasm and interest, so that we realise the role that Europe is playing in Space. Lord Sainsbury the Science Minister is doing a good job of investing in space activities, but with a limited budget it is not possible to enter into as many as is desired. The opinion of human space travel has changed, but if we were to invest in the International Space Station, it may be necessary to close other facilities like the Radio Telescope at Jodrell Bank, which would be greatly missed.

There is a lot of good work being done by the ESA Education Office to promote projects like SSETI, Parabolic flight and the World Space Congress, but these are severely under subscribed and need much more

promotion. For instance there were 600 applications from students wanting to go to the World Space Congress in October, which is great. But this is a free trip to Houston to meet all the top people in the Space Industry and find out about the most up to date research being done – there should be 600 applicants from each country! But if people could see that there is an opportunity to pursue a career with limitless challenges, possibilities and significance, then interest should soar.

In order to try to alleviate the reluctant attitude of students it is important to reach the youth of today at a grassroots level and motivate them into pursuing an interest in the space industry. I intend to contribute to this by organising ‘Introduction to Astronomy’ sessions aimed at school children but open to anyone who is interested. Although this may only reach a small number of children initially, after participating in these introductory sessions I will try to help the children to set up Astronomy and Space Societies and Clubs at their schools so that they can spread the word to their fellow pupils.

AIMS:

- Get the youth of today interested in space forever.
- Give people an introduction to astronomy and space.
- Give people a chance to use a telescope who would not otherwise have the opportunity.
- Educate children about the importance of Space, and the work that is being done.

- Let people have a go of a telescope before committing themselves to purchasing one.
- Inform people about the current space activities being pursued by ESA and other countries at the present, and plans for the future.
- Teach people about the importance of all aspects of the space industry.
- Present a fun way of learning about Space.
- Teach people about ‘Technology Transfer’, how developments in the space industry are being used in industries on earth.
- Educate People that there is more to Space than Science and Engineering, like Law and Medicine.

PLAN:

- Set up an Astronomy Club in my local area, especially welcoming school children but open to all.
- With several clubs set up I hope to persuade local council and telescope shops to help provide telescopes or discount for the Clubs.
- Set up a similar Club at local schools, with the help of teachers to run it.
- Persuade the children who attend the club to set up their own clubs at their schools, to promote the idea to as many people as possible.
- Do a talk at my University to inform people about what is being done in the space industry and the opportunities to get involved through initiatives like SEDS, Parabolic flight, and the World Space Congress. I will also promote the University’s Astronomy Society and SSETI club, so they gain more members.
THE LUNARACE
A PUBLIC OUTREACH, INVOLVEMENT, EDUCATION AND SUPPORT MISSION

Spitzl, H. (1), Bouquet, F., Arafune, K., Contino, M-C., Fontaine, TH., Freihofer J., Grey, I., Leindecker, W.,
Lintchik, E., Meierink, G., Pauly, K., Shen, Z., Simi, N., Summerer, L., Weinmann, G., Yoon, J. (2)

(1) Vienna University of Technology, Austria. Email: helm@spitzl.cc
(2) International Space University, Summer Session 1999, Nakhon Ratchasima, Thailand

ABSTRACT

Today’s level of technology allows for many fantastic missions to space. Funding of these missions is a problem, because governments are cutting space budgets and commercial expenditure in space is minimal. The major obstacle to achieving global involvement into large scale, economically viable space enterprises is the lack of public involvement, education and support.

At the 1999 Summer Session of the International Space University, the LunaRace (LR) mission has been designed. With its extensive public outreach program before, during and after the race, this mission could be the first to bridge the gap between space and public. In national and international design contests, the most promising rover designs will be selected. Similar to the Tour de France and Paris-Dakar, the LR will be a staged event from the Apollo 17 to the Luna 21 landing site and back, during one Lunar Day. During the remaining sunlight after the race the surviving rovers will be used for public outreach purposes.

This LunaRace will be a stepping stone for future human space exploration beyond low Earth orbit. Next to the technology pull it implies, it has a high chance of boosting public support and education that brings the institution of commercially viable space enterprises a step closer.

1. INTRODUCTION

Last summer marked the thirtieth anniversary of Armstrong’s first steps on the Moon, but since the last lunar landing in 1972 humans have not ventured beyond low Earth orbit. The main obstacle to human expansion beyond LEO as well as space exploration in general is neither technological nor economical, it is the chronic and worsening lack of public (political, business) interest and support for such endeavors.

The moment a mission fails to be appealing to the public, a vicious circle is started. People lose interest in space exploration, governments and business who follow public opinion and customer demand allocate less and less money for space exploration activities, and smaller budgets again make it even harder for follow-on missions to be interesting and appealing.

Therefore, the primary goal of the precursor missions is to increase public awareness, education and support.

2. MISSION DESCRIPTION

An intensive brain-storm session yielded a large number of possible missions that were eligible to be designed in further detail. The LunaRace mission was selected as being the most promising mission to boost public interest, support and education, while also presenting the contestants with a huge technological challenge.

With this LR, public interest will be achieved by involving as many people of all ages as possible in all areas of the mission from beginning to end. Schools, universities, companies, individual groups, etc. will participate. The race resembles a mix of the famous Tour de France and the Paris-Dakar Rally. Current rover technology could allow the rovers to travel at speeds of around 10km an hour on the moon. Each rover will carry efficient cameras for teleoperation purposes but also for the general public to watch at home or in specially constructed pavilions.

3. PARTICIPATION

In order to increase public awareness to its fullest, it is desirable for all people around the world to interact as much as possible with the race. This can be best achieved by allowing private persons, universities, schools, the military and companies to have equal opportunities to compete for a position on the lunar race track. Each of the five continents of the world will have a lunar rover representing it and then there will also be another five consisting of the best five runners-up, whatever their geographical location. This way the quality of the rovers will be high. To be able to compete for a position, a registration fee should be


© European Space Agency • Provided by the NASA Astrophysics Data System
charged, the amount of which is dependent on the status of the organization. Companies will be required to pay most and schools will pay least.

Many companies can participate indirectly in the race. In yacht races, etc., most, if not all of the funding comes from sponsors such as alcohol, cigarette, camera film, etc. companies. Companies will wish to sponsor one of the rovers as it will be great promotion of their products and/or service. Telecommunication companies would want to make sure that they get coverage of the event by providing cameras either at mission control or for the rovers. Computer companies can compete to provide sophisticated software and so allowing the rovers to be as dynamic as technology allows. Information systems can also be created so to create dynamic web pages for the ultimate real time link up.

4. PRIZES

In the staged LunaRace there are points for having the best overall time, best points at each stage, being the best climber, for driving the most aggressively and winning halfway sprints during straight-line stages. Each win earns points towards a rover’s overall score. The prize for each of these will be in the form of medals, specially designed according to the nature of the competition. Each stage’s winner gets to name the specific location of that day’s checkpoint. The rover with the highest number of points will win the gold medal and gets to name the created Lunar Race Track. A final medal should also go to the team whose rover was the cheapest to build.

5. ROVER QUALIFICATION

The first step for selecting rovers is through design studies. Each candidate must demonstrate in the study that the rover meets the imposed constraints and requirements. Innovative ideas are encouraged at this point of the contest.

The entrant field will be narrowed down to 20 candidates maximum and each continent shall be represented within that field.

An Earth based Rover Race will determine which rovers can best achieve the race on the Moon. The selection criteria will be endurance, robustness, speed and communication capabilities (including the lunar time delay). As the competition must remain intercontinental, the final selection will include at least one representative of each continent. The entrant field will be narrowed down to 12 candidates. The first 10 will participate in the LunaRace and the other 2 will be left as backups.

6. RACETRACK

The racetrack concept is a round trip rally around two historical spacecraft-landing sites. The selection of the landing sites is based on the following criteria:

- selection of international origin of landing sites
- historical nature of lunar sites
- geological features of sites
- relative proximity of sites one from the other

The following sites were selected:

Apollo 17

Launched by the USA on 7 December 1972, and landed on 11 December 1972 at the Taurus-Littrow site. Apollo 17 was the last manned mission on the Moon. The crew was composed of Eugene A. Cernan (commander), Ronald E. Evans (command module pilot), and Harrison H. Schmitt (lunar module pilot). The crew returned safely to Earth on December 17th. The mission studied geological features of the Moon such as a lunar seismic profile or gravity cartography and tested also life support experiments. The mission included the Lunar Rover Vehicle (LRV).

Luna 21

Launched by the USSR on 8 January 1973 and landed on 15 January 1973 at the Mare Serenitatis / Le Monnier site. The mission, consisting of a Lander and a Rover (Lunokhod 2), was dedicated to measuring various parameters of the light on the moon (solar X-rays, ambient light or laser ranging from Earth), collecting pictures and studying mechanical properties of the lunar surface material. This second USSR lunar rover had solar panels to supply power during lunar daytime, and an isotopic heat source for the cold lunar nighttime. The rover stopped without keeping the laser reflector visible, which is perhaps why it failed.

<table>
<thead>
<tr>
<th>Site</th>
<th>Lunar Longitude</th>
<th>Lunar Latitude</th>
<th>Location Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apollo 17</td>
<td>30.77° E</td>
<td>20.19° N</td>
<td>Taurus-Littrow</td>
</tr>
<tr>
<td>Luna 21</td>
<td>30.38° E</td>
<td>25.51° N</td>
<td>Mare</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Serenitatis (Le Monnier)</td>
</tr>
</tbody>
</table>

1 Note: The landing sites being almost at the same longitude, the race is oriented South-North/North- South.
The round trip race is a twofold challenge for participants. The first half of the race consists of being the first rover to reach an historical landing site. The second half of the race is to get back to the initial departure site. Returning back to the starting point of the race will also permit good media coverage at the finish line with high resolution TV by the lunar race lander.

Figure 1 gives an overview of the race track and the landscape in the area of the two landing sites. The perspective view of the figure causes the distances near the top of the figure to appear smaller than the lower part.

![Figure 1 Sketch of the race track](image)

### 6.1 Schedule

The LunaRace will take place within 14 Earth diurnals. The spacecraft lander will therefore touch down when the sun rises at the Apollo 17 site. The race will start at 18:00 GMT the following Earth day, in order to have the maximum TV audience. The race will end 10 Earth diurnals later. One stage will typically last 20 hours overall, of which approximately 6 to 7 hours will be spent racing. This allows all humankind to watch at least two stages of the race at a convenient time and greatly increases the public outreach character of the race. Figure 2 and Figure 3 shows the Earth locations where the indicated stages start at six o'clock.

![Figure 2 Lunar Day Cycle and Race Schedule Overview](image)

![Figure 3 Stage numbers vs. Starting Time on Earth](image)

### 7. TECHNOLOGICAL CHALLENGES

Robotic rovers, of which Figure 4 gives an example, will play a key role in preparing human habitats on, for instance, the Moon and Mars, before humans actually arrive there. Locomotion on the Lunar surface is a major issue in robotic rover design.

![Figure 4 Mars Pathfinder Sojourner in (a) stowed and (b) extended position (courtesy of NASA).](image)

To get an indication of current status of rover speed: Mars Pathfinder's Sojourner (Figure 4) was designed for a top speed of 1 cm/s. The Lunar Race Rovers will be specifically designed for speed and endurance. The speed and endurance requirements are likely to result in completely differently designed rovers than Sojourner. There are a number of considerations that constrain the maximum speed of the rovers:

---

2 Note more checkpoints appear to be near Luna 21, but this is only due to the perspective of the presented viewing position.

3 Picture Reference: APOLLO OVER THE MOON: "A View From Orbit", NASA SP-362

4 A lunar diurnal is equivalent to 28 Earth diurnals (of 24 hours approximately).
ON-BOARD POWER: This is limited by allowable rover size and mass

2.7s LUNAR TIME-LAG: This will complicate teleoperation at higher speeds

LOW LUNAR GRAVITY LEVEL: This will change the inertial response of the rover. The speed of a rover on a rough course on the Moon will have to be reduced six fold to have it behave in the same manner as it would on a similar test course on the Earth. This will require the rovers to be very stable.

As described previously, teleoperation of a high-speed rover in an environment that is delayed by 2.7 s is very complicated. Next to using predictive displays, an autonomous rover will not suffer from the time lag as it makes its own decisions in local real-time. This requires sophisticated, robust control algorithms presenting major software design challenges.

The difference in temperature between lunar night and noon is approximately 300 degrees centigrade (-180°C to +120°C). The rovers will need a thermal control system to keep excessively high temperatures from damaging the system.

The rovers will communicate with ground control in a direct manner. The up and down link will be required for teleoperation of the rover and to send real-time video coverage of the race at a minimum required frame rate, which will consume the major part of the bandwidth. A trade off will have to be made between high speed and communication performance in the sense of required power.

8. MISSION ARCHITECTURE

8.1 Trajectory

The mission architecture has been designed in a preliminary manner. The design starts with the trajectory selection. The trajectory (Figure 5) is from GTO directly to the Lunar surface, requiring 3.05 km/s.

Figure 5 Trajectory from GTO to the Moon

8.2 Lander

The spacecraft that lands the rovers on the Lunar surface is designed to remain on the Lunar surface. The rovers will not be returned to Earth. Figure 6 shows a landing craft as it could be implemented for this mission.

Figure 6 Lander Outline

At a total estimated mass of 1.6 tons, the lander can be carried by current medium launchers, such as the Delta 2 or Ariane 4.

8.3 Media Vehicle

As the whole mission is based on public involvement, round the clock media coverage is required. Not only are the participants required to broadcast video coverage of their journey, a dedicated Media Vehicle is provided by the LunaRace company which will cover the finish of each stage of the race and the activities between the stages. The MV’s camera will be mounted on a deployable boom (see Figure 7).

Figure 7 Media Vehicle will Provide Media Coverage from 4 km (drawing not to scale)
8.4 Ground support

Telecommunications between the Earth and the Moon shall be carried out directly from each rover, the MV, and the lander. The telecommunication equipment on the rovers, the MV and the lander should be small in size and require little power.

9. PUBLIC INVOLVEMENT, EDUCATION AND SUPPORT

The overall goal of the LR is to stimulate public interest and, moreover, public support for the exploration of space. Therefore, we shall seek to reach as many different parts of human society as possible.

9.1 Activities Before the Race

In the pre-race phase, the main activities used to spread the idea will be a variety of contests, designed to reach different groups of society.

The first approach is Space Olympics, a semi-scientific contest for high schools in space related fields, similar to the Mathematics and Physics Olympics. The final competition, as well as the winner of the Space Olympics can be publicized via intensive media coverage. This would allow advertising of the race to reach a wider public.

In a later phase, universities are invited to participate in a contest designing the best Scientific Experiment that can be carried out on the lander-spacecraft during the race on the Moon. The winning experiment will be mounted on the lander and a member of the designing institution will be principal investigator.

High schools can write proposals for the scientific use of Rovo-Minutes i.e. minutes of steering of one of the rovers on the moon after the race via tele-operation. Again, it is a committee that allocates the minutes, but this contest will be less competitive. It will be probably a few hundred schools that will have the opportunity to steer a rover.

We intend to address younger students via a contest for the Art Design of the Launching Rocket. All elementary schools around the world are invited to participate in designing the painting of the fairing. The winning school class will be invited to actually paint the real launcher in the proposed design. This can be a huge media event, showing enthusiastic children colorfully painting a rocket. Furthermore, it is also an important symbolic statement to demonstrate the peaceful use of rockets.

The most challenging contest will of course be the participation in the race itself. To provide a chance for non-space faring countries to participate, we plan to send the five continental champions (i.e. the rovers that won the qualifying races on each continent) plus the five rovers with the highest scores to the finals on the moon, similar to the soccer World Cup. We hope that private companies, national agencies, universities, as well as small groups of garage-operating "geniuses" will participate. In this way, we hope to stimulate much innovation.

Finally, to involve a part of society that is traditionally left out of space activities, we will hold a contest for artists for the design of moon sculptures. The Camerators, or penetrators that carry a little camera, will be sticks that are ballistically deployed like a javelin from the lander within an area of approximately 1 km in radius. Except for the small payload of camera plus communication device (with the lander), which is in the order of approximately 100g, the artists are totally free to create their sculptures as long as the weight limit of 0.5 kg is not exceeded. This way, we open the first Art Gallery on the Moon, and in addition to that, in the final phase of the race, when the rovers approach the lander, they will provide the Earth with lively pictures of the finish from many different angles. Although this will add another 5 kg to be landed on the moon and therefore will increase launching costs by several percent, we find it is a great opportunity to go one step beyond ordinary thinking in space activities. Figure 8 shows the planning of activities around the LR.

9.2 Media Coverage

To spread the idea of the race as well as gain public support, intensive media coverage is necessary. All the above mentioned activities in the pre-race phase will be intensively covered on TV, especially the final
competitions of each contest. These can be huge media events. Such a program will not only create a variety of marketing opportunities but will also serve as a means of public education in various fields such as telecommunications, robotics, rocket technology, celestial mechanics, and even thermo-dynamics and propagation of light. The aim is to increase general knowledge about space and to prepare the public for human exploration away from Earth. In order to raise interest all over the world, the events will take place in different countries.

Some of the previously mentioned contests, like the national and continental qualifying races, could be linked with highly popular sports events like Formula One racing the World Cup Finals or the "Super Bowl" in order to make the project well-known.

Another media element of rapidly growing importance is a website providing a wide range of facilities to every individual on request. In the pre-race phase the website will basically consist of two parts.

The first is the information part that displays the basic idea of the project, the credo of the organization committee, and the rules of the race. There will be application forms as well as detailed terms of participation available for download not only for the race itself but also for all the other contests planned. Even the Space Olympics will be described and covered here, thus providing the link with the otherwise unrelated rover race. The website will be the information headquarters of the project with all the common features of an elaborate web facility such as newsletters, discussion groups, questions and answers, and web-cam coverage of events, etc. A special feature will be a virtual reality race simulation that demonstrates the 2.7s time-delay in teleoperation on the Moon. In this way even children can learn about things such as the limited speed of light. Finally, prepared lessons about related topics will be provided for download, serving teachers as background information.

The second part is a teaching system in astronomy and astrophysics in general. Inspired by the projected popularity of the lunar race, an interested public may be receptive to introductory to advanced level education in related fields. Many very interesting teaching systems are being developed at the moment, some even equipped with artificial intelligence to individually adapt to students' needs. As a special feature of this teaching system, we will have an assessment section in which we award some rovo-mintes to students who have excellent learning progress.

9.3 Betting

Betting will be allowed in the Lunar Rover Race. People who are not at all interested in space activities will start to study rover technology in order to improve their probability of winning the bet.

9.4 Race-Phase - Media Coverage

When the actual race on the Moon begins, TV coverage will intensify. The rovers will provide ten different views of the racetrack via real time images. In addition to that the Media Vehicle sends the leader board information and lively TV coverage of the start and finish phase of each stage to Earth. The racetrack is chosen to pass historic previous landing sites as well as interesting geological features on the Moon. At the start of every stage the rovers will be head-to-head, thus providing exciting, lively race pictures. The important phases may be broadcast via the major TV stations and perhaps an additional channel with 24-hour live coverage during the 10 race days: a RVN or Rovo-News channel that broadcasts also via the web. The information provided will be enriched by a master of ceremonies (MC) explaining geological features, historic sites, evolution of the Moon and the Solar System as well as the technology used. This MC will serve as an educator for the public.

The information part of the web-site will be enhanced by a detailed description and technical specification of the participating teams. This is an important part of the terms of participation: all the technology and the innovations used have to be made available to the public as soon as the race starts.

9.5 Race-Phase - Public Events

As the rovers are partly or fully tele-operated from Earth, there will be a ground station for every participating team where people can visit and watch the contestants in action. We see this as an important part of the game: fans actually seeing and being in touch with the pilots. The strong connection between high-tech and physical human endurance during the ten days of the race is meant to create heroes, like in any other sport. The intention is to present the pilots as real athletes. We want to place the ground-stations at a public point of interest in each participating team's country. Around those ground-stations we plan to promote a variety of activities like pop concerts, fashion shows, expositions and the like. We want everybody to identify with their favourite team, to get into the 'we participate' spirit like in the Olympic games. In this way the Moon gets much closer in peoples minds. During the race we foresee the ground-station as a fair of public education.
9.6 Wrapping It Up: Activities After the Race

The prize award ceremony will be an event unlike any other award ceremony. This has to be a mega-event crowning the 'rovaunts of the Universe'. The prize will be gold, silver and bronze medals, as given at Olympic games. Furthermore, the racetrack will be named after the winning team. We hope the public will accept the race as a fair, honest contest in both engineering expertise and physical endurance, thus respecting without prejudice a winning team from any country. The credo is to portray the Olympic spirit not only during the race but also in future human space exploration.

After the race some rovers should still be functional for at least two days, even if they do not survive the lunar night. During that time some of them will be available for teleoperation by schools that have won 'rovonutes' in the pre-race activities. Also, museums or entertainment complexes like Disney World could have the opportunity to allow visitors a short period of experiencing real rover teleoperation. As the rovers will be at the landing site of Apollo 17, there will be lots of visually interesting things to visit, not to forget the previously deployed space sculpture garden.

9.7 Educational and Academic Outreach

The participating teams will tour around the world to give lectures as well as question and answer sessions in schools, museums and universities. In this way, youth not only gets into direct contact with the 'scientific heroes', but in addition to that the spirit of the race - that space concerns everybody and is part of everyone’s future.

After the race, the website will be enhanced by any scientific results gained. This is not meant to justify the idea of a race on the Moon. The overall aim is still to have fun and to shape public awareness, but there will still be several unique opportunities to carry out scientific research during the race:

• There are the ten different camera-views of the same area provided by the rovers as well as the view of the media vehicle that can be post-processed and combined to a 3-D mapping of the lunar surface with unprecedented accuracy.

• During the whole race a scientific experiment will be carried out on the lander. The very nature of that experiment is to be defined via the contest in the pre-race phase.

• Material scientists can investigate by using rover cameras the long-duration exposure effects of the Apollo 17 artifacts to the lunar environment

• The art 'camerators' could analyze the material below the lunar surface, by using small geological probes.

The results of these experiments will be published on the web-site as well as in refereed scientific journals. Finally, a scientific documentary about the whole project could be created, similar to the Universal series by BBC. This would be addressed to an interested public without specific expert knowledge.

10. FINANCIAL ISSUES

10.1 Preliminary Cost Estimate

A preliminary estimate of the costs for the entire four year LunaRace activity is $190 million US dollars. The majority of these costs are due to the spacecraft, $63.5 million, the media vehicle, $11.4 million, and the launch on a Delta II or equivalent class vehicle, $50 million. In addition, the cost of insuring the launch and the lander/rover payload separately for failure plus the insurance to cover potential lost revenue in case either the lander or Media Vehicle are lost or malfunction before or during the race is nearly $25 million. Table 2 below provides a summary of all major program costs without considering inflation or interest. In most cases costs were estimated using the analogy method.

<table>
<thead>
<tr>
<th>Table 2 Preliminary cost estimate for the LR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budget Item</strong></td>
</tr>
<tr>
<td>Business and Management</td>
</tr>
<tr>
<td>Salary</td>
</tr>
<tr>
<td>Public Outreach</td>
</tr>
<tr>
<td>Facilities</td>
</tr>
<tr>
<td>Audit</td>
</tr>
<tr>
<td>Advertisement</td>
</tr>
<tr>
<td>Operations</td>
</tr>
<tr>
<td>Qualification Events</td>
</tr>
<tr>
<td>Race Event</td>
</tr>
<tr>
<td>Mission Control Centers</td>
</tr>
<tr>
<td>Telemetry, Data Network</td>
</tr>
<tr>
<td>Equipment</td>
</tr>
<tr>
<td>Spacecraft and Launch</td>
</tr>
<tr>
<td>Spacecraft</td>
</tr>
<tr>
<td>Media Rover</td>
</tr>
<tr>
<td>Launch</td>
</tr>
<tr>
<td>Insurance</td>
</tr>
<tr>
<td>Program Reserve</td>
</tr>
<tr>
<td><strong>Total Preliminary Cost Estimate</strong></td>
</tr>
</tbody>
</table>

10.2 Preliminary Revenue Estimate

At this stage it is difficult to calculate a confident estimate of revenues from the Lunar Rover Race. The
focus of this activity was primarily to determine whether the race ought to be studied further for its revenue generating potential. Revenue sources include event broadcast fees for several events including
- multiple pre-qualification test events
- final qualification race event and the selection of the final 10 race contestants
- events at the mission control centers
- direct broadcast from the moon

These are ranked in ascending order of their revenue generating potential with the greatest being, of course, the actual broadcast of events from the moon. The value of these fees is based on the amount and value of the TV coverage time. A very rough estimate of the revenue potential ranges from over $50 million to well over $400 million. A somewhat conservative approach, which assumes only a few stations provide strong coverage for only race highlights, suggests a revenue potential of around $175 million.

Other significant revenue sources include:
- sponsorship of the event itself
- sponsorship of the lander and media rover
- sponsorship of the Mission Control crew and race officials (patches and clothing)
- merchandizing and copyrights
- sales of rover time during non-race activities

These revenue sources are even more difficult to estimate and would require direct marketing of the idea to potential sponsors. Again, as a very rough estimate, we believe the potential revenue from these sources could range from $5 million to as high as $50 million. These numbers are modeled somewhat loosely on the revenue generated by start-up professional sports teams and the value of naming stadiums and other public and private venues. Part of the problem is that this is a first time event. Should the LunaRace occur on a regular basis it is possible to forecast an increasing revenue stream as its popularity and sophistication grows?

Clearly further investigation is needed, but this preliminary research shows that there is significant potential for a positive return on this event (up to $450 million revenue vs. $190 million in cost, returning nearly $260 million in net cash). We believe that further research and revenue concept development will yield profitable results.

11. REFERENCES

e.a., 'Out of the Cradle', International Space University, Summer Session Design Report, September 1999


PSYCHOLOGICAL ASPECTS OF LIVING IN SPACE
- ARCHITECTURAL CHALLENGES

Sandra Häuplik (1)  Susanne Lorenz (2)

(1) Architect, Austria, 1040 Vienna, Johann Strauss Gasse 40/2/7, Email: Sandra@saka.at
(2) Architect, Austria, 1150 Vienna, Sechshauserstrasse 11/2/24, Email: sannalorenz@gmx.de

ABSTRACT / INTRODUCTION

Space missions have generally involved crews, drawn from a highly homogeneous pool (such as white, educated, young adult males) and functioned for limited periods of time. Future missions may involve crews drawn from a more heterogeneous pool and missions could eventually last years. 3 to 5-person groups are considered appropriate for the Space Shuttle and the first interplanetary missions.

In addition to the above mentioned topics the success of a mission will no longer be dependent only on safety issues due to technological progress, but sociological and psychological aspects will become important determinants off the success or failure of future space missions. To create and ensure the social and psychological balance an adequate spatial planning is essential.

In the following essay notions for a conception basis of designing a space station will be described.

1. ENVIRONMENT

1.1. The Physical Environment
Space itself is totally inhospitable. Specific hazards include risks associated with, radiation, weightlessness, microgravity, and acceleration.

1.2. The Social Environment
The astronaut coming from a diverse macro society is thrown into a micro society, defined by isolation, confinement, dependency and intense contact with very few people.

Withdrawal from the normal social matrix means the loss of variety in social relationships and the loss of variety of different roles. (To exercise one’s typical social roles is important for a complete sense of identity. Withdrawal from existing relationships removes known comparison points, limits the self-evaluation process and can therefore make it difficult to maintain a sense of personal continuity or identity).

1.3. Basic Reactions to Space-like Environments
- Impaired Intellectual Functioning (decreases in alertness, concentration and memory, distorting perceptual judgements → problems in case of emergency)
- Motivational Decline (Confinement begins with high motivation that decreases with time.)
  - increased feelings of helplessness and worthlessness
  - goal-directed activities decrease
- Somatic Complaints (insomnia, headaches, and digestive problems)
- Psychological Changes (depression and anxiety, “eruption of the unconscious”)
- Social Tensions (touchiness and social irritability)

2. CHARACTERIZATION OF SMALL GROUPS IN SPACE

Future missions must not only yield effective individuals, they must yield effective groups. (People, whose behaviours are compatible with one another as well as with the environmental systems)

2.1. Selection
Optimal crew selection, training methods, habitability and communication specification will depend upon such variables as Crew size, Crew heterogeneity, Mission duration and Mission objectives.

2.2. Gender and sexual stereotypes
Space travel has been male-dominated, but in future the role of female astronauts will definitely increase. On the one hand, mixed-sex crews offer social diversity, on the other hand, sex-role-stereotypes (leading, men taking unnecessary risks to impress women, women faking helplessness in the presence of men) and negative attitudes could be disruptive.

2.3. Sexual behaviour
Sexual bonds of any kind are seen as potentially disruptive because jealousies may arise as the result of other crewmembers “pairing off”. In addition a relationship that is functioning well under normal conditions could prove devastating under conditions of isolation and confinement. The Assumption that sex simply would not occur because of the level of

Professionalism of the crew members is not realistic, for within the Navy, shipboard pregnancies have become a problem. The periods of deprivations will be so long, that the effective internal restraints may lose their effectiveness. If sexual activity is banned, this certainly will add to tensions and could lead to violent sexual acts.

2.4. Age
Ideally a crew entering space would be young in the sense that the members have not yet become bored with one another but mature in the sense that they have achieved a high degree of interpersonal coordination.

2.5. Culture
Future missions will include a larger proportion of people from different cultures. Prolonged isolation and confinement could bring long-standing prejudices to the fore. Some conditions associated with life in space may minimize certain ethnic prejudices such as similarities (many interests and values are the same) between crewmembers.

2.6. Crew Size and Social Compatibility
Increasing the Crew Size increases the number of possible social relationships, as well as options for social stimulation, for developing friendship, for exercising varied role behaviours.

Personal qualities, that are of high value to most members of isolated and confined groups, according to Rawls and Hopper (1969) are: Social compatibility, Personal attractiveness, Emotional stability, Technical competence, Cooperativeness and Social Versatility.

But group members don't have to be completely "like-minded". A person with a need to teach might establish a satisfying relationship with a person with a need to learn. Competitive needs could result in frustration.

Need compatibility gains importance under conditions of isolation and confinement.

3. INTERPERSONAL DYNAMICS IN SPACE

3.1. Group Cohesiveness
Cohesiveness refers to the bonds that tie group members to one another. Cohesive groups tend to have tight boundaries. There is a sharp distinction between the "in group" and the "out group". It is also difficult for an outsider or newcomer to join the group.

3.2. Compliance, Conformity, and Independence
Individuals may fear that unorthodox suggestions will incur the leader's displeasure and lead to rejection. Under normal conditions, this may simply result in leaving the group, but under conditions of isolation and confinement, this is not possible. Prolonged rejection may lead to the pathological "long eye" syndrome. This may involve hallucinations, tears, loss of appetite, silence and sloth. This is not only punishing the rejected individual, but it robs the group of the services of one of its members. Fortunately, the effects of the "long eye" are temporary and vanish when the individual is reaccepted by the group. It is necessary to achieve a healthy balance between crewmembers' acceptance of social influence and a willingness to engage in appropriate independent action.

3.3. Group Performance
Group Performance reflects a complex interplay of many different factors. If group norms favour high levels of productivity, as it is likely in space, then high cohesiveness is associated with high performance. Structuring tasks in such a way as to maximizing motivation and commitment is important.

3.4. Leadership and Organisation
Leaders organize direct and coordinate followers. Most people in any multilevel organizational hierarchy fill both leader and follower role. Leadership is a two-way-influence process. There might also be a leader that is not identifiable. One person may organize and direct the group, while another attempts to satisfy the group members' needs.

The outcome of leadership process depends on the characteristics of the leader, the followers and the environment. Heavy demands will be placed upon people performing leadership functions in space. A failure of leadership can lead to severe consequences.

4. WORK AND LEISURE
Performance during spaceflights is important for the health and safety of crewmembers and to the success of the mission. The slightest upset in astronaut performance and behaviour can have a deadly impact on the mission.

4.1. Factor Affecting Work Capacity
4.1.1. Gravity
The accuracy of psychomotor performance in space is affected by the lack of gravitational pull. For example, navigation requires the execution of a number of human abilities, like arm-hand steadiness, finger dexterity, hand-eye coordination, perceptual speed, rapid reaction time. It is very difficult to keep track of small tools in an environment where items can simply drift out of sight.
Under conditions of microgravity, belts, pockets, snaps and acres are the order of the day.

4.1.2. Biomedical changes
Reduced cardiac activity and diminished musculoskeletal strength can affect capacity. Also space sickness is a factor to reduce work capacity.

4.1.3. Perception
Outer Space is characterized by sharper visual contrasts than we are used to on Earth, because the terrestrial atmosphere absorbs a percentage of incoming light. Light transmissions can be very abrupt. Also living under conditions of microgravity makes it difficult to gain a true sense of up and down.

4.1.4. Sleep Disturbances
In isolation sleep difficulties have been a repeated problem. Studies from Arctic/ Antarctic outposts show that sleep quality deteriorates and the men suffer from a dramatic form of insomnia, in which the deeper stages of sleep decrease or disappear. In space the problem appears to be related to high noise and vibration, staggered sleep schedules and bed designs. Sleep loss directly affects the concentration and vigilance, and indirectly the mood, attitude, motivation and performance efficiency. Wilkinson found that sleep loss has a significant impact on decision-making. Background noise must be kept within tolerable levels, especially in areas where astronauts sleep.

4.2. Factors Affecting Work Schedules

4.2.1. Circadian Rhythms
People can run entirely on their internal clocks, but in the absence of external reference points they will gravitate toward 25.4 hour days. Disruption in the body’s circadian activity (desynchronizes), can produce physical symptoms as malaise, insomnia, appetite loss, and nervous stress. If the crewmembers engage in social activities, like having breakfast at a certain time it is able to combat the problems associated with “free cycling”.

4.2.2. Wake-sleep cycles
Also the importance of sleep must be considered. Changes in the quality or duration of sleep can affect performance. When wake-sleep cycles are altered from a 24 hours pattern, this change affect sleep quality and quantity.

4.2.3. Work-rest cycles
Crewmembers might also be given flexibility in determining their work schedules, and a rotation of crew duties.

4.2.4. Stress
The more demanding the schedule, the more severe is the performance impairment because of stresses of lost sleep and additional workload.

4.2.5. Factors Affecting Workload
Workload will be defined as the total work demand placed on the operator.

4.2.6. Overload
Role overload exists, when a position carries a combination of tasks that cannot be completely performed without undue stress. When the operator is overloaded, action is confused. In contrast, attention tends to drift, performance is lethargic.

4.2.7. Fatigue
Fatigue represents a reversible impairment of performance as a consequence of over or under loading. It is necessary to balance work schedules in space to minimize condition of fatigue that reduce work capacity, and result in irregular or disordered performance. Physical fitness, adequate rest, weight control, nutrition and diet and moderate use of alcohol and tobacco and also amphetamine drugs can reduce astronaut’s fatigue. The most studied condition of task under loading involves the requirement for visual or auditory attention.

4.3. Factors Affecting Work role
Conflicts occur, when the person’s expectations disagree with the expectations of another person in related roles. In isolation minor issues may be blown out. Allowings persons to perform many different roles can encourage the interaction between the crewmembers. Improving communication and articulating tasks among occupants of adjacent roles minimizes misunderstandings.

4.3.1. Motivation
The important theme is that some of the conditions that motivate persons on earth are unavailable in extreme environment.

4.3.2. Rewards
The most prominent rewards are pay and social recognition. Under the condition of extend duration spaceflight some of the pay’s utility like luxury get lost because there are no stores of luxury items. Under those condition relatively minor luxuries like a private room, telecommunications become important. Also the social recognition becomes less important because of space travel becomes routine. At some time in the future the risks and discomforts in space are likely to become disproportional to the rewards.

© European Space Agency • Provided by the NASA Astrophysics Data System
Work is important to crewmembers on several reasons. Crewmembers are selected on bases as intelligence, drive and interest in achievement. Such people can be expected to place a high priority on work.

4.3.3. Variety of work available
We must ensure that the training program does not reduce the interest value of mission responsibilities. When all work is over learned, the problem of boredom is exacerbated. One strategy is to over train astronauts on tasks that are crucial for safety and success, while leaving no critical tasks for scheduled for in flight learning. A reduced or no varying work environment can produce dramatic decrements in various aspects of performance. Satisfaction is high when work assignments are congruent with personal interests. To ensure that everyone gets on well with different and changing tasks, a clearly arranged interior design is necessary. A reset function should be built in for equipment that is shared by the group.

4.3.4. Sanctions
A major task is establishing a legal system, which is simple, has clear jurisdictional boundaries, and preserves the interests of security, justice and civil rights.

5. COMMUNICATION
The term communication refers to the exchange of information between one person and another or between one social system and another, for the purpose of reaching a shared understanding.

5.1. Verbal Communication
Verbal communication refers to communication through grammar and syntax. (what is said) The ambient noise level may make oral communication difficult.

5.2. Message Complexity
Astronauts, like other specialists, have developed a special language in which acronyms and other speech habits are used to improve efficiency by condensing words. The use of such a language could lead to misunderstanding with other crewmembers with a different field of interest. In addition the ability to process information is reduced when the environment is complex and demanding. Procedures must be developed to ensure that important messages are received and understood.

5.3. Nonverbal Communication
Communication through the linguistic method (grammar and syntax) is supplemented by:
• The communication through the amplitude, rate and tenor of speech
• The communication through facial expressions and gestures (which is different in a non vertical environment)
• The communication through distancing or placement
• Nonverbal messages can substitute verbal messages, modify verbal messages and contradict verbal messages.
• Weightlessness has the potential of interfering with these aspects of communication. Artificial atmospheres could transform or alter voice quality.

5.4. Crew Heterogeneity
We can expect that crew heterogeneity in space will be associated with an increase in the likelihood of nonverbal and verbal miscommunication. For example interpersonal distance zones are specific vary within certain societies.

5.5. Mediated Communication
The current direction of systems' development is toward greater flexibility and compatibility. Updating a system should be possible without changing the equipment. It has to be ensured that the communication systems will process crucial information, even under overload conditions.

6. CRISES AND COPING MECHANISM

6.1. Individual Response to Threat
It is unlikely that all future space travelers will possess the technical skills to deal with emergencies. But all must possess the psychological constitution to withstand such threatening events.

6.2. Conceptual view of fear
(According to Spielberger 1972)
Fear: is the emotional response to a real or objective danger. It is proportionate to the magnitude of the danger.
Anxiety: is the fear-like reaction that occurs in the absence of specific threatening circumstances. It is unrelated to the danger.

6.3. Consequences of Fear in Space
A lack of understanding or a denial seems to be the first response to threat. For some individuals involved in serious accidents, a complete depersonalization occurs. These individuals describe a perceived slowing of time just before the accident, a heightening of awareness, and a seeming separation from the physical body. They report a feeling of having "left their bodies" and of witnessing the accident from some distance (Mitchell 1981). There appears to be a close tie between fear and impaired performance.
Fenz has shown that when one has some time to consider the threat, the fear response can vary significantly. Studies with parachuting shows that experienced jumpers become fearful well in advance of the event, becoming calm as the event approaches. Whereas for novice jumpers fear increases steadily.

6.4. Training for Danger
There is considerable evidence that a person can be trained to deal with dangers of a particular situation. (see jumpers) There is also an approach to train individuals to deal with danger generally. The astronaut training program has not explicitly adopted the value of generalized stress training. However scuba diving, survival testing and flight training have been part of this program. In Future flight training might be only available for certain candidates.

6.5. Group Response to life-threatening situations
Due to a spaceflight accident, panic could develop within the space crew, especially if the group were large and diverse (Hartmann and Flinn, 1964).
Rosengreen (1974) identified 3 conditions which could lead to panic:
An individual
• sees a threat toward his or her own existence
• sees a possibility to escape, and
• believes this possibility is soon to disappear (e.g. because it is not sufficient for all wish to use it).
Due to the fact that panic does not contribute to an individual’s chance for survival in space, the probability of panic on-board is remote. (Although the crew might experience intense fear).

6.6. Responses to Threat
A crisis is accompanied by a general state of instability within the group (Hartman and Flinn, 1964). The quality of leadership will determine to a large extent how quickly stability can be restored (Llano, 1955).
Attribution of blame is a common occurrence during post-crisis recovery. Blaming serves a psychological need. It results from seeking an explanation to something which cannot be explained satisfactorily. It is an attempt to control the future by creating a structure whereby inexplicable events become explicable. Blaming has also an instrumental effect. A person is more likely to offer help if the injured individual implies that the person is to blame. (Schwartz and David, 1976)

6.7. Environment Adaptation
Some psychiatric conditions are thought to represent a failure of the individual to adapt to the environment. Andrews (1976) has found that suppression, sublimation and humour lead to a healthy adjustment. Defences such as fantasy, projection, or passive aggression can lead to disruption in the individual’s relationship to reality.

6.8. Selection
The first defence against the occurrence of a psychological disturbance in space is personnel selection.
It should be remembered that standardized psychological tests are intended to select out individuals (mentally sound persons, to identify those with potential pathology). The more difficult and NASA’s primary task is to select in individuals (i.e., those with high levels of psychological and emotional health).
But what is a mentally sound person? There seems to be a link between maturity and emotional health. Lindemann describes a mature person as one who is able to perform required tasks, meet ordinary stresses of life without disintegration, operate without making others sick, and adapt his or her own perceptions to reality.

6.9. Crisis within a Group
The threat posed by an individual or in the relationship among individuals in space could have a devastating effect on the entire crew and mission.
It has been observed that a person experiencing a severe psychological disturbance will show impairment in at least one of four functions:
• Thinking (disordered, as in schizophrenia)
• Perception (distorted, as in paranoia)
• Mood (elevated, as in mania or despondent as in depression)
• Impulse (violent, as in hebephrenic excitement or frozen, as in catatonic stupor)
Impairment in mood expressed as depression is the most commonly observed form of psychological disturbance. Depression is marked by lack of ability to concentrate, feelings of worthlessness, irritability, weight loss, hopelessness, guilt, etc.

It has been found that events or series of events can aggregate or trigger a psychological episode. Depression is sometimes interpreted as anger turned inward. Under normal Earth conditions some individuals turn their anger inward; some direct their anger toward others. In confinement, even normally outward-directed individuals could be subject to depression.
For future space travellers, an especially worrisome form of depression is one following the relational breakdown between the individual and the group.

6.10. Treatment
If a person should experience psychological problems in space, the requirements of other crewmembers should be taken into account in selecting a treatment strategy.
Therefore procedures must be developed before flight. It can be expected that drugs, commonly used in treating on earth, will play a similar role in space. For standard psychotherapy a trained therapist would have to be on board. Alternatives range from totally impersonal computer counselling to telephone therapy, to two-way video phone (anonym) or closed-circuit TV sessions. Awareness and Sensitivity training could be very helpful and essential in avoiding relational problems or in dealing with such problems when they arise. Exercise has been found to have positive therapeutic effects in helping an individual maintain a sense of vitality. Meditation increases sense of calm, having greater control of live and leads to higher levels of awareness. Other techniques to control negative reactions to stress include progressive relaxation, autogenic training, and biofeedback. The value of interacting with and caring for pets could be considered.

6.11. Substance Abuse
It is unlikely, although not impossible, that there will be problems with illegal drugs. Much more likely would be the abuse of drugs brought aboard for medical purpose. Alcohol is presently the only drug accepted for recreational purposes in our society (except for Marihuana in a few countries). We know that isolation and confinement lead to increased emphasis on food, or heighten oral needs generally.

6.12. Grief
Grief in its extreme form is among the most profound of all psychological and physiological stressors. Grief can be experienced in response to any significant loss, such as the loss of status, the loss of income, or the loss of valued possession. This loss may be temporary or permanent. Grief is divided into certain stages:
It has been found that there is a relationship between grieving and mortality. The rate among survivors due to a variety of causes was found to be more extreme for younger than for older persons, and for men than for women.
Time for recovery varies with the severity of the loss. It is agreed that experiencing grief is essential to recovery. Because of the extreme pain of grief, many people try to avoid it, resulting in delayed or distorted reactions.

6.13. Death aboard the Spacecraft
A death aboard the spacecraft can be expected to be an extremely traumatic event and will have a profound effect on the crew. Crewmembers will be required to deal with both the physical and the psychological demands of the situation.

6.14. Homicide and Suicide
Confined groups seek to avoid confrontation. But also acting-out tendencies have been observed. The likelihood that a crewmember would injure him or her and the possibility of a suicide occurring on an extended flight should be considered.

6.15. Crisis Intervention
Steps in crisis intervention are assessment of the problem and the person’s response for it, planning of the therapeutic intervention, intervention (exploring and testing coping strategies and replacing some of the roles disrupted by the loss) and Reaffirmation of the progress made (reviews the strategies that have been successful and helps the individual plan the future). Some or all crewmembers could receive intervention training as part of their overall orientation.

6.16. Implications for Space
A crisis in space could arise from any number of psychological and/or relational disruptions occurring with the space crew. Depression poses a particular threat. Being ostracized from a group that is itself isolated can be expected to result in severe pressures on a space traveller.
One suggestion is the use of a “buddy” system, where each individual is assigned a partner whose responsibility it is to understand that person’s perspective and to defend him or her, if needed, to the larger group.

7. HABITABILITY

7.1. The Social Space

7.1.1. Personal Space
A simple definition was offered by Robert Sommer, 1969: “Personal space refers to an area with invisible boundaries surrounding a person’s body into which intruders may not come”.
Researchers found out that greater interpersonal distance was required for holding a conversation to one another in the usual ways (both “right side up”)

7.1.2. Privacy
Privacy regulation is a culturally universal process. It is necessary for individuals and group survival. Privacy is controlled by four behavioural mechanisms:
- verbal content and structure
- nonverbal behaviour (body language)
- environmental mechanism
- culturally base norms and customs

Basic functions of privacy are:
- management of the interaction between one person and others
- define self in relationship to others
- most central function: self identification
• helpful for achieving concentration when other people are distracting
• Promotes “rest and recuperation”
• Helps us control the images, that we project to other persons and thereby regulate the relationships we have with them Individuals use the environment and the physical objects in the environment to define themselves.
Visual access and visual exposure are the two key aspects of privacy regulation.
In See-lab II, some aquanauts commented that teammates who didn’t always see eye to eye were able to get along for the period of the mission.

7.1.3. Crowding
Crowding is a more complex variable than density, and a person in a high density situation may or may not feel crowded.
There are three perspectives on crowding:
• stimulus overload
• behavioural constraint
• ecological formulations
High density in private or working spaces is more closely related to stress than high densities outside or in public areas. Friends require less distance than strangers, informal groups less than formal. If individuals believe that they control the situation the feeling of crowding is reduced.
Self directed, high self esteem persons have lower spatial needs than individuals searching for identity, and laissez faire individuals have lower spatial needs than authoritarian persons. Crowded men were found to respond more negatively than women. Women become more cohesive and cooperative, whereas men become more remote and competitive. Over time, crowded women get more health related problems and showed less group stability.

7.1.4. Territoriality
Personal and exclusive use of an area is probably less important than the privacy requirement which such space would address. So, in space we have to search for solutions for the privacy problem which do not depend on the location.
Occupants can use personal items to “stake out” areas in the same way that students use textbooks and other personal items to gain temporary control over an area.

7.1.5. Complex and Multiple Stressors
Environmental pressure can result in complex patterns of response. Effects may first appear after the offending stimulus has been withdrawn.
Environmental stressors interact in three general ways:
• Addition
• Synergism
• Antagonism

7.1.6. Privacy in Space
In Space, personal property is necessarily limited. There is a lack of opportunity for personal territory. Close contact leads to problems of information management. This situation increases the probability that the individual will be caught unaware in behaviour he would keep private.
In space, astronauts must depend on each other for their day to day survival. Withdrawal behaviour in space is likely to be curtailed or specialized.
The individual privacy is important for the effective functioning of the group.
Another major aspect of privacy in space is the intrusion of outsiders on the privacy of the group. Group privacy rights can be expected to become an important issue.
Some spaces (library–small group recreation room) should provide for small groups of 2 or 3 individuals and should screen them from visual distractions and provide auditory privacy.
Anything that would help emphasizes a person’s individuality help offset privacy loss. (Decorating an area, Personally selected cloth)

7.2. The Physical Space
Habitability is a term which connotes a level of environmental acceptability. The conditions change with circumstances. Over the long term, conditions must support individual’s psychological health.
The most recent developments in this area have shifted the focus away from the components of habitability to the relationship between them.

7.2.1. Organisation of Interior Space
In the near term, astronaut’s habitats must be restrictive. A basic question concerns the minimum space. Important variables are the length of confinement and the number of individuals sharing confinement. More space per individual is needed as the number of individual’s increases.
A further question concerns the utilization of the available areas within the spacecraft. Weightlessness does permit astronauts to use space more efficiently than on Earth. Some astronauts prefer a room with a defined up and down.

7.2.2. Food
The most important question is how food will be used to fulfill psychological and social needs. Space travellers should share at least one meal a day together to maintain group cohesiveness.
Experiences in the Arctic and Antarctic conclude that men in confinement enjoy meal more and take twice as long to eat as men in general population. The challenge for future spaceflights is to provide sufficiently engaging activity options, so that the individual's
psychological and social needs can be met without a harmful over reliance on food or mealtime.

7.2.3. Hygiene
Limitations on bathing facilities and waste management problems have been high on the list of discomforts reported by individuals in confinement experiments. Although vastly improved over early systems hygienic facility still did not allow the comfort which is needed in a long duration fly. Several personal hygiene facilities should be provided and should insure complete visual and auditory privacy.

7.2.4. Temperature and Humidity
Variations in temperature affect human performance in diverse ways. Cognitive tasks are more affected by a high temperature than motor performance tasks. The ability to withstand heat is limited by the amount of physical effort required. High temperatures have also been related to temper outbursts and negative reactions to others. The visual reaction time and declines in skilled motor performance have been reported.

7.2.5. Interior Design
Attractive surroundings were found to be more important when crew were composed of both men and women, and when all male crews had free time. Décor was found to be most important when all crewmembers, men and women were without meaningful work. Interior design of a spacecraft should have build in flexibility. Flexibility is enhanced by differing between “hard” architectural features (wall, hatches…) and “soft” features (screens, moveable partitions…). There is an aversion to sameness. There is also a recognized aversion to clashing designs. In space the number of designs must be limited.

Visual variety can be achieved through
• Texture
• Material
• Colour

Windows are crucial, and photographs of landscapes can “enlarge” the environment psychologically.

7.2.6. Lighting
In this environment lightning becomes important. Proper lightning is important to safeguard vision, to minimize annoyance and to enhance the visual environment.

• Day-cycle and night-cycle
• Artificial light: for creating different atmospheres
• Day-light in work areas
• User-controlled lightening in private areas

• Use intensive lightning shifts to help the crewmembers adapt rapidly to a new time zone

7.2.7. Odour
Materials used in spaceflights are subjected to testing for odours as well as for flammability and toxicity. Odours present problems in future flights, especially when humans and animals live together in the same area. Ample airflow and air revitalization techniques are important. Odour could also be used as a tool to create diverse atmospheres: Smell of forest, flowers in recreation areas.

7.2.8. Sound - Noise
In the extreme noise can cause pain and damage to the inner ear. Also noise can be a psychological stressor. It can cause higher error rates and greater variability of performance. Noise is also thought to impact performance negatively by lowering motivation and moral. Noise over which a person has no control results in high level of stress. Typical noise levels in spacecraft are 60 -70dB, but for long term missions a desirable and safe maximum sound level is 45 dB indoors. Background noise must be kept within tolerable levels, especially in areas where astronauts sleep. One the other side the presence of noise is necessary and beneficial. Russian experiments have shown that unfamiliar and unusual music had a positive effect on the astronauts. Soothing music has been found to lead to helping behaviour.

7.2.9. Health and Leisure
People in confinement adopt an extreme work orientation and exhibit little interest in leisure activities. Until recently it was believed that a proper exercise program could reverse the significant physiological changes associated with the body’s response to 0 g. However, physical activity relieves depression and lessons anxiety. People who are actively engaged in their free time are psychologically healthier than passive individuals. So we must determine how to motivate astronauts to perform the needed exercises. Cognitive strategies are very important in maintaining an exercise commitment. It is likely that individuals in space will develop interest in movements unique to space. Astronauts on Skylab were enthusiastic about acrobating in space and suggested that all future space stations include a facility for acrobatics.
8. RESUME

Within a crew, consisting of 3-5 members living in such an extreme environment, stressful situations are likely to occur if there is not enough privacy, no adequate relation between work and leisure time and communication problems intern or with mission control. These factors also influence our live on earth, but are more amplified in isolated and confined environments and are endangering the success of the whole mission. To cope with these situations not only the training of the astronauts must be improved through incorporating psycho sociological aspects but also architects have to create a spatial environment, which decreases or eliminates such stressors.

9. REFERENCES

Connors Mary M, Living Aloft, Library of Congress Cataloging in Publication Data, (NASA SP;483)


Margaret J. King, Apollo 13 Creativity: In-the-Box-Innovation, Fourth Quarter 1997, Number4, Volume 31, p.229-308


John Sturgeon, Sex in Space, www.space.edu/LibraryResearch/sex.html

Human Needs in Space, www.belmont.k12.ca.us

Accidents in space, www.exn.ca/ISS

Constanze Adams, Light System Design Studies for Space Habitats (NASA, 2000-01-2464)

Soviet Psycho-Physiological Investigations in Experiments with simulated Isolation- some Results and Prospects, www.geocities.com/CapeCanaveral/Launchpad/1033/splviews.html

Erwin Riefler, Menschliche Existenz und Grundbefindlichkeiten in Extremsituationen,1989
Fischer Verlag, ISBN 3-8323-942-9

Bruno Bettelheim, Erziehung zum Überleben, 1980


THE CONTEXT OF SPACE ARCHITECTURE

Sandra Häuplik (1), Susanne Lorenz (2)

(1) Architect, Austria, 1040 Vienna, Johann Strauss Gasse 40/2/7, Email: Sandra@saka.at
(2) Architect, Austria, 1150 Vienna, Sechshausenstraase 11 2 24, Email: sannalorenz@gmx.de

ABSTRACT

So far architects were mainly working with the organisation of singular project-flows and their integration into the space mission system. The emerging tourism is changing the view of our society concerning space missions. Up to now they have been always seen as high-technologized and scientific processes but in the future more and more humans will be able to participate in trips into space which will become part of our everyday life.

Within this paradigm-shift the role of the space-architect changes as well.

By investigating into design concepts of space habitats new key issues will be found and through their interpretation a change of the designing process takes place. That creates a new context for space architecture.

For a successful accomplishment of a space mission, it not only requires, that the engineering secures the safety issues, but it also requires a specific and distinct design of the technological space, which is also characterised by a distinguished atmosphere, designed by incorporating issues such as light, material, texture and colour.

1. SHORT CHARACTERIZATION OF SPACE AND THE MOON

1.1 The Moon
Average day/temperature high 134°C
Average night temperature low - 170°C
Day / Night Length 13.66 Earth days
Albedo 0.07 (approx. ¼ of Earth)
Atmosphere hard vacuum
Magnetic field less than 10-3 of that on Earth
Seismic Activity magnitude of 5 or more possible
Surface Material fine, silica-rich sand and dust

Macrostructure: MARIA, HIGHLANDS (TERRAE), CRATERS (Mares), Rilles
Microstructure: Regolith, fine sand-like material

1.2 Topics
Gravitation: Moon: 1/6 Earth gravity, Space: 0 gravity
Protection from Radiation (solar radiation and galactic, seismic radiation), Micrometeorites and other Hazards to Health, Safety (Robotic Systems)
Dust Surface: Coating Abrasion of Mechanical Components
Temperature: -170°C – 134°C on the Moon
Lighting: 14 day/night cycle lunar except on the South Pole (eternal peak of light)
Communication: Constant communication is possible on the near side (70° eastern and 70° western longitude). Regions on the far side cannot communicate with Earth without the aid of radio satellites or relay antennas on the lunar surface.
Hard Vacuum: internal pressures of habitats: 0.6 to 1.0 atm

2. CONSTRUCTION/FORM

2.1 Static
Different gravity conditions and the lack of wind pressure consequences in new construction abilities.

2.2 Rigid Modules
The advantages of rigid modules are immediate utilization and standardization. They are based on international space technology.
Volume and mass limitation, form limitation and need for one site linking are the disadvantages. Also a site preparation is required. Typical materials are aluminium, lithium and composites.


© European Space Agency • Provided by the NASA Astrophysics Data System
2.3 Membrane Structures
They offer some advantages such as low weight, low cost, low storage volume and a free form.
The required layers of a membrane structure are an outer thermal protection coating, a micrometeoroid barrier and a pressure bladder, inner flame/gas barrier.

Other construction methods include self-erecting structures, foam rigidized structures and structures using in situ material.

2.4 Development
Important approaches have to be made in the perception of the relation between outer and inner space of the habitat. Furthermore, the investigation has to follow into the spatial structure which in the future should more adopt to the conditions and laws of space, leaving the paradigm Earth.

One suggestion would be an ongoing design workshop with professionals of different fields, (Engineers, Architects, Designers) to achieve a complete and comprehensive solution.

3. SPATIAL STRUCTURE

3.1 Human Psychological and Social Factors
The psychological and social factors like lighting, lower gravity conditions, safety, stress, community and privacy, sensation, place and identity will become important determinants offset the success or failure of future space missions.

3.2 Lighting
The need for adequate lighting in living or working spaces is one of the most important human factors.

Symptoms of inadequate lighting include eyestrain, reduced productivity, Seasonal Affective Disorder (SAD), depression, more sleep needed, moodiness, increased appetite, and weight gain lethargy.

People can run entirely on their internal clocks, but in the absence of external reference points they will gravitate toward 25.4 hour days.

An artificial day-night cycle could prevent disruption in the body's circadian activity (malaise, insomnia, appetite loss, and nervous stress).

A satisfactory light system should supply natural sunlight during the lunar and a sunlight-color balanced system for the lunar night. During the lunar night it might not be necessary to light the entire habitat, but some areas should be as bright as daylight.

3.2.2 Design Constraints
There must be some areas with bright light (>2500 lux).

In case the lunar base is built under the surface and during lunar nights a light-guide system could supply sunlight. Sunlight could be piped with heliostat mirrors for windowless areas.

Light areas appear less crowded than dark areas.

In work areas do not light the room, light the task. Pools of light define areas of interest and importance.

An effective exterior lighting minimize accidents.

Light can help to achieve a vertical orientation in space.

Adjustable lighting emphasizes a person's individuality and helps offset privacy loss. Light grazing adds texture, patterns, and expands one's sense of space. Dimming provides a soft and relaxed atmosphere.

3.3 Gravity and Ergonomic requirements
Lower Gravity changes body movements, posture, locomotion, reach capability, inclination of the body, reduction of friction to the ground and vision.

The inclination of vision on Earth (1-g) is 10° downward, in space (0-g) 25° downward and on the moon (1/6-g) 22° downward (interpolation).

Reactions to lowered gravity include dizziness, space motion sickness, and feeling of rotation, cardiopulmonary changes and muscle tissue breakdown.

3.3.1 Requirements in 0-Gravity
The architecture should provide an adequate access to all areas including mechanical areas.

Disorientation is to be avoided. Weightlessness does permit astronauts to use space more efficiently than on Earth. Some astronauts prefer a room with a defined up and down.

Due to the fact that in the neutral body position the toes are pointed, flat floors should be avoided.

The work station should be designed for this neutral body position. The arms tend to float out and away from the body. So, waist high work stations should be avoided. Also situations which require the astronaut to bend forward must be avoided. It is much easier to pull the feet to the chest than bend over.

Windows should be located in the lower 45° sector.

3.3.2 Requirements in 1/6-Gravity
It is easier to reach upward and more difficult downwards.

Humans
- walk and run 40% slower on the moon (Earth 1.2m/sec; Moon 3m/sec)
- the stepping rate is less than on Earth
- lean further forward and swing their legs further forward
- tend to walk stiff legged with very little flexing of the knees
- have a reduced ability to change direction quickly
- stopping and turning are difficult
- sleep in horizontal position, requiring more space.

3.3.3 Design Constraints
Equipment must be easily accessible. Corridors should be free of obstruction with rounded corners. Floor surfaces are to be used to increase friction. Ceiling heights, stairs, and seating should be appropriate to gravity.

3.4 Safety
All design considerations should attempt to reduce the chances of body injuries. Specific safety measures include minimizing risks associated with acceleration, weightlessness, or microgravity, and radiation.

3.4.1 Design Constraints
Protection from radiation could be achieved through "layers" (the personal clothes, the skin of the habitat and a flexible additional protection during solar flares)

Equipment should be properly designed and properly used (perhaps with a reset button build in).

People can easily hurt themselves in uncommon lower gravity environment. Soft furnishing and floors and no hard edges are easy to install. Provide handholds and food restraints for friction.

Every crewmember should have easy access to all parts of the habitats and to the emergency vehicle or shelter. The safety system includes a search and rescue system, Emergency lighting, clear visibility of hazardous places, clear coding.

3.5 Stress
Environmental stressors are vibration, noise, air pollution, light vibrations. Noise can cause higher error rates and greater variability of performance. It impacts performance negatively by lowering motivation and moral. Noise over which a person has no control results in high level of stress.

3.5.1 Design Constraints
Minimizing the sources of Noise and vibrations throughout the habitat minimizes stress. (Work areas < 44 dBA, Crew quarters and sleeping areas < 35 dBA). In addition material that dampens noise is useful.

Increase control of crew members over sound producing machines (heating and cooling), noise in the crew quarters should be under control of individual crewmembers.

One the other side the presence of noise is necessary and beneficial. Russian experiments have shown that unfamiliar and unusual music had a positive effect on the astronauts.

Soothing music has been found to lead to helping behaviour.

Odour could also be used as a tool to create diverse atmospheres. Smell of forest, flowers in recreation areas

3.6 Community and Privacy

"Personal space refers to an area with invisible boundaries surrounding a person's body into which intruders may not come" (Robert Sommer, 1969)

Uncontrolled loss of privacy results in stress and social tension.

Visual access and visual exposure are the two key aspects of privacy regulation. A room to accommodate all crew members as well as disclosure of the individual private areas is important for the functioning of the group. To have the control over space and being an observer instead of being observed decreases privacy and territory problems.

Important is the relationship of the interior not the components.

As future space mission will involve a heterogenous group, private space for couples is demanded. This could be a flexible semi-private space. If needed crewmembers could expand their private space and share it with others.

3.6.1 Design Constraints
Some spaces (library - small group recreation room) should provide for small groups of 2 or 3 individuals and should screen them from visual distractions and provide auditory privacy.

Interior design of a spacecraft should have build in flexibility. Flexibility is enhanced by differing between "hard" architectural features (wall, hatches...) and "soft" features (screens, moveable partitions...).

Anything that would emphasize a person's individuality helps offset privacy loss. Decorating an area seems to have positive connotations; personally selected cloth could also reinforce individual separateness.

Plants can be added to provide positive amenities such as aromatics. Also a viewing port is crucial. It should be located in the lower 45° sectors.
3.7 Workstations
For ease of integration, maintenance, and change out, most fixed interior space station systems are incorporated into equipment racks and functional units. Racks contain: health maintenance equipment, food preparation appliances, storage units, life support, and waste management systems, environmental controls and experiment hardware. Functional Units: for sleeping, leisure activities, personal hygiene, showers and toilets.

The current direction of computer systems' development is toward greater flexibility and compatibility. Updating a system should be possible without changing the equipment.

To ensure that everyone gets on well with different and changing tasks, a clearly arranged interior design is necessary.

3.7.1 Design Constraints
It is very difficult to keep track of small tools in an environment where items can simply drift out of sight. Under conditions of microgravity, belts, pockets, snaps and acres are the order of the day.

A reset function should be built in for equipment that is shared by the group. Adequate work and circulation space is needed. The design should be standardized and uncomplicated to minimize potential for error.

3.8 Recreational Activities
Confinement begins with high motivation that decreases with time. The intention of engaging in creative activities (writing, reading...) mostly end with spending time on time-marking activities such as solitary.

The environment should provide an oasis or relaxation, pleasures and interaction and relief from the day to day activities.

Earth–Viewing is the most popular recreational activity offering visual attraction and psychological reassurance.

3.8.1 Design Constraints
Cognitive Strategies to enhance physical and social activities should be built in. Activities in space include hopping, sliding and climbing.

The sleeping habitats should provide privacy, undisturbed sleep and individuality. Individuality is reached by diversity in colors and texture, by lighting and odor and by personal objects such as plants. Each quarter should contain a sleeping area/ sleeping bag, storage facilities and a multimedia center. Each room should be surrounded by soundproof partitions.

4. CONCLUSION
We are now entering the stage where utopia - u-topos - no-place - becomes space to live. Architecture must always have the whole as its goal. Architecture is another bubble surrounding our body. In addition to the technical and functional abilities of this body, sensuality completes its. Our sensual perception - mainly visual, auditory and tactile - is influenced by form, color and light.

New perceptibility requires new ways of looking. This changes the context of space architecture.

Leaving the paradigm Earth and appreciating the new material and environment leads to a distinct and specific design of space architecture.

5. REFERENCES

Psychological Aspects for Living in Space - Architectural Challenges, Häuplik S., Lorenz S.


Connors Mary M, Living Aloft, Library of Congress Cataloging in Publication Data, (NASA SP,483)


Building for Space Travel, 2001 ISBN 086559-188-1

Inflatable Space structures, SICSA outreach Vol 1 No. 7, 1988

Space Architecture, Architectural Design Vol 70 No. 2, 2000

Spektrum der Wissenschaft, Raumfahrt No. 4, 1999

Piers Bizony, Die Internationale Raumstation, ISBN 3-405-15249-6


The Lunar and The Lunar Base Handbook, Peter Eckhart

Portable Architecture, Robert Kronenburg, Inflatable Lunar Habitat, Kriss Kennedy, NASA, p. 116-126

LUNAR OUTPOST, Alred, Bufkin, Kennedy, Roberts, Petro, Stecklein & Sturm, Systems Definition Branch,
Advanced Programs Office, NASA/Johnson Space Center [1989]

SICSA Outreach, Special Design Project Issue, Vol 2, No 4, 1989
SICSA Outreach, Special Design Project Issue, Vol 2, No 2, 1989
SICSA Outreach, Special Design Project Issue, Vol 2, No 3, 1989
Planetary Institute, Houston, ISBN 0-942862-02-3
http://spaceflight.nasa.gov/spacenews
http://spaceflight.nasa.gov/station
http://www.thespacestation.org
http://www.abo.fi
http://www.ssl.umd.edu
http://www.exn.ca
http://library.thinkquest.org/29033/begin/earthsunmoon.htm
Late Papers
TOOLS FOR RADIO-WAVE PROPAGATION SIMULATIONS IN PLANETARY SCIENCE

Alexandre Piot
SCI-SB, Research and Scientific Support Department (RSSD), ESTEC, 2200 AG Noordwijk, The Netherlands
E-mail: Alexandre.Piot@rssd.esa.int

ABSTRACT

As radar soundings become common in planetary science, new needs in simulating radio-wave propagation appear. This article takes the example of the radio experiment CONSERT on-board ROSETTA to introduce the Pseudo Spectral Time Domain (PSTD) algorithm which was first presented by Liu (1997). This algorithm aims at solving the Maxwell equations as the well known Finite Difference Time Domain (FDTD) but is less memory and CPU demanding to achieve accuracy computations.

Key words: Simulations of radio-wave propagation; Maxwell equations; Pseudo-Spectral Time Domain method.

1. INTRODUCTION

Radar sounding experiments are commonly used in planetary science. They range from ionospheric sounding to permafrost search on Mars (NETLANDER on MARS PREMIER, MARIS on MARS EXPRESS) via comet nucleus sounding (CONSERT on ROSETTA). To be developed, the inversion algorithms need some synthetic data which are provided by simulations of radio-wave propagation in the heterogeneous propagating media.

Most of these applications are low-frequency based. Simulations using geometrical optic approximations (wavelength small compared to the scale of the heterogeneities) like ray tracing code for example might be sufficient. However, some applications, like CONSERT, lie on the edge: needs appear to simulate radio-wave propagation in coarser comet nucleus models where the geometrical optic approximations are not valid anymore. Thus, a Maxwell equation solver, called PSTD algorithm, was developed at the Laboratoire de Planétologie de Grenoble during the Ph-D thesis of Piot (2002).

This article briefly presents the CONSERT experiment and its special requirements. Then, the PSTD algorithm is introduced and some of its properties presented.

2. THE COMET NUCLEUS SOUNDING EXPERIMENT BY RADIO-WAVE TRANSMISSION

CONSERT is part of ROSETTA, the ESA comet chaser to be launched in January 2003 (Schwehm and Schulz, 1999). It consists in two radio emitters: one on ROSETTA main spacecraft, the other on RoLAND, the ROSETTA Lander. The frequency is 90 MHz to which comet material is thought to be transparent. The full description of the experiment can be found in Kofman et al. (1998) and Barbin et al. (1999).

CONSERT aims at determining the electrical properties of the nucleus and revealing its internal structure. It is a major experiment as it will constrain the theories of the formation of comet nuclei which are closely related to the formation of the Solar System.

This experiment has never been run before, and so intensive research is conducted in order to prepare the data analysis. A complete tomography is believed to be possible if enough data are collected. See Barriot et al. (1999), Benna and Barriot (2001) and Benna et al. (2002).

3. MAXWELL EQUATION SOLVER TO SIMULATE PROPAGATION OF RADIO-WAVES

To solve Maxwell equations (2) in 2D (to start with: 3D will come later), two methods were considered: the Finite Difference Time Domain (FDTD) and the Pseudo Spectral Time Domain (PSTD) methods.

Both are leap frog algorithms. It means the electromagnetic field is computed from its value at previous


© European Space Agency • Provided by the NASA Astrophysics Data System
time step and the values of its space derivatives.

Both need Absorbing Boundary Conditions to simulate an unbounded medium. One of the most efficient available is the Perfectly Matched Layer (PML) from Bérenger (1994). These can dampen signals coming back into the computational domain as much as 150 dB.

\[
\begin{align*}
\frac{\partial E_x}{\partial t} &= \frac{1}{\epsilon} \left( \frac{\partial H_y}{\partial x} - \frac{\partial H_z}{\partial y} - J_x \right) \\
\frac{\partial H_z}{\partial t} &= -\frac{1}{\mu} \frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{1}{\mu} \frac{\partial E_z}{\partial x}
\end{align*}
\]

(1)

3.1. FDTD

The Finite Difference Time Domain (FDTD) algorithm relies on the central differential operator Eq. (2) to compute the time and space derivatives (Yee, 1966).

\[
\frac{\partial f}{\partial x}|_{x_0, x_f} = \frac{f(x_j+1) - f(x_i-1)}{2\Delta x}
\]

(2)

The FDTD algorithm is easy to implement but it suffers from anisotropic numerical dispersion. It means that the phase velocity in the "numerical" vacuum is not equal to that of the light in vacuum, and moreover it is not constant all over the computational domain (cf. Taflove (1995)). This is a major handicap in our application, as we will see in Sec. 4.2.

3.2. PSTD

The Pseudo Spectral Time Domain (PSTD) relies on the Fast Fourier Transform (FFT) to compute the space derivatives. Time is integrated using Finite Differences or Runge-Kutta like methods.

We used the well known property of the Discrete Fourier Transform concerning the derivatives. We consider a discrete function \( f(x) \) of \( N_x \) samples on the interval \([0; L_x]\). Let us note \( v \) a vector containing the value of \( f \) and \( v' \) a vector containing the derivatives of \( f \).

\[
v = \begin{pmatrix} f_0 \\ \vdots \\ f_{N_x-1} \end{pmatrix} \quad \text{and} \quad v' = \begin{pmatrix} \frac{\partial f}{\partial x}|_{x_0} \\ \vdots \\ \frac{\partial f}{\partial x}|_{x_{N_x-1}} \end{pmatrix}
\]

then

\[
v' = \mathcal{F}^{-1} \left\{ \begin{pmatrix} 0 \\ 1 \\ \vdots \\ N_x/2-1 \\ -N_x/2 \\ \vdots \\ -1 \end{pmatrix} \otimes \mathcal{F} \{v\} \right\}
\]

(3)

where \( \otimes \) is an element by element multiplication.

4. PERFORMANCES OF THE ALGORITHMS

4.1. Size, accuracy and dynamic requirements

CONSORT operates at 90 MHz. The nucleus of Wirtanen is thought to be 1200 m across (Lamy et al., 1998) and of relative permittivity of 2 (Kofman et al., 1998). Thus the expected wavelength (\( \lambda \)) is 2.4 m. The computational domain has then to be somehow 600 \( \lambda \) large.

4.2. Limitations of the FDTD algorithm

FDTD has then a major handicap: the numerical propagation is anisotropic and that leads to unbearable phase errors as it accumulates over the propagation distance. The only way to reduce this anisotropy is to reduce the grid lattice to as small as \( \lambda/20 \) or even smaller. But the computational domain is then to large to cop with, considering the computing means found at the Observatoire des Sciences de l'Univers Grenoble (OSUG) to which the LPG is linked.
4.3. Advantages of the PSTD Algorithm

PSTD allows to work with a $\lambda/2$ grid lattice, in theory, without any disturbing phase distortion. Thus, the PSTD algorithm is less memory demanding. In practice, we used a $\lambda/5$ grid lattice to enhance the signal to numerical noise ratio. This ratio is kept as low as (-120 dB) which is appreciated to simulate the full dynamic of the instrument (Piot, 2002).

Nevertheless, the PSTD algorithm still requires a computational domain containing 3072*3072 points and increments in time a few tens of thousands steps. Fortunately, it can be easily parallelized, thus reducing the execution time.

For example, to simulate a 25 $\mu$s experiment, 54,000 time steps are needed. It takes 6 days to complete, running 6 Power 3 processors on a shared memory IBM SP 9070 computer.

5. SIMULATIONS RUNNING THE PSTD ALGORITHM

5.1. Comet Nucleus Model

We run the PSTD algorithm on a smooth model, but it can deal with much coarser permittivity profiles. The relative permittivity here extends from 1.8 to 2.2.

The source is located at the star on the surface of the nucleus. The circle around the nucleus is called the orbit. It is not a realistic orbit but it is only to optimize the computations. It is useless to use the PSTD algorithm to simulate the propagation in vacuum. From knowing the electro-magnetic field on a circular orbit, it is possible to propagate the signal forwards using a technic similar to the retro-propagation described in Herique (1999).

5.2. Snapshots of radio-waves propagating in a comet nucleus model

These snapshots, Fig. 3 show the intensity in dB of the electro-magnetic field with respect to time. The nucleus surface and the orbit are also drawn. Time $t = 0$ refers to the emission of the maximum of the pulse.

5.3. Signal on the orbit

The source on the nucleus emits a modulated pulse which propagates inside the nucleus. The waves are then recorded on a circular orbit. This produces the patterns seen on the Fig. 4: intensity of signal function of time and position on the orbit. These patterns carry information about the internal structure of the nucleus. This output is fed to the inversion algorithm prototypes (Benna et al., 2002).

6. CONCLUSION

Among other tools to simulate the propagation of radio-waves, the PSTD algorithm is efficient to provide simulated data sets to develop inversion algorithm. It needs no assumptions, in theory, on the propagation media and is less memory and CPU time demanding than the FDTD algorithm, at least for large computational domains with respect to the wavelength.

However, it is still quite demanding in memory and then CPU because of the requirement in space sampling. Thus 3D applications are for the moment out of reach. But progresses in computers enable us to be optimistic in a near future. Full 3D simulations will surely run before ROSETTA arrives at Wirtanen in 2011.

REFERENCES


Figure 3. Propagation in a comet nucleus model.

Figure 4. Signal on orbit


LUNAR BASE DESIGN AND OPERATION STUDY I (LB-DAOS I)

Barbara Imhof1, Susmita Mohanty2, Hans Jurgen Rombaut3, Paul J. van Susante4, Jim Volp5

1 Architect, Institut fuer hochbau II, TU-Vienna, Karlsplatz 13/270 2, A-1040 Vienna and Principal, ESCAPE*spHERE, grosse morengasse 38/1, A-1020, Vienna, Austria
e-mail: imhof@hb2.tuwien.ac.at
2 Industrial Designer, Principal, Moonfront LLC, 550 Battery Street, Suite 804, CA94111, San Francisco, USA
3 Architect, Principal, Lunar Architecture, Utrecht, The Netherlands
4 Civil Engineer, Graduate Research Assistant, Colorado School of Mines, department of Engineering, 1500 Illinois street, Golden, CO 80401, Colorado, USA
5 Astrophysicist, Young Graduate Trainee, ESA/ESTEC, Keplerlaan 1, 2200 AG, Noordwijk, The Netherlands

This paper focuses on the actual case studies - scenarios of the Lunar Base Design And Operation Study I (LB-DAOS I) for further lunar exploration. The study started with the Lunar Base Design Workshop, a two-week programme conducted in cooperation with ESA-ESTEC, held from June 10th-21st 2002. Initially three scenarios considered as the most relevant for further Human Lunar exploration with a permanent Lunar outpost were investigated:

1. Scenario: Lunar Telescope Mission
3. Scenario: Lunar Ice Mission

These case studies developed into six different mission scenarios including ice mining, solar cell production, lunar telescopes, He3 mining, research and commercial operations. Each of these scenarios ranging from initial small scale scientific research outposts to commercial larger bases focussed on a timeframe between 2020 to 2090 to establish permanent habitats on the lunar surface.

APPROACH

The first approach was through student design teams which were multidisciplinary and comprised of about 50 students from 16 different countries, with a bachelor’s degree or higher from a variety of disciplines such as engineering, architecture, industrial design, mining, applied physics and medicine. The students were coached by the Workshop Managers and experts from ESA, NASA, space companies and universities during the two-week workshop. Amongst them were, Apollo-17 moonwalker - Harrison Schmit, ESA astronaut - Wubbo Ockels, NASA moon expert - Wendell Mendell, ESA’s SMART-1 Project Scientist - Bernard Foing.

ESA’s head of life sciences – Didier Schmitt and the chief of ESA’s Aurora Program - Franco Ongaro.

Workshop Program Overview

Week 1

- First 6 days
  - Lectures
  - Invited experts from ESA, NASA and other organizations
  - Remaining 3 days
  - Working Sessions
  - Students split into 5 teams
  - Introduced to Design Brief
  - Mission Case Studies
  - Research
  - Tour of ESTEC facilities

Week 2

- All 7 days
  - Working Sessions
  - Mission Case Studies
  - Research
  - Design Brief
  - Final Project Presentation

End of Week 1
- Design Critique
- By Expert Panel

End of Week 2
- Final Presentation
- To a Professional Jury

INTRODUCTION

The first lunar base designs date back from even before it was declared by President Kennedy that the USA were going to land a man on the moon and bring him safely back to Earth.

The ideas and designs that have been proposed are very versatile and diverse. Many issues need to be addressed in the designs such as life support, power generation, heat radiation, radiation shielding, landing and take-off and transportation. The design of space hardware is a very technical, relatively new area and the space design industry has been dominated since the beginning by engineers and scientists. In the designs for Lunar bases this is very visible. Many designs have a very high “tin can” level or just the opposite they are complete cities with clear domes and terra-formed regions below the domes. It is very difficult to find any designs of lunar bases that go beyond the first stage construction shack, tin can designs and into the realm of long duration durable expandable lunar bases. Some official studies
have been performed within NASA towards lunar base designs but due to the political influences it has sometimes even been forbidden to talk about plans for the moon. This resulted in highly fragmented work and loose studies during times when it was allowed to talk about the moon. All over the world individuals (at universities) and incidentally companies have pursued studies towards lunar bases.

Recently the interest in returning to the moon has increased worldwide. Europe is going to send its first small mission for advanced research and technology (SMART-1) to the moon, Japan is planning 2 missions and India and China both have short term plans to go to the Moon. At the moment Europe has the only official program for human planetary exploration called the aurora program.

The study will define what concepts will be applicable for missions described below. A preliminary conceptual approach will be set and key parameters defined for testing with the Concurrent Design Facility in ESTEC as a next step in 2003/2004. Aspects such as functionality, safety, psychology, and architectural design topics will be issues to be focused on. It will be done from a human-robot synergy viewpoint.

MAIN CONCEPTS – CASE STUDIES

Previous lunar base designs
Many designs for lunar bases have been proposed over the years. One of the remarkable things is that almost none of the studies relates to the human movement in 1/6G. This can be seen from the ceiling height, the design of vertical transport (usually earth type ladders) and the organization of the inside horizontal movement. This probably has a lot to do with the fact that the Apollo landers did not have enough room to move freely about and the only movement we have seen on the moon was the hopping of the Apollo astronauts in spacesuit on the lunar surface. This however is not a good reference to movement inside a lunar base in shirt sleeve environment.

Case study 1: Lunar Telescope Mission
It is the year 2020, a small permanent human outpost exists on the lunar surface near the south pole. Some elementary production capacity exists which can make simple forms and moderate sizes of construction material from lunar regolith. E.g. baked tiles of 50x50cm2 and aluminum rods with a length of 1 meter and diameter of 5 cm. there are more possible construction elements that can easily be manufactured from lunar regolith.
The telescope is to be placed at the bottom of a crater that is permanently shadowed and will be an infrared/visible light telescope with a light collecting area of at least 1000m2 either in one telescope or using more smaller ones connected to each other using interferometry.

"Define what you would be needed for a mission like this, design the lunar outpost that should support the construction of the telescope, give some ideas of what the machines should be able to do, make a comfortable but functional design for this lunar outpost and the infrastructure needed to support a project like this. Do this from a human-robot synergy viewpoint. So humans and robots working together, keep in mind the different needs of both parts of the synergy and the project. Make sure that your outpost can be expanded to accommodate scientists and astronomers that work on and with the telescope. Maybe it can be expanded to a full-size scientific base or maybe it can expand to a construction oriented base once the telescope project is built. Define criteria for each choice that you make and document your ideas and directions you took."

(LUNAR BASE DESIGN WORKSHOP, Scenario 1 – Brief)

The other two scenarios will be described as follows:

Scenario 2: Lunar Solar Power Factory Mission
This outpost is supposed to be extended to a full-scale factory for the production of solar cells and other required elements for the production of solar power stations in orbit around the Earth. Therefore the elements must be produced, tested on the moon and shipped to Earth Orbit or Lunar orbit where they will be assembled to start producing solar energy for beaming down to either the Earth or the Moon. For the launching of the elements an electro magnetic levitation accelerator (maglev) should be build. In this case the amounts of materials that can be send to space cheaply can be increased.

Scenario 3: Lunar Ice Mission
Scientific research to the composition and origin of the lunar ice that can be found in the permanently shadowed craters around the south pole. Lunar ice mining for the production of water and volatiles for human consumption and recycling and for the production of rocket propellant.

After the start of the workshop 6 groups were formed, each group chose a name and (or defined) a scenario. The groups and their chosen scenario are listed below:

Anaxagoras: Ice-Mining and Research
Kepler: Scientific Research in the areas of mining, self-sustainability and social and psychological aspects of living on the moon.
Tsolkovsky: Solar Power System elements production
Gagarin: Water usage research on the moon
Copernicus: Provide service to other lunar bases
Tycho: Helium-3 mining

Anaxagoras
Their chosen mission was Ice-Mining and Research. Since the possible presence of ice on the moon is limited to the lunar poles this group chose to build their base on the lunar south pole in the rim of Shackleton Crater. They plan to dig in the crater rim and support the excavated spaces with sintered regolith panels. Several alternatives for the base lay-out were made and the models were transformed into hands-on scale-models that were shown during the final presentation. One of the most important features of this study has been the study to movement inside a lunar base in a shirt sleeve environment. This appears to be significantly different from moving on Earth or on the lunar surface in a spacesuit. This is true for both horizontal and vertical movement. Since the human body structure is designed for 1G, movement in 1/6G requires less energy, but is less efficient if standard moving techniques are used.

Regular walking must be done very slow because the friction of the floor is less than we are used to and in the regular walking movement, a large part of the energy would be wasted by going in the air instead of forward. Because of this another way of "walking" was
proposed. For this a running pose would be assumed, leaning forward like a fast runner. However the position of the feet would not be correct anymore and thus it was proposed to make a “waving” floor so the feet would be aligned with the angle of this floor. Another option would be to put a stairs with large steps flat on the ground such that the steps would have an angle of 45 degrees with the normal floor. For vertical movement three modes of transportation were identified:
- regular movement
- “coffee-cup” movement
- heavy or large element transport

All three modes of transportation must be accommodated in the possible vertical transport designs, up as well as in downward direction. The solution came in the form of a vertical tube with 4 possibilities for moving. A spiral of small platforms with periodically a larger platform with on the opposite side a rails and a wide open middle area. The small platforms are for smooth (coffee-cup) transport using small controlled steps/jumps. The large periodically placed platforms are for passing but also for jumping up/down in regular movement. The rails on the opposite side is for smooth but fast rides down (like skating/gliding down). The opening in the middle is for transport of large elements or for just jumping down in case the height is not too high. Another important item in their design was the “contact” with the green in their base. The plants and recycling would be in a different room, but still visible by transparent panels. Also in their base design they suggested the use of colorful flowers and scented herbs. Especially in the recreation/relaxation room. A last statement of this group: “As we will never be able to survive on the surface and terraform it, we will have to move beneath the ground.”

Kepler
Their self defined mission was scientific research in the areas of mining, self-sustainability and social and psychological aspects of living on the moon. They came up with several metaphors for their activities: Beehive wise (modular structure and mobile activities performed from the nest., Kangaroo wise (hopping, mobile, but able to carry a lot).

PERSONAL CELL – ergonomically minimized private unit

The thoughts about modularity (e.g. one frame on which more modules can be fitted according to the need of the situation) stayed with them during the whole process.
Another example would be one big rover for long distance reconnaissance carrying small rovers for more local research. One of their lunar base ideas was the importance of visual contact with the green. In this case from their personal cabins. This would be in the form of a transparent piece in their cabins that would look out on/into the farm. The form chosen for the personal cabins allows a very effective way of stacking them on top or next to the farm.

Tsiolkovsky
This group chose the solar power system elements production facility. The site location was chose near the equator for reasons of communication and launches using a mass-driver. The transport issues were a very important part of this design. There are mining operations, production operations, storing and shipping activities. Next to that there of course also are living areas and such. A monorail system was chosen to provide the necessary transport capabilities between the separate elements of the base. An ingenious inflation sequence for the habitat areas was thought up. The areas where humans would have to be present would be connected with regolith covered inflatable tunnels. As final comment this group gave: "After two weeks we start to understand what is possible."

Gagarin
This team picked water usage research on the moon as their topic. At first all kinds of uses for the water were thought of. Some examples are: a submerged base, ice for shielding, ice for construction, etc. This teams metaphors consisted of among others: termites, continuous organic growth and the idea that with small uniform forms many different shaped and strong structures can be constructed. They chose an inflatable bubble as the smallest element for their base design. With this element all kinds of larger shapes can be constructed. The location of their base would be close to the production facilities of water, thus near the polar areas. Their base would be located in the top of a fairly flat crater rim. This means that some parts of the base would be dug in and some parts on the surface, but covered with regolith. It is spread out in horizontal as well as in vertical way. Another idea this group had was to make the surroundings around the base recognizable by placing lunar-made-regolith-tiles on the surface. They would have a slightly different color than the landscape, but not such that it would be totally off, just recognizable. This base would be very flexible because of its organic setup.

MonoRail
SubSystem design

Composition interface
Center component connection ring

After transport regolith from mining facilities to the production facility.

© LUNAR BASE DESIGN WORKSHOP TEAM

Kopernikus
This group decided to go more into the future and design a base that would provide service to other lunar bases. The services they decided that they wanted to provide were: Medical care, personal/cargo transport, repair/maintenance, spaceport, lounge facilities. Metaphors this group used were: chameleon, symbiosis, transformers. Their base idea would look like a spine on which inflatable units are attached. The services are provided using a number of vehicles that
have interchangeable modules for one or more of the required services. Those modules would be stored under the spine and fixed modules of the base itself. In case a large group of people would arrive or gather for a party a special section would be inflated for the duration of the event. All traffic from Earth would go through this location and from there be distributed to its final destination.

Tycho
Helium-3 mining was the topic that was chosen by Tycho. Their strategy was to be mobile. The whole base consists of 3 spheres that roll. The largest sphere contains the greenhouse and hydroponics, the medium contains the habitat and the smallest one contains the power generator which is a helium 3 fusion reactor that beams power to both other spheres. The mining itself is done by mining robots that collect the helium 3 and also other volatiles embedded in the regolith. The helium 3 is launched to Earth and the rest (like oxygen, hydrogen, nitrogen) is stored on the mining area. The mining areas are square and have a certain size. After mining these areas are prepared for further development and will have a supply of stored volatiles waiting for use. The rolling base will slowly move on and mine more areas while creating an infrastructure. Inside the habitat ball, the philosophy is that noise is grouped together and the rest is as shielded as possible. This applies to both activities as well as machinery. Another idea this group had was the way that personal space was shaped would show your privacy mood to the rest. E.g. if you have guests over for tea or so you can extend your floor space somewhat and if you want privacy you erect your tent and close it.

CONCLUSION

Novel ways for looking to a design for a lunar base have been proposed, mainly considering the permanently evolving lunar base concept and the interior aspects that so far often have been neglected. The interior movement is very important for the design of the base and how to put everything together such that it will be comfortable, efficient, safe and psychologically satisfactory is a challenge. These designs give direction in many ways and the next step will have to incorporate these issues. A very important tool that is still missing in the design of lunar bases is a systems model where all the systems are interconnected and interaction can be simulated. This tool could be used in ESA’s Concurrent Design Facility to design a lunar base interactively using a multitude of disciplines simultaneously. This would
pose a more engineering oriented approach but the model would be based upon the findings of previous studies, including the Lunar Base Design Workshop results.

ACKNOWLEDGMENTS
For this work we would like to acknowledge the assistance of European Space Agency / ESTEC for hosting the event, Colorado School of Mines, Lunar Architecture, Moonfron, ESCAPE*spHERE, TU-Vienna, Institute for Design and Building Construction, Austrian Space Agency, Umbilical Design, International Lunar Exploration Working Group and the Lunar Explorers Society.

Sponsoring organizations for the workshop included ESA-ESTEC, the Austrian Space Agency, and the International Lunar Exploration Working Group (ILEWG).

Last we would like to thank all participants for their contributions.

REFERENCES

Woodcock, G., Space transfer concepts and analysis for exploration missions, Boeing space & defense group, NASA report no. NAS8-37857, 1993
“3001: The Final Odyssey”, Arthur Charles Clarke, Ballantine Books (Mm.) March 1998
URL: http://www.lunararchitecture.com/lunarbase.html
Closing Remarks
CLOSING REMARKS ON ESLAB36 SYMPOSIUM ON
‘EARTH-LIKE PLANETS AND MOONS’

Bernard H. Foing* and the organisers of ESLAB36**

*Chief Scientist, Research & Scientific Support Department,
ESTEC / SCI-SR, P. O. Box 299, 2200 AG Noordwijk, NL, Europe
Email: bernard.foing@esa.int

ABSTRACT
We give a summary of ESLAB36 symposium on ‘Earth-like Planets and Moons’ that took place at ESTEC, Noordwijk on 3-7 June 2002. The different sessions included:
- A family portrait of Earth-like Planets and Moons
- The contribution of space missions for understanding Earth-like Planets and Moons
- Earth as a planet
- Methods for comparative planetology
- Interiors, surfaces, exospheres and impact processes
- Comparing atmospheres and fluids
- Earth-Like Planets and Moons in the galaxy
- Habitable Earth-like Planets and Moons
- ESLAB Symposium summary and roundtable discussion
- Robotic and Human exploration
- Young Planetary Explorers (YPE) special session

1. ESLAB36 ORGANISING COMMITTEES

**The ESLAB36 organisers included:

Scientific Organising Committee: B.H. Foing (chair), A. Chicarro, M. Fridlund, R. Grard, J.P. Lebreton (ESA RSSD)
+ A. Herland (ESA Earth Science Division), C. Barbieri (I),
+ T. Encrenaz (F), G. Neukum (D), J. Perez-Mercader (E),
+ T. Spohn, (D), J. Head, T. Johnson (US), A. Basilevsky, E.
+ Galimov (Russia), H. Mizutani (J)

Local Organising Committee: B.H. Foing, C. Bingham, C.
+ Nilsson, D. Heather, D. Koschny, P. Martin, G. Schwehm,
+ O. Witasse, B. Bartrick, R. LoVerde, M. Sanders

(see http://ssd.esa.int/resources/conferences/eslab36/)

2. HIGHLIGHTS FROM ESLAB36 SESSIONS

We summarise some key points from the ESLAB36 sessions:

2.1 Keynote Lecture

In his talk (also ESTEC Colloquium), ‘A family portrait of Earth-like Planets and Moons: similarities and differences’, J. Head discussed how Earth-like planets form, work and evolve. He compared the various physical and evolutionary parameters that shape each Earth-like planet in the Solar System. It is important to learn from these comparisons, also to understand the conditions on these planetary bodies relevant to the origin and evolution of life.

2.2 The contribution of space missions for understanding Earth-like Planets and Moons

(session chaired by J. Head)

E.A. Herland described ESA Missions to study planet Earth, including Earth-watch and Earth explorers.

B.H. Foing described the heritage from Apollo, Luna, Clementine and Lunar Prospector, and the perspectives with SMART-1, Lunar-A, Selene and future lunar missions.

D. Titov gave the rationale and history for missions to Venus, and the rationale for further studies.

A. Chicarro reviewed the results of recent missions to Mars, and the potential of Mars Express, to be launched in 2003.

R. Grard emphasised how Mercury is a key to understanding terrestrial planets, and how the US Messenger mission can help to prepare for ESA’s Bepi Colombo cornerstone, which will have global surface coverage, higher and uniform spatial resolution, long integration times and 50 times more telemetry capability.

J.P. Lebreton described ‘Missions to outer moons’ with emphasis on Galileo results on Jupiter’s moons, and
perspectives for the future exploration of Europa, and the upcoming studies of Titan with Cassini/Huygens.

M. Fridlund described ‘Missions to Earth-like exoplanets’, including indirect photometric detection methods that will be used by COROT, Eddington and Kepler. He presented the Darwin infrared Space Interferometer for the direct characterisation of Earth-like exoplanets and for investigating their capability to sustain life as we know it. Synergies between all these space missions were discussed in the area of remote-sensing, in-situ instruments, miniaturised technologies, and the challenge of large and complex global data sets from multi-missions.

2.3 Earth as a planet
(session chaired by E.A Herland and M. Rast).

H. Paulsen showed what is known about the Earth interior through seismology. She uses seismic ray path reconstruction, enabling velocity derivation leading to shear velocity models and different speeds in the upper mantle. Seismic tomography distinguishes higher velocity and lower velocity regions, which are correlated with surface tectonics. There is strong potential for collaboration here in the future not only with SAR (Synthetic Aperture Radar) interferometer, but also gravity missions such as GOCE (Gravity Field and Steady-State Ocean Circulation) and of course CHAMP (Changing Minisatellite Payload for Geophysical Research and applications) and GRACE (Gravity Recovery and Climate Experiment).

R. Wortel described deformation processes at various scales, in relation to Earth mantle structure and processes. He showed detailed evidence on the plate motions which are not in steady state, on the sea-floor spreading, on the subduction zone in the Adriatic (Mediterranean tear migration), and how sediment basins and their change may be used as indicators, with potential for remote sensing.

H. Laasko gave information on solar winds as a primary source of energy and momentum to the Earth plasma environment. He discussed solar proton episodes and Sun-Earth connection, the ionosphere's formation and layers, as well as the ionosphere-thermosphere coupling. He presented magnetospheric results from Cluster. A synergy with the future SWARM mission can be nicely constructed and be relevant to inter-disciplinary research (planetary/Earth sciences).

S. Bakan described in his talk ‘Global atmosphere and climate studies’, the energy balance in the atmosphere which shows anthropogenic effects. Scenarios of CO₂ emission predict temperature rises of 2-4 degrees, and global sea level rises up to 1 m. He described the relevance of these results for the International Pannel on Climate Change. Problems are in the description of water fluxes overland (evaporation), the importance of freshwater flux for thermo-haline circulation, and the relevance of the hydrological cycle for our climate.

H. van Leeuwen in his talk on ‘Ocean, ice and fluid envelope’, described difficulties to make accurate measurements in-situ. He described the oceanic circulation and energy transport, properties of seawater and interior dynamics (salinity, internal waves, Rossby vs. Kelvin waves, etc), Moon tidal relationships, tidal winds and thermo-haline forcing needed to explain present day ocean dynamics. Remote sensing can help a lot (Scat, SAR, SMOS, GOCE) and better inverse models are desperately needed.

M. Rast gave a talk on ENVISAT’s first highlights with an overview of its status and performance, illustrated with lots of spectacular data.

In summary for the session on Earth as a planet: Earth sciences can help us to further our understanding of planetary phenomena (interior and exterior dynamics); Remote sensing is one of the tools that can enable quantum leaps in gaining knowledge; Earth is probably a good testbed to develop and test sophisticated observation techniques; Planetary and Earth sciences have also in common the permanent battle for better understanding and thus improvement of our models.

2.4 Methods for comparative planetology
(session chaired by H. Waenke and A. Chicarro)

P. Martin’s talk on ‘Surface Mineralogy’, discussed the signature of the crust and mantle on the Moon and Mercury, the effects of weathering and alteration on minerals, and the comparison between basaltic and andesitic spectral features on Mars. He also addressed spectral features from icy outer moons.

S. Maurice reviewed ‘Methods for Remote Sensing of Elemental Composition’, with emphasis on neutron, X-ray, gamma-ray techniques, and recent results obtained with Lunar Prospector.

L. Colangeli’s talk on ‘Laboratory and simulation studies’ described the small-scale processing (such as activation, hydrogenation, radiation) and the new techniques used for study of extraterrestrial samples (IR microscopy, X-ray, Field Emission Scanning Electron Microscopes).

D. Koschny’s talk on ‘Science operations of ESA planetary missions’, and A. Marini’s companion SMART-1 paper, described comprehensive tasks and planning for planetary missions such as Rosetta, SMART-1 and Mars-Express.

A number of flash presentations of posters on related topics were given.
2.5 Interiors, surfaces, exospheres and impact processes  
(chaired by R. Grard)
H. Wanke described the 'Geochemistry of Mars from in-situ and meteorite analysis', showing the presence of 2 main components (A highly reduced, B fully oxidized) with different mixing ratios for the Earth (A=85%) and Mars (A=60%).
D. Breuer showed computer simulations for the modelling of the interiors and evolution of Earth-like Planets and Moons.
L. Wilson and J. Head discussed the 'History of volcanism on Earth-like Planets and Moons'. They showed the different styles of volcanism on Earth. They compared early Moon volcanism, Venus volcanic resurfacing over 100 Myr, and Mars extensive Hesperian volcanism and recent flows.
G. Cremonese's talk on 'Exospheres of Earth-like Planets and Moons' described the sodium and potassium exospheric detections for the Moon, Mercury and Europa, and discussed the sources and processes needed to sustain them.
M. Grande gave a tutorial on 'Magnetospheres in the Solar System', showing the different cases pending on the existence of an atmosphere, the magnetisation or the rotation properties of the body.
H. Hoffmann & G. Neukum, in their talk 'The Cratering Record on Planets & Moons in the Solar System' presented the new standard crater frequency distribution on the Moon, on Mars (including the effects of eolian resurfacing), and on outer bodies in the Solar System.

2.6 Impacts, interaction with Planets and Moons  
(session chaired by L. Colangeli, and G. Schwehm)
Talks by Ph. Claeys on 'The Earth impact cratering record' and A. Ocampo on 'Large impact cratering processes and their stratigraphy' discussed Earth crater impact structures and processes, and in particular the in-situ studies of Chixculub layers in relation to the Cretaceous-Tertiary KT event.
O. Botta's talk on 'Exogenous material delivery to Earth-like Planets and Moons' discussed the volatile and organics inventory contribution by asteroids, comets, meteorites and Interplanetary Dust Particles. Some of these components might have been important ingredients for prebiotic chemistry on the Early Earth and Mars.

2.7 Comparing atmospheres and fluids (with emphasis on Earth, Mars, Venus, Titan, Europa)  
(chaired by J.P. Lebreton, L. Becker, T. Encrenaz)
T. Encrenaz's talk on 'Atmospheric structure, composition, and diagnostics' reviewed atmospheric properties of terrestrial planets and outer satellites. Infrared spectra are used to derive elemental abundances. She discussed reservoirs of volatiles (chemical elements and D/H) in the Solar System nebula, and limits to rocky big Earths in the context of exoplanets. She discussed the case of Titan's special chemistry. She showed the specific case of the Earth's composition indicating departure from thermal equilibrium, and the rise of biogenic oxygen.
C. Sotin showed how, as a result of diverse sources of heat (accretion, radiogenic, tidal heating) large rocky planets can sustain an ocean.
C. d'Uston presented 'Preliminary results on the chemical composition of the Mars surface as observed by Mars Odyssey Gamma Ray Spectrometer' showing evidence of near-surface water.

2.8 Earth-Like Planets and Moons in the galaxy  
(session chaired by S. Volonté)
Some 100 exo-Jupiters detected with the velocity technique have been reported. The next steps for detection of Earth-size exoplanets will make use of high accuracy photometry of transits: this is the goal of space missions such as Corot, Eddington and Kepler. Visible coronography and IR nulling interferometry will follow, as complementary techniques to detect and characterize terrestrial atmospheres through spectroscopy of CO₂, water, CH₄, O₂ and O₃. The rise of terrestrial oxygen in the last billion years has led to life-searching strategies using O₂ or O₃ absorption bands. Hypertelescopes with baselines of 100 km, as proposed by A. Labeyrie and colleagues would allow us to resolve well the surface of terrestrial exoplanets.

2.9 Habitable Earth-like Planets and Moons  
(session chaired by B.H. Foing)
P. Ehrenfreund's talk on 'Complex organics and prebiotic chemistry in space and on planets' reviewed the abundances of ices and organics in the interstellar medium and in comets. She showed the ingredients necessary for life, exogenous delivery, and the earliest fingerprints of life on Earth through the ¹³C record and stromatolites. She argued whether life is a cosmic imperative: how "easy" is it to start it and what are the prospects for life elsewhere?
L. Becker's review on 'Impacts and mass extinctions' discussed the Cretaceous -Tertiary (65 Myr) and Permian - Triassic (250 Myr) boundary layers. She described the analysis of samples containing fullerenes as diagnostics of the large impactor bodies. Besides the direct impact and sustained winter effects, possible mass extinction factors may include poisoning with CO₂-rich water, and the possible association of impacts with massive floods of basalt. How does life re-invent and complexify itself after extinctions?

M. van Loosdrecht, in the talk 'Extremes of life on Earth' discussed the chemistry and energetics of life, using organics, water and redox reactions. Humans are extremophiles, they live in extreme environments and need defense against oxidising compounds. They resist to doses of 500 rads, but some species resist up to 1 billion more. With the large range of extremophiles on Earth, we need to think new strategies to search for life, and to develop sensors or biomarkers.

2.9 ESLAB Symposium summary & roundtable discussion

Y. Langevin, (chair of ESA Solar System Working Group), discussed 'Perspectives for collaborations and future missions'. He showed the comprehensive approach and synergies between ESA missions to Earth-like Planets and Moons. He insisted on the need for the Venus Express mission as a science opportunity for the planetary community during the long period between Mars-Express and Bepi Colombo.

In his ESLAB 36 closing remarks, D. Southwood, ESA Director of Science, repeated the need of cross-talk between Earth, planetary, astronomy and space sciences. He thanked the symposium organizers and participants for this interdisciplinary effort. ESLAB36 came at the right time, when the community is preparing for Mars-Express, SMART-1, Cassini-Huygens, Rosetta and Venus Express. He also called us to involve the youth and the public to participate in this adventure of knowledge and exploration. We should remember when we were kids, fascinated and inspired by the Apollo programme.

In the follow-up discussion, it was proposed, at the invitation of Prof L. Colangeli, to have another symposium Earth-like Planets and Moons II to be organized possibly at Naples in 2004. Other follow-up actions would be to promote a European network of comparative planetology allowing the exchange of scientists and post-docs, as well as summer schools for the training of post-graduate students in related topics.

2.10 Future robotic and Human exploration

(session chaired by B.H. Foing)

In his Keynote lecture 'Human exploration of the Moon and Mars', lunar astronaut Harrison 'Jack' Schmitt, described the adventure of the lunar exploration programme, and some highlights and documents from the Apollo 17 mission. For the future he insisted on the key assets of humans for exploration:

1) the brain, a reprogrammable and retrainable supercomputer; 2) the eyes, a wide and high definition optical system directly connected to the brain; 3) the fingers, a challenge in dexterity, especially if supported by advanced gloves; 4) the legs, giving necessary mobility.

Future robotics have to improve quantitative near real-time, long distance mobility, and tele-operations. Special developments are needed for light extravehicular suits.

2.11 Young Planetary Explorers (YPE) special session

This session contained a Discussion and Recommendations for ESLAB 36 follow-up, and "Flash Presentations" of Young Planetary Explorers posters. This covered topics ranging from Lunar and planetary science, public outreach projects, Psychological and Architectural challenges for living in space, and the inauguration of Lunar Base Design Workshop taking place at ESTEC on 10-21 June.

Lively Poster sessions (chaired by O. Witaske and D. Heather) during the whole ESLAB36 week allowed to present recent science results, specific space missions to Earth-like Planets and Moons, as well as space and ground instrumentation studies. On Wednesday 5 June, the participants could visit ESTEC Test facilities, as well as RSSD Laboratories and Planetary Operation Center.

3 Acknowledgements

We acknowledge the Scientific and Local Organising Committees, the session Chairs, the participants, the ESTEC services and Conference Bureau, that helped to make ESLAB36 a memorable event for science and exploration. We also extend our thanks to the colleagues that provided comments for this summary.