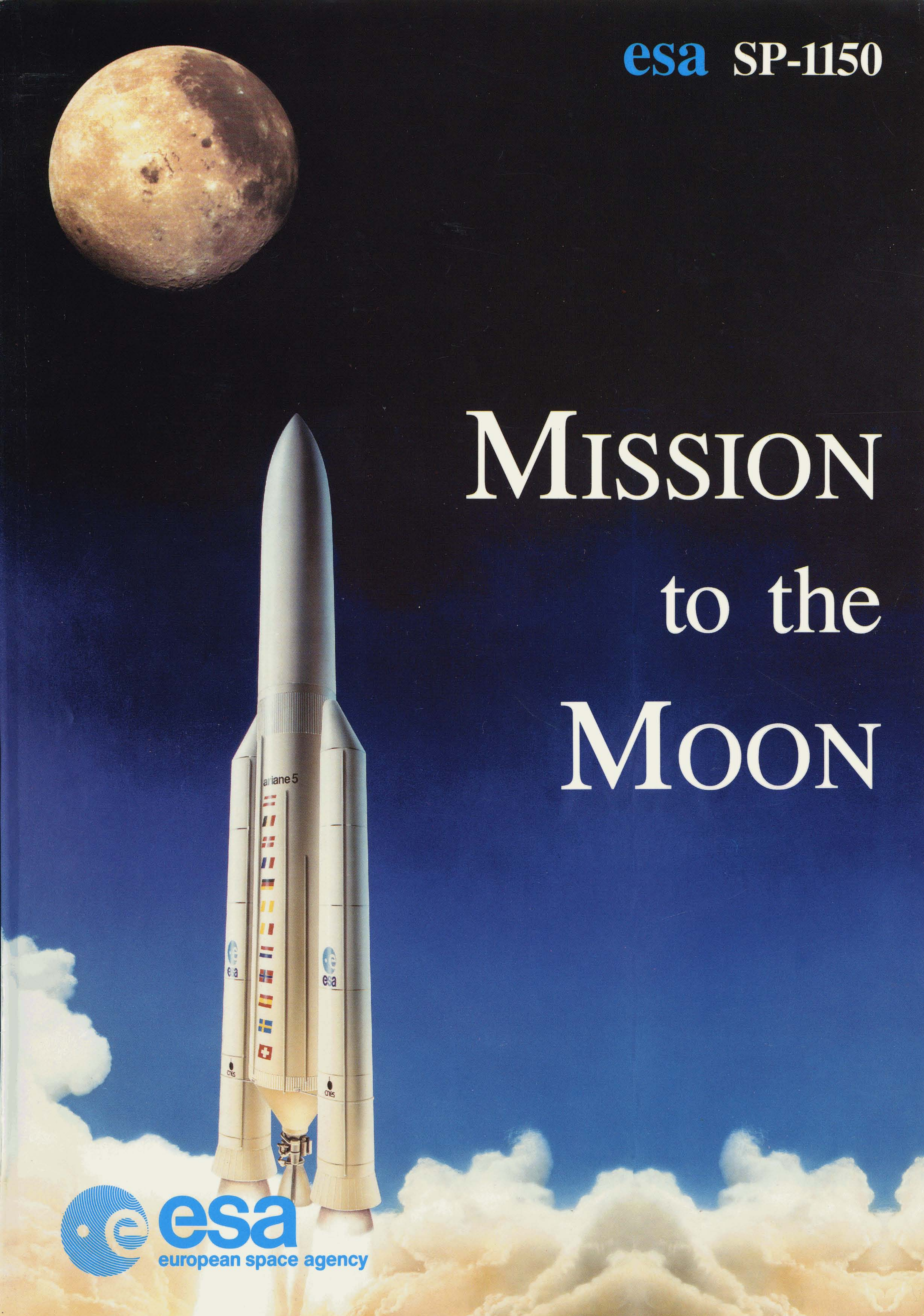
esa SP-1150

Mission to the Moon



esa
european space agency

esa SP-1150
June 1992

MISSION to the MOON

Europe's Priorities for the Scientific Exploration and Utilisation of the Moon

Report of the Lunar Study Steering Group

european space agency
agence spatiale europeenne

ILLUSTRATIONS

The use in the preparation of this report of photographic material furnished by CERN, ISAS, and NASA is gratefully acknowledged.

Published by:	ESA Publications Division, ESTEC, Noordwijk, The Netherlands
Editors:	B. Battrick & C. Barron
Layout:	C. Haakman
Cover:	P. Berkhout
Price:	50 Dutch Guilders
ISBN	92-9092-037-8

Contents

Preface

At a time when all of the World's major space programmes are being forced to re-evaluate their aims in the context of the quasi-elimination of competition between East and West, and with the increasing concern regarding the safeguarding of the Earth and its environment, this may seem a strange moment for ESA to embark on the study of a lunar base and to publish a document such as this.

Those who might question the opportuneness of this work should, however, be re-assured. The publication of this study does not reflect any change in the Agency's overall long-term goals, as recently reaffirmed by the Council Meeting at Ministerial Level in Munich in November 1991, nor does it reflect a change of priority in the 'Space Science: Horizon 2000' programmes. Rather, this study stemmed from the increasing interest being expressed in the worldwide scientific community in a return to the Moon which, in the context of the USA's Space Exploration Initiative, is also coupled with the manned exploration of Mars. Clearly, such ambitious goals can only be fulfilled through broad international cooperation, involving all of the World's major space agencies including ESA. It was therefore important for ESA to analyse how it could participate in this new and challenging venture, whenever it might materialise.

In 1990, ESA's Director General delegated responsibility for the coordination of the study to the Agency's Director of Scientific Programmes. This approach was taken in order to ensure that whatever ESA might contribute to any future lunar base, it would not be contrary to the interests of the scientific community and, on the contrary, would attract its support. The study reported here, which concentrated on the Moon only, with no connection to Mars exploration, was split into two phases.

In Phase-1, which is essentially the subject of this report, the scientific justifications had to be established. This task has been assumed by a steering group called the Lunar Study Steering Group (LSSG), the composition of which reflected all of the scientific disciplines that might benefit from the existence of a lunar base, such as studies of the Moon itself, astronomy, fusion, life sciences, etc.

The mandate of the LSSG, chaired by Professor Hans Balsiger, was to assess which scientific objectives can be uniquely fulfilled on the Moon, and not via other space- or ground-based means, and how they could be implemented. It was also asked to identify the potential role of Europe, in both science and technology, particularly in those mission areas where Europe has well-established and recognised expertise. Cost and schedule considerations were deliberately not pursued in this first phase, in order to encourage a brainstorming-type approach.

Phase-2 of the study will now focus on the broader policy issues and identify the technology requirements, the role that man might play, the means of transportation, and the infrastructure that should be considered. It will also look at the international-cooperation and legal aspects, and make recommendations on the approach to be adopted in order to preserve what we already know to be a unique environment in which to conduct scientific studies in the next century.

The LSSG has fulfilled its mandate, which was to complete its report, including its set of Recommendations (Chapter 7), within about a year and is to be richly complimented on the excellent quality of its work. The issuing of this report does not close the activities undertaken by ESA and the assistance of experts from the LSSG and its subgroups might well be needed in Phase-2, which should form the logical extension to the endeavours reported here.

Judging by the work accomplished by the Group and the substance of their Recommendations, we can be satisfied that the bottom-up approach that they have followed has produced a sound assessment of the consensus within the scientific community regarding the Moon's value as a scientific base.

R.M. Bonnet
Director of Scientific Programmes, ESA

Foreword

The increasing worldwide interest in the continuation of lunar exploration has prompted ESA to conduct an investigation of the motivations for returning to the Moon, establishing a permanent or semi-permanent base, and to consider the possible role that Europe could play in such an undertaking.

ESA's Director General delegated coordination of the study to the Agency's Directorate for Scientific Programmes. A *Lunar Study Steering Group (LSSG)* was therefore set up consisting of scientific experts representing a broad range of disciplines. This Group was charged with the task of formulating a set of Recommendations for ESA's possible scientific involvement and requested to report back to the Director General within approximately one year. In carrying out its task, the LSSG considered a number of key issues, such as the uniqueness of the lunar surface for scientific investigation/utilisation, the scientific areas of potential European involvement, and the impact of such an endeavour on the Moon's environment. The LSSG activity started in June 1990 and continued until August 1991.

As the first step in its work, the LSSG considered disciplines that would benefit significantly from the lunar environment. Due to the variety and complexity of the questions addressed, associated Study Teams were set up to consider more deeply those disciplines that appeared most promising. Much effort was devoted to the identification of the unique scientific objectives, the establishment of suitable strategies for European involvement, and the identification of the mission studies needed. A number of fields that appeared suitable only for later-stage activity were not addressed by specific teams, but were studied by the relevant members of the LSSG.

In the second step, the Study Team reports (reproduced in extenso as Chapters 1 to 6 of this document) were reviewed by the Steering Group, resulting in a set of priorities embodied in the Recommendations presented in Chapter 7, and reflected in the Executive Summary.

The members of the various Teams were as follows:

Lunar Study Steering Group

H. Balsiger *Chairman*	Physikalisches Institut, Universität Bern (CH)
J. Beckers	European Southern Observatory, Garching (D)
A. Carusi	Istituto di Astrofisica Spaziale, Roma (I)
J. Geiss	Physikalisches Institut, Universität Bern (CH)
G. Grieger	Max-Planck-Institut für Plasmaphysik, Garching (D)
G. Horneck	Institut für Flugmedizin/DLR, Köln (D)
P. Hoyng	Stichting Ruimteonderzoek Nederland, Utrecht (NL)
M. Jacob	CERN, Genève (CH)
Y. Langevin	Institut d'Astrophysique Spatiale, Orsay (F)
J. Lemaire	Institut d'Aéronomie Spatiale, Bruxelles (B)
J. Lequeux	Observatoire de Paris, Meudon (F)
P. Pinet	Observatoire de Midi-Pyrénées, Toulouse (F)
T. Spohn	Institut für Planetologie, Münster (D)
H. Wänke	Max-Planck-Institut für Chemie, Mainz (D)

The Team members from the ESA Directorate for Scientific Programmes, responsible for the study, were:

A. Chicarro	Study Scientist, ESA Space Science Dept., ESTEC, Noordwijk (NL)
M. Fridlund	Study Scientist, ESA Space Science Dept., ESTEC, Noordwijk (NL)
S. Volonte	Executive Secretary, ESA Head Office (D/Sci), Paris (F)

Technical support was also provided by personnel from ESA's Technical, Operations, Space Station and Microgravity, Space Transportation Systems, and Scientific Programmes Directorates.

Science of the Moon Study Team

A. Carusi	Istituto di Astrofisica Spaziale, Roma (I)
J. Geiss	Physikalisches Institut, Universität Bern (CH)
Y. Langevin *Chairman*	Institut d'Astrophysique Spatiale, Orsay (F)
J. Lemaire	Institut d'Aéronomie Spatiale, Bruxelles (B)
P. Masson	Université de Paris-Sud, Orsay (F)
P. Pinet	Obervatoire de Midi-Pyrénées, Toulouse (F)
T. Spohn	Institut für Planetologie, Münster (D)
H. Wänke	Max-Planck-Institut für Chemie, Mainz (D)
A. Chicarro	ESA Space Science Dept., ESTEC, Noordwijk (NL)

Lunar Interferometry Study Team

J. Beckers	European Southern Observatory, Garching (D)
J.C. Dainty *Chairman*	Imperial College, London (UK)
M. Faucherre	European Southern Observatory, Garching (D)
A.H. Greenaway	Royal Signals & Radar Estab., Malvern (UK)
A. Harris	Max-Planck-Institut für Astrophysik, Garching (D)
A. Labeyrie	Observatoire de la Côte d'Azur, Grasse (F)
J. Lequeux	Observatoire de Paris, Meudon (F)
J.E. Noordam	Netherlands Foundation for Radioastronomy, Dwingeloo (NL)
G.P. Weigelt	Max-Planck-Institut für Radioastronomie, Bonn (D)
M. Fridlund	ESA Space Science Dept., ESTEC, Noordwijk (NL)
A. Hawkyard	ESTEC, Noordwijk (NL)
S. Volonte	ESA Head Office (D/Sci), Paris (F)

The above Team acknowledges the contributions of:

G. Gahm, Stockholm; M.A.C. Perryman, ESTEC; P. Roussel, ESTEC;
J. Surdej, Liège; J.P. Swings, Liège; M.H. Ulrich, ESO; F. Verbunt, Utrecht; and
E.J. Wampler, ESO.

Very-Low-Frequency Astronomy Study Team

R.S. Booth *Chairman*	Space Observatory, Onsala (S)
J.L. Bougeret	Observatoire de Paris, Meudon (F)
A. Hewish	University of Cambridge (UK)
A. Lecacheux	Observatoire de Paris, Meudon (F)
J.L. Steinberg	Observatoire de Paris, Meudon (F)
M. Fridlund	ESA Space Science Dept., ESTEC, Noordwijk (NL)
S. Volonte	ESA Head Office (D/Sci), Paris (F)

Astrometry Study Team

P. Bernacca	Osservatorio Astrofisico, Asiago (I)
E. Hög	University of Copenhagen Observatory (DK)
J. Kovalevsky	Observatoire de la Côte d'Azur, Grasse (F)
R. Le Poole	Sterrewacht Leiden (NL)
M. Fridlund	ESA Space Science Dept., ESTEC, Noordwijk (NL)
M. Perryman *Chairman*	ESA Space Science Dept., ESTEC, Noordwijk (NL)

Planetary Astronomy

A. Carusi	Istituto di Astrofisica Spaziale, Roma (I)

Solar-Physics Study Team

J.P. Delaboudiniere	Institut d'Astrophysique Spatiale, Verrières-le-Buisson (F)
P. Hoyng *Chairman*	Stichting Ruimteonderzoek Nederland, Utrecht (NL)
S. Solanki	ETH, Zürich (CH)
O. von der Lühe	European Southern Observatory, Garching (D)
M. Fridlund	ESA Space Science Dept., ESTEC, Noordwijk (NL)

Solar-Wind Plasma

J. Lemaire	Institut d'Aéronomie Spatiale, Bruxelles (B)

High-Energy Physics

M. Jacob	CERN, Genève (CH)

M. Jacob acknowledges communications with A. Abashian, G. Cocconi, G. Domokos, S. Kovesi-Domokos, G. Garvey, Ph. Goret, P. Grieder, F. Halzen, C. Jarlskog, H. Meyer, L. van Hove, G. Yodh and T.T. Wu.

Life-Sciences Study Team

V. Blüm	Ruhr-Universität Bochum (D)
F. Bonde-Petersen	Aeron. & Space Medical Centre, Copenhagen (DK)
A. Brack	Centre Biophysique Moléculaire/CNRS, Orleans (F)
H. Bücker	Institut für Flugmedizin/DLR, Köln (D)
G. Horneck *Chairwoman*	Institut für Flugmedizin/DLR, Köln (D)
J. Kiefer	Justus-Liebig-Universität, Gießen (D)
C. Lindberg	Institut für Flugmedizin/DLR, Koln (D)
H. Wänke	Max-Planck-Institut für Chemie, Mainz (D)
A. Chicarro	ESA Space Science Dept., ESTEC, Noordwijk (NL)
H. Oser	ESA Head Office (D/SSM), Paris (F)

This Team acknowledges the contributions of:

M. André (St. Paul-lez-Durance), K. Dose (Mainz), A.J. Heidmann (Meudon), K.H. Kreuzberg (Köln), F. Raulin (Créteil).

Helium-3 Fusion

G. Grieger	Max-Planck-Institut für Plasmaphysik, Garching (D)

Exospheric Contamination

J. Lemaire	Institut d'Aéronomie Spatiale, Bruxelles (B)

Lunar Archive

J. Beckers	European Southern Observatory, Garching (D)

Executive Summary

Executive Summary

The Moon, our closest planetary neighbour, is the best known object in the Solar System, with the obvious exception of the Earth. Our current knowledge is essentially derived from past exploration, both automated and human, carried out within the Apollo and Lunar programmes of the seventies. These have already demonstrated that the Moon is within reach by exploiting our present capabilities. However successful these programmes were, major questions still remain unanswered concerning the origin and characteristics of our 'satellite'. Renewed interest in the Moon has developed in recent years, resulting in studies, mainly performed in the USA, to define scenarios leading eventually to the establishment of a manned lunar base.

The aim of the present study was to determine Europe's potential role in the future exploration and utilisation of the Moon. In this first phase, the Lunar Study Steering Group (LSSG) concentrated mainly on scientific questions, leaving technological issues such as transportation, the role of humans, infrastructure and policy matters to a later phase. It only partially considered questions relating to the exploitation of lunar resources and the impact of human activities on science. The two main questions addressed in this first phase were:
(i) Are there scientific investigations that can only (or best) be performed on the Moon?
(ii) Can Europe/ESA play a key role in some of them?

Both questions received a positive answer.

For the purposes of this study, the scientific issues were divided into three main areas:
- *Science of the Moon*, including all investigations concerning the Moon as a planetary body;
- *Science from the Moon*, using the Moon as a platform and therefore including observatories in the broadest sense;
- *Science on the Moon*, including not only questions relating to human activities in space, but also the development of artificial ecosystems beyond the Earth.

Science of the Moon focusses on geophysical, geochemical and geological observations from a variety of technical elements, leading to a better understanding of the origin of the Earth–Moon system. The Moon also provides a unique record of the history of the interplanetary medium. In this context, the LSSG proposes the following sequence of events for a renewed lunar exploration programme:
- A lunar polar orbiter, to provide complete geophysical, geochemical and geological mapping, and full high-resolution coverage of the lunar surface.
- Surface stations and rovers to determine the internal structure of the Moon, the chemical and mineralogical composition of selected sites, and to perform traverses across geological boundaries.
- Sample-return missions to investigate further the most relevant sites on the Moon, such as the highlands and the far side, of which our knowledge is very limited.
- A lunar outpost/base to provide support to field geologists in sampling and in-situ observations of the lunar surface, and to allow the refurbishment of surface stations and rovers.

Obviously, lunar exploration must be conducted on the Moon itself, but other disciplines will also benefit from studies performed on the lunar surface. This is generally true for observatories in the broadest sense (from traditional astronomy to high-energy cosmic-ray detection). The LSSG has therefore tried to identify those activities that can be carried out more effectively from the Moon rather than from free-flyers or from Earth. A good preliminary assessment has already emerged.

Science from the Moon takes advantage of the stable lunar ground, its atmosphere-free sky and, on the far side, its radio-quiet environment. The Moon provides an attractive platform for the observation and study of the Universe. Two techniques that can make unique use of the lunar platform are ultraviolet to sub-millimetre interferometric imaging, and very-low-frequency astronomy.

The improvement by order(s) of magnitude in angular resolution and sensitivity at all wavelengths allowed by very-long-baseline lunar interferometers will have a tremendous impact on imaging and astrometry, and will lead to scientific breakthroughs in galactic and extra-galactic astrophysics, as well as in planetary astronomy and the physics of the Sun. However, it is apparent at this stage that separate systems will be required for the UV to IR and the sub-millimetre domains, as well as for solar physics.

A possible strategy allowing the deployment of a lunar-based interferometer emerges from a variety of considerations: technological maturity, scientific impact, realistic time scales and programme flexibility in relation to developments both on the ground and in space. A logical course would be:
- to undertake an in-depth design study of an intermediate baseline (~ 100 m) space-based optical interferometer providing milli-arcsecond resolution;
- to initiate conceptual studies for very-long-baseline (> 100 m) lunar interferometers to cover the UV to IR and the sub-millimetre domains.

For very-low-frequency astronomy, the ability to deploy large arrays on the far side of the Moon represents the only means of avoiding the interference effects caused by electromagnetic radiation from the Earth. The installation of large structures on the far side will require a mature lunar base and infrastructure, and is therefore not feasible during the early stages of lunar exploration. An alternative option is to deploy simple antennas at surface stations remotely, to gain early access to this unexplored domain of the electromagnetic spectrum.

For high-energy physics, the Moon appears to be an interesting location for the study of very energetic cosmic-ray-related events. Since the Earth's atmosphere acts as a convenient converter, this research can be expected to remain Earth-based for the foreseeable future. Only with the advent of a well-established, mature lunar base will it become advantageous to deploy future-generation specialised detectors, which will open up new capabilities.

The lunar surface is very stable because seismic activity is extremely low, plutonic processes ceased to play a role three billion years ago, and topographic changes due to impact cratering are slow. Apart, therefore, from its suitability for present-day *Science from the Moon*, the lunar surface offers a unique opportunity to investigate the history of radiations such as the solar wind, solar-flare particles, galactic cosmic rays, and solid particle fluxes. As a multitude of investigations conducted within the framework of the Apollo Programme has shown, the lunar record covers the last three to four billion years. Whereas the scientific aim of studying the histories of these radiations is in the category of *Science from the Moon*, the experimental methods and measurements required are linked to the *Science of the Moon*: most of the studies need returned samples, intelligently collected, making full use of the Apollo experience. In order to optimise the needs both of the *Science of the Moon* and the investigations of radiations that permeated the Solar System in the past, the requirements and techniques for the collection of lunar samples need to be studied carefully and in great depth.

One of the goals of life-sciences studies (*Science on the Moon*) is obviously to provide the prerequisite information for establishing a manned lunar base. This includes studies of human physiology under reduced gravity, radiation protection and life-support systems, and feasibility studies based on existing hardware. These results will be of direct interest in the context of human exploration of Mars and beyond. The studies that should be performed on the Moon rather than on the Space Station should be identified during the next phase by a life-sciences study team. The second goal of the life-sciences studies is the establishment – for the first time – of an artificial ecosystem on a celestial body beyond the Earth. This could begin with a simple, remotely controlled system to be built up as the lunar base is developed.

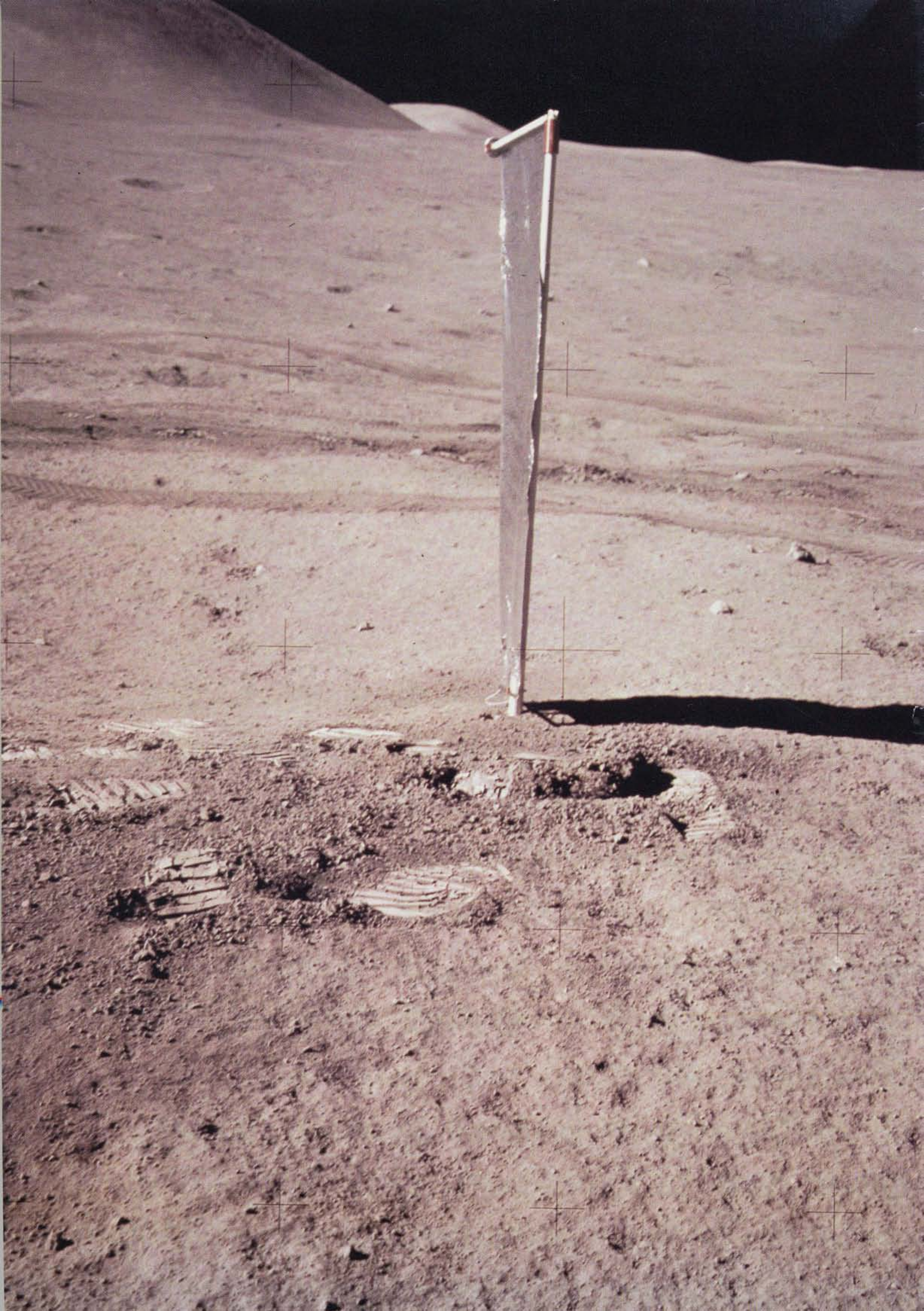
The infrastructure required for man's permanent presence on the Moon will make use of available resources to support the exploration activities. This implies that exploitation of lunar resources will eventually have to be considered to provide self-sufficiency at the lunar base, and in principle the raw materials are available for building such an infrastructure. As far as fuel for fusion energy is concerned, the important question of mining large amounts of lunar soil to extract the available helium-3 (^3He) can await a later stage of lunar exploration. Continuous improvement of the database on the ^3He resources on the Moon's surface and on their distribution is expected as a byproduct of other lunar exploration programmes. This type of fusion technique is unlikely to become available for quite a number of decades.

A question that needs to be kept in mind, and studied further, is whether human activity, and in particular mining, will disturb the lunar environment in such a way that the unique scientific activities on the Moon are badly compromised. This problem must be taken into account when policy issues are discussed at a later stage.

The first phase of the study has already confirmed that there is not only significant interest in Europe in all lunar-science disciplines, but that there is also significant expertise already available in many fields. In the past, this community has shown itself able to use analytical data to derive important conclusions about the composition and structure of the Moon and the time-scale of formation of the lunar land-forms, as well as about the origin of the Moon itself. On the one hand, we have a community of researchers with a long and successful tradition of analysing meteorites and lunar samples with excellent laboratory equipment. On the other, Europe can compete with anyone when it comes to building space experiments or ultra-modern telescopes. The necessary technological capabilities have already been demonstrated by past scientific missions undertaken by ESA and the European national agencies. Depending on the financial resources devoted to the exploration of the Moon in Europe, and on the time frame in which these could become available, a European contribution to lunar exploration could take the form of moderate elements (e.g. a lunar polar orbiter, surface stations, a precursor interferometer) early in the sequence of missions, or be part of a larger undertaking later (e.g. sample-return missions, fully-fledged interferometers, a lunar base).

The overall recommendations of the Lunar Study Steering Group are summarised in Chapter 7 of this report. They are essentially to set up specific study teams for those fields that have been judged to be the most promising for Europe, with the aim of providing more detailed scientific and technological specifications. It is also suggested that the scope of the overall study activities be expanded in order to derive mission scenarios for a viable ESA lunar exploration programme, and to consider economic, legal and policy matters. Finally, the need for international coordination early in the study phase is emphasised.

1. Science of the Moon



Chapter 1. Science of the Moon

1.1 Introduction

A new lunar exploration programme, with the prospect of establishing a permanent lunar base, represents an outstanding opportunity for scientific study. Indeed, while the Moon, after the Apollo and Luna Programmes is now the best known planetary object except for the Earth itself, important questions have still to be answered.

Prominent among such questions is the origin of the Earth–Moon system. The extensive lunar study programme of the 1960s and early 1970s has led to a good understanding of lunar history, stretching as far back as a few hundred million years after the Moon's formation. It is now known that the Moon and the Earth were formed within the same few million years. However, the observed orbital evolution and the physical, chemical and mineralogical properties of our satellite seemed to rule out each of the three major groups of models proposed for its origin: capture, accretion in Earth orbit, and fission from a rapidly rotating Earth. A few years ago, 'mixed' scenarios such as the hypothesis concerning a giant impact raising mantle material into orbit, provided a ray of hope, but much is still unknown. Acquiring new knowledge about the origin of the Earth–Moon system is considered to be a unifying theme for renewed efforts in lunar science.

In the field of geophysics, much is yet to be learned about the internal structure of the Moon. An upper limit to the size of a possible lunar core has already been estimated, but its existence remains to be proven and its size is yet to be measured. Similarly, the thickness of the crust has only been measured directly on the central near-side, where the Apollo seismometers were installed. The thickness of the crust and the size of the core are essential parameters for assessing the global chemical composition of the Moon, particularly its strongly depleted iron content compared with that of Earth. A related issue, paleomagnetism, involves the question of whether the Moon has generated a dipolar magnetic field. This question is difficult to answer from the restricted geographical distribution of available lunar samples. Three steps are required for these objectives:

(i) a global mapping of altimetric, gravitational, magnetic, electric and thermal properties from a low lunar orbit
(ii) ground stations equipped with broad-band seismometers and heat-flow probes with a wide geographical distribution (far side, polar regions), so as to observe as many seismic events as possible from distant sites
(iii) a representative set of sample-returns for deciphering the paleomagnetic record.

The broad chronological framework for the geological history of our satellite has been established from isotopic studies on lunar samples. Over a few hundred million years, a crust developed, mainly constituted by calcium-poor pyroxenes and anorthosite, a calcium-rich feldspar. Current scenarios suggest that over the next 500 million years, heat from long-lived radio-nuclides accumulated in the interior, while giant basins were formed by impacts. The formation of mare resulted from the filling up of these basins, mostly on the near side, by basaltic flows, followed by a tectonic phase due to settling. The origin of this lunar asymmetry is still a mystery. Over the last three billion years, the evolution of the surface has been dominated by impact cratering. Detailed investigation of the tectonic and geological history of the Moon is hampered by the limited coverage and performances of the remote-sensing techniques of the seventies. Only 14% of the surface was observed, in the equatorial zone. The main goals for geology are therefore to extend the remote-sensing coverage to the whole lunar surface, and to carry out traverses in selected areas, first with automatic rovers, and then by field geologists based on the lunar surface.

The bulk chemistry of the Moon represents a major challenge for formation modelling. The information needed is dependent on geophysical studies (mass of a possible iron-rich core, relative volume of mantle and crust). Constraints on the composition of the crust and of the mantle can be derived from a comprehensive study of the composition of highland material and mare basalts. These units exhibit significant variation, preventing the extrapolation the limited set of remote-sensing and in-situ information to the whole Moon. In particular, the concentration of titanium and highly incompatible elements such as P, K, REE, Hf, Th and U, vary greatly in lunar-surface units. For these objectives, global mapping of the chemical and mineralogical composition is again required, together with a representative set of in-situ analyses combined with sample-returns, which will provide the composition in terms of minor and trace elements, as well as a very valuable 'ground truth' for interpreting remote-sensing data.

The interaction of the Moon with its environment (meteoroids, solar wind, cosmic rays) has dominated its surface evolution over the last three billion years. The result of this interaction is the ubiquitous lunar regolith, a layer of dust and rocky fragments 5 to 10 m thick. The formation of a regolith is expected for many Solar System bodies, including Mercury, small rocky satellites and asteroids. This layer, when considered as a resource, contains significant proportions of all but a few key elements, such as hydrogen and carbon, which represent the major challenge for the autonomous operation of a lunar base. The lunar regolith represents a unique record of the evolution of the interplanetary medium over several billion years. In this case, the best strategy is to drill cores extending to the base of the regolith (hence, as far back in time as possible) in accumulation regions such as foothills, where the depositional history is relatively simple. Such a programme would benefit strongly from a human presence, as demonstrated during the Apollo years.

Based on this summary of the scientific objectives, a balanced lunar exploration programme should include the following four steps:

- A lunar polar orbiter, providing complete geophysical, geochemical and geological mapping, and full high-resolution coverage of the lunar surface with up-to-date remote-sensing techniques.
- In-situ studies at selected representative sites: ground stations, including seismic observatories, and rovers performing traverses across geological boundaries, will determine the local chemical and mineralogical composition, as well as providing site documentation. Studies conducted by ESA for the Marsnet Phase-A study are likely to yield relevant technical information for eventual in-situ lunar investigations.
- Sample-returns, from selected representative sites; the highlands, the far side in general, and high-latitude regions are not (or poorly) represented in the present set of nine sampling sites (six Apollo, three Luna). Deep drill cores extending to the base of the regolith and beyond would also be of high scientific value.
- A lunar base would allow field geologists to conduct investigations that would be extremely valuable for selecting and documenting new lunar samples. A lunar base would also be of help in placing and maintaining surface stations and rovers.

A lunar polar orbiter is considered a mandatory precursor mission for any high-level lunar exploration programme. Such a programme should also include the implementation of surface modules such as surface stations and rovers. The duration of these programmes could be extended at a later stage, exploiting the refurbishment possibilities provided by the presence of a lunar base. A sample-return programme would benefit strongly from the presence of geologists on the Moon, as good site selection and documentation are essential for making the most of sample investigations in the laboratory.

The European scientific community can play a major role in almost every aspect of a lunar study programme, given its acknowledged excellence in three key areas:
- Geosciences: this community is very strong in Europe, and is becoming increasingly involved in space activities, both for studying the Earth itself (ERS-1, Aristoteles) and for planetary exploration.
- Remote sensing and in-situ analysis techniques: Europe's participation in planetary missions received a major boost with ESA's highly successful Giotto mission and the recent decision to participate as a major partner in NASA's Cassini mission to Titan and Saturn. European expertise now extends to all major remote-sensing instruments, and has developed to include in-situ studies in the framework of the Huygens probe for Cassini, and the surface modules for the Russian Mars-94 mission.
- Sample analysis: this community has developed via the exploitation of the rich European meteorite collections and, more recently, the lunar sample-return programme. It is actively involved in the preparation of future sample-return missions, such as Rosetta (comet-nucleus sample return), as well as American and Russian plans for Mars sample returns.

1.2 Origin of the Earth – Moon system

1.2.1 Scientific background

The Earth, with the Moon as its only satellite, is one of only two examples of binary systems orbiting the Sun. The other is the Pluto – Charon pair. The Earth – Moon system, however, is unique in several respects:
- The major component is the largest of the terrestrial planets, and the minor component is the fifth largest satellite in the Solar System. In comparison, the dimensions of Pluto lie between those of 1 Ceres and Triton.
- The Earth – Moon system is located very close to the Sun, in a region where a strong depletion of gases and volatiles has taken place during accretion.
- Geochemical data suggest a genetic relationship between the Earth and the Moon.
- The dynamics of the system are unique: the Moon's orbit is substantially inclined with respect to both the Earth's equatorial plane and its ecliptic plane. Furthermore, the system is just within the limits of dynamic stability.
- The Moon is easy to reach from the Earth and hence can be extensively studied.

■ Theories of formation

The various theories about the origin of the Earth – Moon system, although very different in many respects, agree basically on one point: that the two bodies originated from the same bulk material and hence in the same region of the Solar System (Wood 1986). Four scenarios have been advanced so far:

- *Capture of the Moon by the Earth.* The Moon could have been a 'planet' formed elsewhere in the Solar System and became gravitationally trapped in the Earth's orbit (Singer 1986).
- *Fission of the Moon* from a parent body, whose remnant is the present Earth. The fission process could have resulted from spontaneous emission of proto-Earth upper-mantle material, which then condensed in orbit around the proto-Earth, or from the fission of a molten, viscous, rapidly rotating proto-Earth (Boss & Peale 1986).
- *Co-accretion of both bodies* from the same material, with important differentiation processes that led to the present difference in the internal compositions of the Earth and Moon (Weidenschilling et al. 1986)
- *Impact-triggered formation* of a circumterrestrial swarm, from which the Moon accreted. A variant of this hypothesis considers the disruption of a passing body within the Roche limit of the Earth, and accretion of the Moon from debris (Cameron 1986; Hartmann 1986).

The capture hypothesis seems basically implausible, because a purely gravitational permanent-capture event is very unlikely. Since a direct capture, via only a gravitational interaction, is almost impossible, dissipative processes have been put forward in order to transform a temporary capture (already difficult per se) into a definitive capture. They range from tidal dissipation to multiple encounters, to sudden variations of the masses of the Sun and the Earth, and to collisions with existing satellites. None of these mechanisms, however, seems convincing.

The classical fission hypothesis needs to explain the high energy and angular momentum required to form the swarm of ejected material. Rotational fission, in turn, often compared with the mechanism of formation of double star systems, does not seem to be applicable to a solid or highly viscous body which would not reach dynamic instability. A rapidly rotating molten and inviscid body, on the other hand, would produce an excess angular momentum difficult to remove.

Co-accretion, although very successful in many respects, needs complicated and concurrent mechanisms – whose probabilities and efficiencies have not yet been satisfactorily assessed – to account for the differentiation processes at an early stage of the bodies' formation. Moreover, the hypothesis is unable to explain convincingly the origin of the large angular momentum possessed by the pair of bodies.

The last hypothesis, which is also the most recent, considers the impact of a large body (500–3000 km) on the Earth, followed by the ejection of a large cloud of upper-mantle material. In a sense, therefore, it provides the fission hypothesis with the necessary energy and angular-momentum inputs. However, the presence of large impactors at the end of the Earth's accretion has not yet been proved. Indirect evidence of such an event (mainly the obliquity of the Earth's polar axis) suggests an impact of a body the size of Mars. This poses delicate problems for the theoreticians who are confronted with the formation of such a large body in the same 'feeding zone' and at the same time as the Earth. It has been argued that, after the impact, most material would fall back on the Earth, leaving little chance of forming a circumterrestrial, proto-lunar disc. A solution to this problem would require that a major portion of the ejecta be vaporised; gas pressure on this material would accelerate it, so that its motion would not be strictly ballistic. The collision hypothesis has been proposed to account for other anomalies in the Solar System, such as the rotation rate of Venus and the orientation of Uranus' axis.

1.2.2 Scientific objectives

The formation of the Earth–Moon system remains essentially an open question. However, although all the suggested theories have their advocates, there are strong indications that most of them are impossible, or very difficult to accept. We need to understand much more about the internal structure and composition of the Moon, its heat flow, the geological processes that have shaped its surface, the dynamic history of its orbit, and the relative abundances of its constituents, from both the mineralogical and elemental points of view.

Moreover, recent investigations into the orbital evolution of the Moon (and hence the roational speed of the Earth in the past) seem to indicate that this may have taken place discontinuously, with periods of fast evolution separated by phases of rather marked stability. This may have strong implications for the physical conditions at the time of accretion (constraints on the orbital inclination).

It should also be emphasised that the presence of such a large satellite as the Moon (with respect to the Earth) and its orbital evolution (the Moon used to be much closer to the Earth) have played an important role (e.g. tidal interactions) in the evolution of the Earth itself and, quite possibly, in the evolution of life.

In conclusion, there is still much to learn about the processes that have led to the formation of the Earth–Moon system. The Apollo and Luna Programmes provided a wealth of data which placed severe constraints on the previously proposed formation theories; indeed, none of these is consistent with all of the available information. A satisfactory theory on the origin of the Earth–Moon system requires an interdisciplinary effort to provide the essential information still missing, such as the internal structure and the global chemical composition of the Moon. The relevant fields of investigation are discussed in more detail in the following sections. A comprehensive strategy for lunar scientific investigation is then proposed.

1.3 Geophysics of the Moon

1.3.1 Scientific background

As discussed in the previous section, probably the most interesting scientific question about the Moon is its origin (and subsequent evolution). Although there appears to be a general consensus among experts that the Moon is closely related chemically to the Earth's mantle, and must in some way have been formed from the Earth, this hypothesis has certainly not been proven. The currently most widely discussed scenario advances the theory that the Moon formed as a consequence of a giant impact on the Earth (Hartmann et al. 1986).

A significant body of geophysical evidence relating to the structure and composition of the lunar interior was acquired during the Apollo missions and included seismic, gravitational and topographic data, heat-flow measurements, dynamic parameters and paleo-magnetic data. These data provide important information concerning the chemical arguments for a genetic relationship between the Moon and the Earth's mantle. Unfortunately, as discussed below, the geophysical database is still incomplete, and some of the most fundamental questions about the Moon are still unanswered. Relatively recent reviews on the subject have been given by the authors of the Basaltic Volcanism Study Project (BVSP 1981) and by Hood (1986).

The most valuable geophysical data for determining the interior structure of a planet are seismic and gravity data. Seismic velocities can be inverted to provide core size, crust thickness and the thicknesses of mantle layers. To a certain extent, seismic velocities are also sensitive to chemical reaction. The Moon is the only body apart from the Earth for which a significant seismic database is available. The low signal-to-noise ratio of lunar seismic signals, however, requires a long-term database to determine the velocity structure of the interior accurately. Unfortunately, the Apollo seismic experiment was terminated in 1977. In addition, the selenographic distances between the seismometers were short and the array covered only the near side of the Moon, where most of the seismic sources were also located. Consequently, only the seismic signal from a single far-side impact event that may have passed through a core was recorded. The poor areal distribution and the restricted operation time of the array have severely limited the accuracy of the data. Gravity data, together with topography data, can be used primarily to map the thickness of the crust by interpolating between sparse seismic measurements. Coverage of the Moon by precise gravity measurement is presently confined to some areas on the near side. Dynamic properties such as libration parameters, magnetic fields, surface heat flow, electrical conductivity as a function of depth, the chronology of igneous activity, and the nature of tectonic surface features are planetary properties that further constrain the interior structure and composition.

Figure 1.1 shows an equatorial section of the Moon and helps to illustrate our present knowledge of the interior structure, based on the Apollo geophysical data set (Goins 1978). The Moon has a crust, the thickness of which varies between about 30 km beneath the mascons to about 110 km beneath the highlands, with an average thickness of about 70 km (Bills & Ferrari 1977). These crust thicknesses were derived from gravity and topography spherical-harmonic models and from near-side seismic data. The quoted values of crust thicknesses are dependent to some extent on the assumed and probable Airy isostatic compensation mechanism (Haines & Metzger 1979). Airy isostasy assumes a laterally constant density. Variations in gravity are then produced by variations in crust thickness. This interpretation of the gravity data is not unique, however, since the same gravity signal can, in principle, be produced by a laterally varying crust thickness and by a laterally varying crust density.

There appears to be a crust thickness dichotomy, the far-side crust being an average of about 40 km thicker than the near-side crust (Bills & Ferrari 1977). This dichotomy is also the simplest explanation for observed centre-of-mass to centre-of-figure offset of 2.0 to 2.5 km (Kaule et al. 1972 & 1974). The crustal-thickness dichotomy is consistent with the observable dichotomy in the distribution of the mare and the highlands. The mean crustal density is about 3 g/cm^3. The crust may itself be differentiated and two- and three-layer models have been proposed with an anorthositic gabbro layer on top, underlain by a basalt layer and a mafic anorthosite layer (Ryder & Wood 1977).

The known structure of the lunar mantle is constrained by seismic data to a depth of about 1000 km. Beneath this, the structure of the Moon, as the authors of the BVSP put it, 'remains a mystery'. The depth limitation stems mainly from the small number of seismic stations and their areal distribution, from the finite available data set, and from the poor signal-to-noise ratio.

According to the seismic velocity data of Nakamura (1983), the mantle may be divided into upper and lower mantles separated by a low-velocity transition zone. These data have been inverted by Hood & Jones (1985) using laboratory elastic-constant data to provide density versus depth profiles. According to these models, the upper-mantle density is about 3.4 g/cm^3 and the velocity data are consistent with an Mg/(Mg+Fe) ratio in the range 0.75−0.80, originally derived from petrologic inferences. In the lower mantle and in the transition zone, the density is much less well constrained. Taken at face value, the inversions by Hood & Jones (1985) suggest a high-density transition zone and a low-density lower mantle, a structure that is unlikely because the lower mantle would have been gravitationally unstable to convective overturn during the early history of the Moon. The upper mantle and the transition zone are about 200 km thick.

The seismic data contain only very limited evidence for the existence of a metallic, iron-rich core. A single seismic event on the far side was recorded by the seismometers. Its p-wave travel times were interpreted as suggesting a high-absorption, possibly metallic, core of 170—360 km radius (Nakamura et al. 1974). The event was weak and the signal-to-noise ratio small, however, and the evidence remains inconclusive. The moment-of-inertia factor I/MR^2, the accepted value of which is 0.391, suggests a concentration of mass towards the centre of the Moon. Models that satisfy the mass, moment-of-inertia factor, and cosmo-chemical

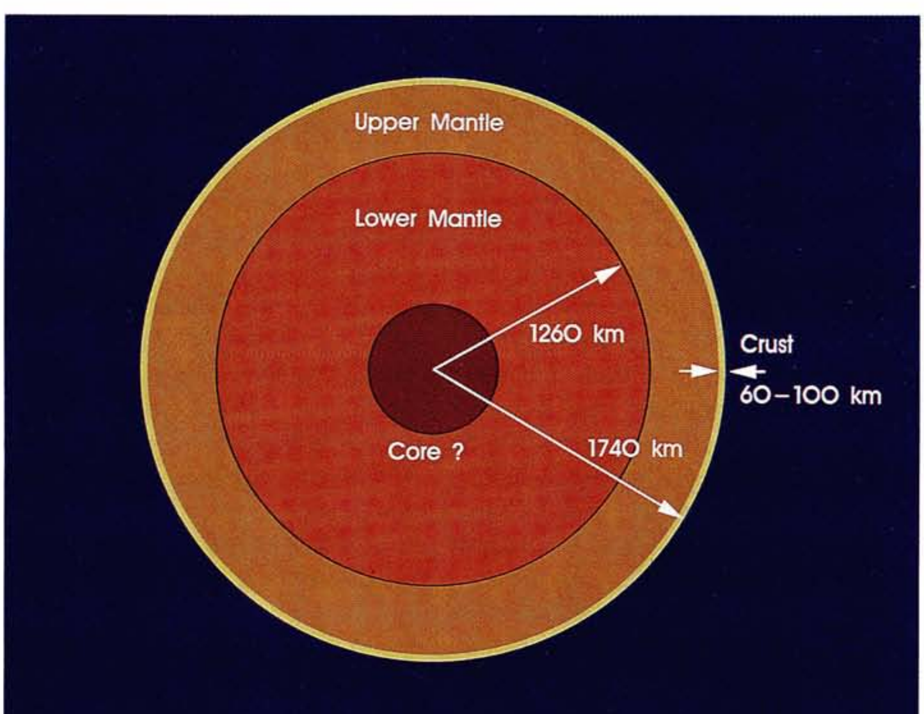


Figure 1.1 Interior structure of the Moon, based on Apollo results

constraints feature core radii of 100 to 500 km (BVSP 1981; Hood 1986). However, the mass and moment-of-inertia factor constraints can also be satisfied by models with density increasing from the upper to the lower mantle. The latter models are somewhat difficult to reconcile with the most recent seismic-velocity profiles for the lower mantle described above. Independent, though still not conclusive, evidence for the existence of a metallic core was derived from electromagnetic sounding data, including that of the measured negative lunar-induced magnetic-dipole moment in the geomagnetic tail, from the measurement of physical libration parameters to laser-ranging data (Yoder 1981), and from paleo-magnetic data that suggest the existence of a core due to the existence of a paleo-magnetic field. The laser-ranging data, which show a 0.2 s advance of the lunar spin axis from the Cassini alignment, even suggest that the present core, or an outer core layer, is still liquid.

1.3.2 Scientific objectives

Some of the chemical arguments for and against the proposed genetic relationship between the Earth's mantle and the Moon can be tested with better means than the Apollo data. The refractory element enrichment derived from the sparse heat-flow data and the Mg/(Mg+Fe) ratio have both been cited as basic tests for the hypothesis that the Moon originated primarily from material of the Earth's mantle (Kaula 1977; Taylor 1982). Among the most important data are the size of the core and the thickness of the crust. Core size is directly related to the Mg/(Mg+Fe) ratio, and it has been proposed that a geophysical determination of the lunar core size may help to distinguish between models of lunar origin. If the Moon was formed entirely from terrestrial mantle material, which was already depleted in siderophiles relative to chondrites, then a relatively small core would be sufficient to explain the lunar siderophile abundances. On the other hand, if the Moon was formed from undifferentiated solar-nebula material, then a larger core (~ 730 km radius) would be required. This criterion assumes complete core differentiation, which may not have occurred. It is also possible that metal was added to the proto-Moon by a terrestrial impactor. An accurate determination of core size may therefore not settle the issue entirely. However, the accurate determination of the size of the lunar core is among the most fundamental geophysical tasks yet to be accomplished.

A second fundamental question addressable by means of geophysical exploration is the thickness of the lunar crust and the depth to which the Moon melted in the past. A small planetary body such as the Moon cannot be geologically active for a long time unless there is some unusual heat source such as tidal heating. The ages of surface features on the Moon demonstrate that geodynamic activity ceased a long time ago. If the Moon has a crust as thick as 30–35 km beneath the mascons and 90–110 km beneath the lunar highlands, then there must have been an outstanding early thermal event leading to almost total melting of the Moon. Although models by Schubert et al. (1990) suggest that it may be possible to completely differentiate a planetary body by pressure-release melting during early vigorous mantle convection in a few 100 million years, these models have been criticised by Spohn (1990) for neglecting the almost inevitable crustal recycling and for being incompatible with the apparent long period of volcanic activity on Mars.

1.3.3 Strategy

Seismic data can be recorded and transmitted to Earth or a lunar base via a network of automatic 'Lunar Science Surface Stations'. The stations should form a seismic array, or even several arrays, and should be arranged such that an accurate determination of the core is possible. It is of particular importance to place some of the stations on the far side to allow the recording of signals that pass through the core, and in turn allow an accurate determination of the thickness of the far-side crust. Almost all seismic sources placed during the Apollo mission were on the near side. The stations should be equipped with advanced high-gain seismometers and could also be equipped with other relevant scientific instruments of interest. Gravity and topography should be measured from a low-altitude satellite such as the proposed 'Lunar Observer'. This satellite could also record radio brightness-temperature data as a means of determining surface heat-flow with extensive global coverage, but some classical in-situ heat-flow measurements are probably required to calibrate the satellite data. These measurements could perhaps be carried out automatically by the surface stations prior to the emplacement of the seismometers or by astronauts during lunar field work. Electromagnetic sounding could be carried out with magnetometers on the satellite and on the surface stations. Paleomagnetic data as well as chronological data require the acquisition of samples. This could probably best be accomplished by manned field excursions from a lunar base. For seleno-geophysics, a lunar base can be used as a relay station for data transmission and as a place where some processing and pre-processing of data could be carried out. The station could also serve as a base for lunar fieldwork. Further aspects of lunar-base science have been discussed by Mendell (1985).

1.4 Geology of the Moon

1.4.1 Scientific background

A considerable amount of knowledge concerning lunar geology has been gained from both in-situ rock sampling and Apollo fieldwork, and remote-sensing observations either from lunar orbit (e.g. γ-ray, X-ray ALSEP experiments) or Earth-based (radar and visible near-infrared spectroscopic reflectance data). The processing and analysis of various types of data has resulted in the idea that the lunar crust might be much more complex in its origin, evolution and composition than the earlier view derived from simple magma ocean models suggests.

Basically, the lunar surface comprises an anorthositic highland crust and basaltic maria. In principle, planetary crusts may form in two ways, either as a consequence of early melting and differentiation, or by derivation from the planetary mantles by partial melting, during the remainder of the evolution of the planet. When the lunar samples were examined, the composition of the highland crust was found to be highly differentiated, while mare basalts resulted from partial melting inside the Moon's mantle. The highland crust forms the oldest accessible area on the Moon and is saturated with giant multi-ring impact basins (diameter >300 km) and large craters ($50-100$ km diameter). Consequently, the formation of the lunar highland crust has been strongly affected by the heavy bombardment of the Solar System.

Lunar-relative chronology is based on the number of craters per surface unit, impact-ejecta stratigraphy and crater morphology (Young 1977; Soderblom & Lebofsky 1972; Head 1975). The absolute time scale was established by radiometric measurements based on U-Pb, Sm-Nd, Rb-Sr and Ar-Ar isotopic ratios, applied to

Table 1.1. Periods of lunar geological history

Stratigraphic Age	Time (billion years)
Copernician Era	0—1
Eratosthenian Era	1—2.5
Imbrian Era	2.5—3.8
Nectarian Era	3.8—4.3
Pre-Nectarian Era	4.3—4.55

the lunar samples (Young 1977; Tera et al. 1974). Lunar geologic history is accordingly divided into five periods, listed in Table 1.1 with their corresponding time spans.

The surface that has been sampled is presumed to date from 4.2 – 4.3 billion years ago, the earliest time at which the growth of a crust solidifying from a possible existing magma ocean could overcome destruction by intense bombardment. The estimated thickness of the crust currently ranges between 50 and 100 km. Following closely, and indeed overlapping, the terminal (3.9 billion years) earlier stages of the intense cratering, the episode of inundation of the impact basin depressions by lava flooding occurred for at least 800 million years and created the lunar mare on the near side by successive filling stages. Presumably, some of the earliest filling stages might subsequently have been obliterated by impact-basin ejecta.

Two prominent features of the mare are their tendency toward a circular form, caused by the flooding of the large ringed basins, and their spreading on the same equipotential surface of the lunar gravity field. Based on an inversion of gravity and topographic data from the near side of the Moon, models for the structure of the crust and upper mantle beneath impact basins provide a maximum thickness estimate of 5 km for the mare basalt fills (Bratt et al. 1985). The Moon is believed to be a one-plate planet, and this belief requires that the lunar tectonism developed in a predominantly vertical sense and was mainly controlled by the depression and loading of the lithosphere caused by the impact basin's formation and subsequent evolution (Solomon & Head 1980).

1.4.2 Scientific objectives

■ Highland crust

Although the composition of the uppermost portion of the highland crust (1 – 2 km of mega-regolith) is dominated by one mineralogical rock type (noritic composition, the major mineral being low-calcium pyroxene), less than a quarter of the remotely-sensed areas that display deep-seated material are noritic in composition. The dominant mineral assemblages of stratigraphically deeper (5 – 10 km) crustal materials are of gabbroic, anorthositic, noritic and troctolitic composition (Fig. 1.2). While noritic compositions occur across the entire lunar near-side crust, with no apparent clustering associated with any of the major basins, the spatial distribution of gabbroic compositions shows a concentration in the western hemisphere, i.e. within the region of the Oceanus Procellarum impact basin. Based on the limited data available, there is thus some evidence that the lunar crust, at least on the near side, shows both large-scale lateral heterogeneities in composition and a pronounced vertical petrological stratification (Stöffler et al. 1980; Pieters 1986; Lucey & Hawke 1987). This remark in the planetary geodynamic context of the Moon, a one-plate planet, leads to the idea that the Moon is a unique 'laboratory' for acquiring an understanding of the fabric of a planetary crust and its formation processes (vertical differentation, partial melting, magmatic intrusions, etc.).

In fact, the Moon, given its small size, cooled relatively rapidly, and endogenic surface activity related to magma generation ceased about one billion years after accretion. In addition, the absence of horizontal motions (no organic belt formation)

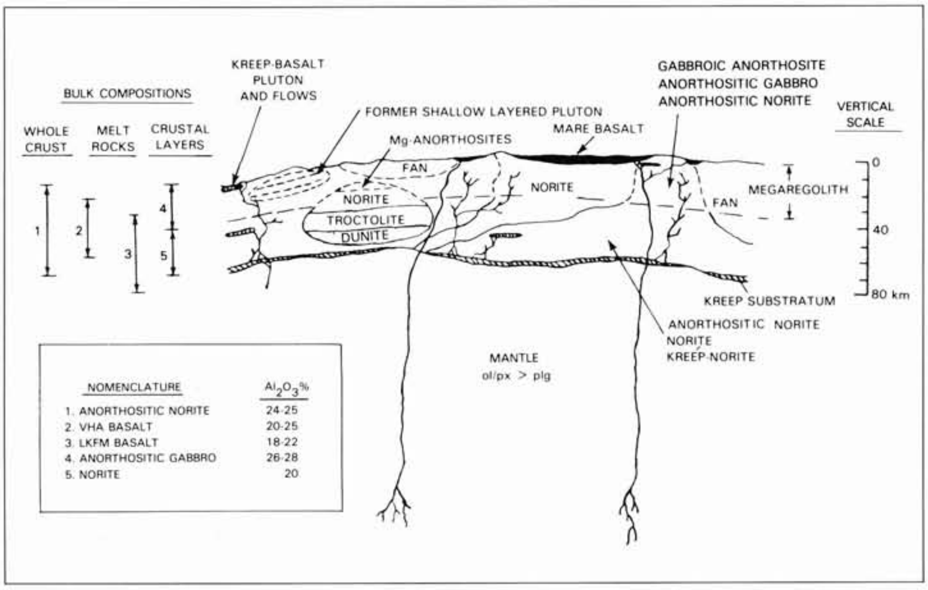


Figure 1.2 Schematic cross-section of the lunar crust

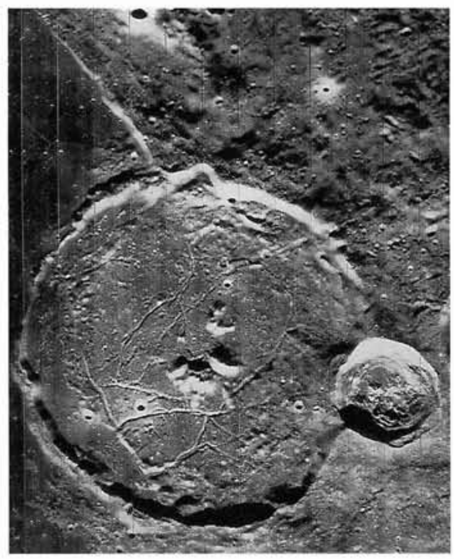


Figure 1.3 Apollo-16 oblique view of Gassendi crater

and of surface-transport processes such as erosion (with the exception of impact ejecta) should have protected the crust from major structural modifications and from recycling within the mantle. In this context, the use of large impact craters and multi-ring basins as natural vertical probes allows access to the petrological stratigraphy of the crust, ultimately down to the mantle (e.g. within Oceanus Procellarum), from mainly structural and compositional surface observations. In particular, valuable knowledge could be gained and constraints put on the lower crust and mantle chemical properties from the systematic study of the old floor-fractured craters (Nectarian or pre-Nectarian time) located at the edge of large impact basins where mega-concentric scarps occurred, and of the morphologically identified volcanic structures (Aristarchus, Marius Hills, etc.). Good examples of craters are Gassendi (Fig. 1.3) located at the edge of Mare Humorum, Piccolomini at the edge of Mare Nectaris, and Petavius at the edge of Mare Fecunditatis. Both central peaks and likely magmatic intrusions within the floor fractures (Fig. 1.4) would be compositionally very interesting.

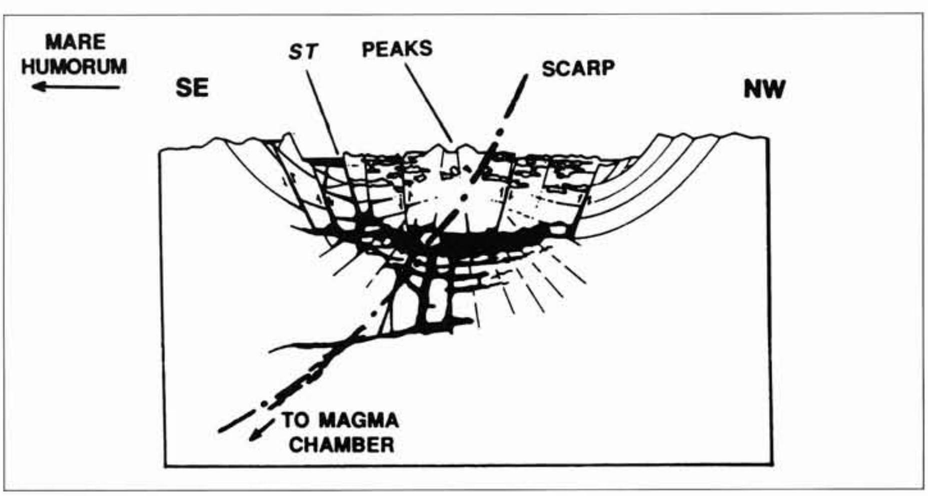


Figure 1.4 Southeast – northwest cross-section of Gassendi crater

■ **Impact basins and mare basalt emplacement**
Basalts are produced by partial melting within planets. Hence, lunar mare basalts provide records of the geochemical and thermal processes that have affected the Moon's mantle through time. In addition to the characterisation of the surface mineralogy and composition information, a significant objective would be to obtain a good description of the geometry, stratigraphy and time emplacement of the successive mare basalt infilling layers, separated by variable regolith thicknesses within the major impact basins in relation to the basin tectonic patterns (multi-rings, concentric mega-scarps, mare ridges), subsidence and downwarping effects. During the Apollo Programme, only a limited investigation was carried out in the case of southern Mare Serenitatis.

1.4.3 Strategy

A comprehensive strategy for studying lunar geology would combine geochemical, seismic and gravimetric observations:

- Geochemical remote-sensing techniques (γ-ray and X-ray spectrometry, visible/infrared imaging spectroscopy) will provide global mapping of the chemical and mineralogical surface compositions. The detection of volatiles in the polar regions would be of significance for establishing a lunar base.
- Seismic observations (via a network of penetrators or semi-hard-landing surface stations) will be required to determine the lunar crust's vertical structure
- Gravimetric techniques (gradiometry or satellite-to-satellite tracking) from a low-altitude orbit (~ 60 km) will allow the detection of local intra-crustal mass anomalies.

Such a strategy would result in a detailed description of the characteristics and history of the lunar crust. Altimetric and subsurface radar-sounding techniques would be very useful for complementing our understanding of mare basalt emplacements and multi-ring impact-basin tectonics.

At a second stage, in-situ studies should document the geological units that were not sampled during the Apollo and Luna programmes, which were restricted to the equatorial regions of the lunar near side. The most important scientific goals would be determination of the chemical composition and the absolute ages by radio-chronology. These investigations could be implemented by deploying surface stations at selected sites. It would also be important to obtain geological and geochemical profiles across boundaries between major units, using a 'mobile laboratory' such as a rover.

Global surveying of the lunar surface by employing powerful remote-sensing techniques, proposed as a first step in a lunar geology programme, will enable potential sites for in-situ studies to be selected. In the longer term, it is also essential for preparing for lunar-base activities. Lunar geology studies would benefit significantly from the implementation of a lunar base from which field geologists could carry out more detailed investigations (traverses, sample selection and documentation, deployment and maintenance of scientific equipment).

1.5 Geochemistry of the Moon

1.5.1 Scientific background

■ **Surface chemistry**

It has long been known that the surface of the Moon consists of two types of unit which are quite different in many respects:
- the dark lowlands or maria
- the brighter highlands or terrae.

The difference in albedo reflects the difference in chemical composition. The highlands represent the ancient lunar crust in which feldspar is highly enriched. This crust solidified at a time when the accretion process of the Moon was not yet complete. Impacts of objects with a wide range of dimensions produced impact craters of all sizes and formed the pock-marked surface of the lunar highlands. The mare regions represent large basins formed by impacts of very large objects, and were later filled with magmas from the deeper interior of the Moon. These magmas were rich in FeO and TiO_2 and upon crystallisation formed basalts containing up to about 25% by weight of ilmenite, an opaque mineral responsible for the low albedo of the regolith formed by the crushing of these mare basalts by later impacts. For the mare basalts, crystallisation ages in the range 3.2–3.8 billion years were found, while the highland rocks are generally older. This difference in age is also reflected by the higher crater density (approximately an order of magnitude) in the lunar highlands compared with the mare regions.

Table 1.2 shows the chemical composition of lunar soil samples typical of the lunar surface. All soil samples reflect the mean composition of the rocks from which they were derived by diminution. They also contain a small contribution from the meteoritic influx which caused the diminution. The meteoritic component never exceeds a concentration of 2–3%. Nevertheless, the meteoritic component is of great importance because it contributes fine-grained metallic iron with 6–8% Ni, a potential source of a highly valuable raw material.

Table 1.2. Composition of lunar soils

Compound	Sample 10084 Apollo-11 Mare Tranquillitatis	Sample 12070 Apollo-12 Oceanus Procellarum	Sample 60601 Apollo-16 Descartes Highland
SiO_2 %	42.1	46.0	45.4
TiO_2	7.2	2.7	0.55
Al_2O_3	13.0	12.7	26.3
Cr_2O_3	0.27	0.33	0.11
FeO	15.4	16.3	5.6
MnO	0.20	0.21	0.07
MgO	8.0	9.7	6.5
CaO	11.3	10.6	16.9
Na (ppm)	3150	3090	3510
K	1090	1900	890
Co	27	42	31.2
Ni	280	200	400
La	15	33	13
U	0.35	1.7	0.57

Not only the highland soils, but also most of the highland rocks collected during the Apollo missions, are breccias, i.e. mechanical mixtures of various components, including the meteoritic component just described. Of these components, only the most abundant, the feldspathic component, is observed in pure form as anorthosites. Another important component is KREEP, a component rich in K, REE and P, but also containing all other large ion lithophile (LIL) elements, i.e. Zr, Ba, Th, U etc., in highly enriched abundance. Although the genesis of KREEP is not yet fully understood, it is clear that it originates from a melt in which all incompatible trace elements were highly enriched, representing either a very low degree of partial melt from the lunar interior or a late residual melt of an originally largely molten Moon.

Another important component in lunar-highland breccias found at the Apollo-16 landing site, although never observed in pure form, is an Mg-rich component which might – after the loss of some olivine – represent either primitive material from which the Moon was formed, or a kind of komatitic liquid produced by a high degree of partial melting of unfractionated regions of the lunar interior.

■ Bulk composition of the Moon

As with the Earth, it is possible to estimate the composition of the lunar interior and hence the bulk composition of the silicate portion of the Moon from the composition of certain rock types found at the surface. As on Earth, komatites produced by a high degree of melting of the mantle play the most important role in this respect. The bulk composition of the silicate portion (mantle) of the Moon derived in this way is listed in Table 1.3, together with that of the Earth.

Whereas the central metallic core of the Earth comprises about one third of its total mass, the lunar core can only make up a few percent of the Moon's mass. The fact that the Moon can only have a comparatively small core has been known for a long time because of its considerably lower density compared with the Earth. The upper limit for the lunar core shrank even further when it became clear that the density of

Table 1.3. Bulk composition of the silicate portions of Moon and Earth

Compound	Moon			Earth
	Ringwood, Seifert and Wänke (1987)	Wänke and Dreibus (1982)	Jones and Delano (1989)	Wänke, Dreibus and Jagoutz (1984)
SiO_2 %	43.24	44.2	42.4	45.95
TiO_2	0.30	0.18	0.19	0.23
Al_2O_3	3.72	3.76	3.7	4.20
Cr_2O_3	0.32	0.37	—	0.44
FeO	12.24	12.7	13.6	7.58
MnO	0.16	0.16	0.19	0.13
MgO	36.85	35.5	37.0	36.85
CaO	3.03	3.15	3.0	3.54
Na_2O	0.06	0.06	—	0.39
Ni (ppm)	2487	1630	—	2080
Co	95	80	—	104
V	79	82	—	77
Sc	14	14.3	—	17.2

the lunar mantle exceeds that of the Earth's mantle due to the higher FeO content on the Moon.

In the early days of the Apollo investigations, it was thought that the Moon might be considerably enriched in refractory elements such as Al, Ca, Sc, REE or U. Since then, however, it has become clear that there is no geochemical evidence for such a proposition. On the other hand, it is evident that the Moon is depleted relative to the Earth in all elements more volatile than Na. It is also likely that the lunar mantle has lower abundances of siderophile elements. However, as the estimated relative abundance of siderophile elements in the lunar mantle decreases with the increasing siderophile character of the elements (lowest for the noble metals; Fig. 1.5), it is likely that this depletion is due to the segregation of small amounts of metal, in the order of 1 % or less, which extracted all siderophile elements according to their metal-silicate partition coefficients.

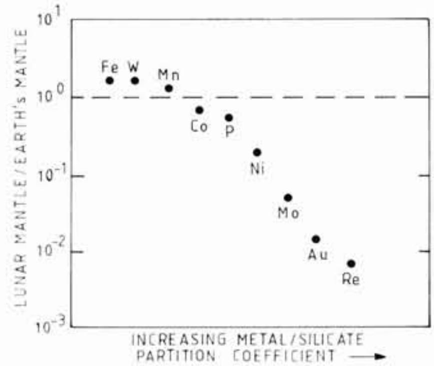


Figure 1.5 Abundance ratio of siderophile elements in lunar and terrestrial mantles

Despite some differences in the chemical composition of the lunar and the terrestrial mantle, most of which can be explained straightforwardly by secondary processes, the similarities are striking. In particular, the depletion of Mn, Cr and V – a feature of the Earth's mantle that is not yet understood – is also observed in the lunar mantle. Like the Earth's mantle, the latter contains surprisingly high concentrations of moderately siderophile elements such as Co. It should also be noted that terrestrial samples and lunar samples plot along the same oxygen-isotope fractionation line, while practically all other extraterrestrial samples (meteorites) do not (Fig. 1.6). Among the differentiated meteorites, the eucrites and the SNC meteorites (the latter probably representing rocks from Mars) definitely plot off the lunar–terrestrial oxygen-isotope fractionation line.

Although there are strong opinions to the contrary, for many geochemists the chemical similarity of the terrestrial and lunar mantles is taken as a strong indication of a clear genetic link between the Moon and the Earth (Ringwood 1986; Wänke & Dreibus 1986). A model for the origin of the Moon via a collision of a Mars-sized

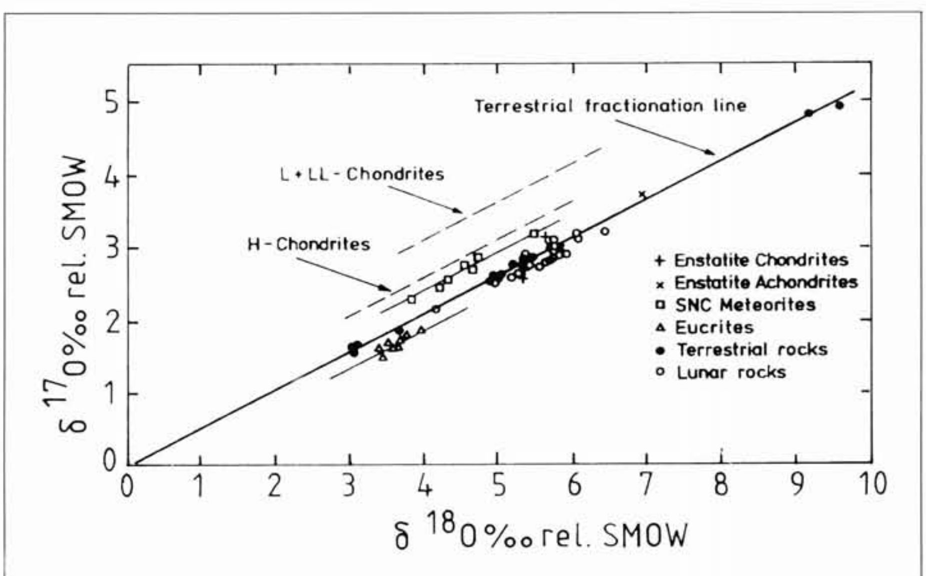


Figure 1.6 Oxygen isotopes in terrestrial, lunar and meteorite samples

object (so-called 'giant-impact model') has become popular during recent years. Such a model can explain some of the observations, but not all. Several smaller high-velocity impacts, rather than one giant impact, would fit the geochemical observations better.

1.5.2 Scientific objectives

Despite more than twenty years of research on returned lunar samples, major scientific questions concerning the origin, constitution and evolution of the Moon are still unresolved. The detection of eleven lunar meteorites among thousands of recovered Antarctic meteorites has provided a stimulus for renewed lunar research, since they dramatically reflect the limited coverage of the lunar surface by the Apollo and Luna missions. The composition of lunar meteorites is significantly different from that of the lunar rocks sampled at what were thought to be typical highland sites, such as that of Apollo-16. Lunar meteorites with highland chemistry probably represent four different events. If they are random samples, they may provide a better estimate of the average composition of the lunar crust than samples from the Apollo missions. Hence, conclusions derived from Apollo-16 samples may not be valid for the whole lunar crust. Further investigations of highland samples from different sites are essential to obtain a better estimate of the lunar composition.

This example demonstrates that a small amount of additional data may significantly improve our knowledge of the composition of the lunar crust and therefore of the differentiation processes that resulted in the present structure of the Moon.

The most important scientific question to be answered is that of the origin of the Moon, for which there are several hypotheses. Although a giant-impact hypothesis has gained wide support, like all other models it is not unanimously accepted. Precise knowledge of the bulk composition of the Moon would be of great value in constraining models of its origin. For example, some models, such as the giant-impact hypothesis, require a lunar composition similar to that of the Earth's mantle after formation of the Fe-rich core. A major difficulty in making a proper estimate of the Moon's bulk composition stems from the fact that it is a differentiated planet. Any lunar sample recovered on Earth has gone through at least one melting process. Therefore, any derivation of lunar bulk composition is circumstantial and requires chemical data from many lunar samples and a sound understanding of early lunar differentiation processes.

The most obvious result of the differentiation of the Moon is the Ca- and Al-rich lunar crust. However, as noted earlier, the overall composition of the lunar crust is still not sufficiently well-known. Any new chemical data on highland rocks would help us to answer crucial questions about the crust's origin and constitution. Did the crust form by differentiation of a magma ocean? How deep was that ocean, i.e. what fraction of the Moon was involved in the melting event? How thick is the present lunar crust? To what extent is it layered? What is its average composition?

The brecciated nature of most lunar highland rocks attests to the importance of impact processes on the Moon. The most significant bombardment took place between 4.5 and 3.9 billion years ago, with a possible peak at 3.9 billion years. The identification of this peak is based on a large number of ages clustering at around 3.9 billion years.

The areal distribution of this age cluster is a very important factor in understanding its origin. For this type of research, it is essential to gather rock samples for laboratory investigation because precise age determinations and other sophisticated analyses can only be carried out in laboratories back on Earth.

Although only 1% of the lunar surface is covered by mare basalts, this rock type is of great importance in deciphering the composition of the lunar interior. Mare basalts come from depths greater than 200 km and thus provide direct information on the composition of the lunar interior. The samples with the most primitive compositions directly encountered on the Moon so far are from glassy materials that came to the lunar surface via pyroclastic eruptions. These glasses have the lowest Fe/Mg ratios and the most unfractionated patterns of refractory elements among lunar samples. It also appears that they reflect a more volatile lunar interior than previously thought. A major goal of any further lunar-sample recovery must be an attempt to gather additional samples of this component.

1.5.3 Strategy

A permanent lunar base would allow the recovery of a large variety of rock types that all have some bearing on these questions. As there is some lateral mixing on the Moon, rocks from an extended area of the surface may be sampled using a single station.

Manned as well as remote-controlled field excursions to various geological units are essential for the recovery of all kinds of lunar surface samples. Unmanned rovers controlled from Earth seem to be the best choice for long-range traverses. For the collection and transfer of samples for sophisticated analyses at Earth-based laboratories, both manned and unmanned rovers will need to be used. Preliminary on-site analysis of lunar matter by instruments onboard the rovers will reduce costs, as only selected samples must then be transported back to Earth. In certain cases, in-situ analyses under prevailing environmental conditions on the Moon would result in less risk of terrestrial contamination. Apart from sample collection, automated long-range rovers will be very useful for the installation of geophysical stations. Important information about the abundance of radioactive elements may be obtained from heat-flow measurements, assuming that there is thermal equilibrium within the bulk Moon.

1.6 Interaction of the Moon with its environment

1.6.1 Scientific background

The interaction of the Moon with its environment is one of the most interesting fields of investigation in the context of a renewed lunar exploration programme. The record of past external radiation (solar and solar-wind) is also of particular importance.

The Moon is almost completely devoid of atmosphere and intrinsic magnetic field, which means that its surface is in direct contact with the interplanetary medium (meteoroids, micro-meteoroids, solar wind, solar and galactic cosmic rays). Moreover, for more than three billion years its geological evolution has almost completely ceased, and consequently the evolution of the lunar surface has been dominated by external processes for the largest part of the Moon's history.

■ **Meteoroid and micro-meteoroid impacts**

Meteoroids of all sizes (from a few micrometres to several tens of kilometres) have impacted on the surface of the Moon, forming craters of all sizes, the largest, Copernicus and Tycho, having diameters of 90 and 85 km, respectively. This resulted in the formation of the lunar regolith, a layer of dust and rocky fragments 5 − 10 m thick on level areas. Extensive mining operations aside, all the material available at the surface for processing, shielding, etc. has been modified by external processes, the outcome of which is well known from lunar-sample studies (Langevin & Arnold 1977).

These modifications are very significant in terms of mineralogy, petrology and (to a lesser extent) chemistry. Up to 50% of lunar dust is composed of either glassy spherules or glassy agglutinates. Liquid droplets ejected during an impact either solidify in flight (spherules), or weld together surface grains (agglutinates). The number of glassy particles increases with the exposure-time, or 'maturity', of a given soil sample. On a larger scale, most rocky fragments at or near the surface (a few percent of the volume of the regolith) are breccias, large aggregates of grains welded together by the shock wave of an impact. These breccias are called 'polymict' when constituted by many different mineralogical species. The 'monomict' breccias result from the fracturing of the bed-rock by large impacts. Below the highlands, which witnessed the intense bombardment phase following planetary accretion, this fractured layer, or mega-regolith, extends over several kilometres. In summary, almost all materials available near the surface have experienced shock processes.

■ **Solar wind, solar flares and galactic cosmic rays**

The role of particle fluxes is also significant in relation to a few key aspects of lunar petrology and chemistry. The solar wind is by far the most intense of these particle fluxes (10^8 particles cm^{-2} s^{-1}), consisting mainly of hydrogen and helium, with small contributions of heavier elements. These large fluxes of low-energy particles (1 keV/amu) result in the formation of an amorphous layer a few tens of nanometres thick on small grains in mature soils. This layer has the interesting effect of promoting welding under pressure.

Solar-wind contributions to the lunar surface chemistry, although relatively small, dominate two key elements which are almost completely lacking in the bed-rock: hydrogen and carbon. These two elements (particularly hydrogen, hence water) represent the major challenge for an autonomous lunar base. Both H and C are present at a level of about 100 ppm in mature soils. This equates to 100 g/t, or 1 kg/m^2 of regolith (assuming a thickness of 5 m and a nominal density of 2). Therefore, obtaining the quantities of hydrogen and carbon required for the operation of a lunar base represents a formidable mining and processing task, even assuming that a lunar base operates as much like a closed system as possible. High-energy ions from solar and galactic cosmic rays induce nuclear reactions with regolith atoms as well as the formation of nuclear particle tracks. The stable and radioactive isotopes produced by these high-energy particles have been used for determining the exposure ages of lunar samples. Conversely, the observed irradiation effects can help unravel the history of external radiations over several billion years.

The role of the lunar regolith as a long-term record of interplanetary conditions in the Solar System has provided one of the major scientific breakthroughs of the lunar-sampling programme. In this respect, the Moon presents a unique opportunity. These radiations are also recorded on asteroid surfaces, and hence in meteorites which are mostly fragments of asteroids. In fact, the stability of the cosmic-ray intensity over the last several million years was first established on the basis of spallation product abundances in meteorites. However, when longer time scales of up to a few billion years are investigated, the meteoritic record is much harder to read than the lunar record. For samples taken at or below the lunar surface, information about the geologic and stratigraphic context was available, which made it possible to reconstruct the exposure history. The results obtained to date from lunar samples and meteorites pertaining to the history of radiations are summarised below.

The composition of the solar wind in earlier epochs is best recorded in the mineral ilmenite (titanium-iron-oxide), found exclusively in mare basalts. No substantial variations over the last few billion years have been detected. There is evidence of a modest secular decrease in the ^4He / ^3He ratio and in the xenon abundance. These observations signal a change in the dynamics of the corona. The change in the helium isotopic abundance could also indicate a modest degree of slow mixing by the Sun to depths below the boundary of the outer convective zone (cf. Geiss & Bochsler 1991). Studies of the amorphous layers on lunar grain surfaces produced by solar-wind irradiation seem to indicate a systematic increase in solar wind speed during the last two billion years. The total number of solar particles collected on the Moon and the distribution of solar noble-gas isotopes reveal that either the solar-wind flux or the flux of suprathermal particles, or both, were higher in the past than they are today.

The solar-flare record in lunar material consist of tracks, trapped flare particles, and spallation products (Walker 1980). Flare frequencies and sizes are highly variable. The eleven-year periodicity is well documented, but so far the lunar record has not revealed other periodicities. Thus, it remains to be decided whether variations on longer time scales are purely statistical. There is evidence in the composition of trapped solar noble gases that can be interpreted in terms of high fluxes of suprathermal particles in the past. Gradients of particle tracks within large grains showed no strong variations in the energy spectrum of solar cosmic rays. It was also possible to determine abundances of broad groups of elements in solar cosmic rays (iron group, VVH elements) up to 1.7 billion years ago.

The intensity of galactic cosmic rays over the last several million years has been constant (Fisk 1980). However, short intensity peaks, as may be expected after nearby supernovae explosions, cannot yet be excluded. Average fluxes over the last few billion years were not very different from the present-day flux. The record of these very ancient radiation epochs, which cover many galactic radiation periods, could be improved by means of a sampling programme focussed on radiation studies.

While tantalising results were obtained, these studies were limited to the last 1 to 2 billion years, due to the maximum coring depth, as well as the limited documentation in the close vicinity of the core. These important objectives would benefit greatly from a dedicated sampling programme, taking into account the vastly improved understanding of regolith structure and dynamics.

■ **The lunar exosphere and its evolution**

Another interesting component of the lunar environment is its exosphere, which results mainly from the outgassing of volatile species formed by solar-wind implantation (H, H_2, He, CO, H_2O) and radiogenic gases (^{40}Ar, Rn). The main loss process for hydrogen and helium is gravitational escape. For the heavier gases, it is the photo-ionisation and charge exchange with solar-wind protons which results in their acceleration by the solar wind $V \times B$ force, which is much stronger than the Moon's gravitational field. The lifetime of these species in the lunar exosphere is only a few months. This steady-state envelope therefore has an extremely low density (a few 10^8 cm^{-3}), and is nearly collision-free. The molecules can travel over large distances on ballistic trajectories, which results in a continuous distillation process: species evaporate continuously on the day-side and condense on the night-side. It has been proposed that polar regions may constitute cold traps with significant accumulation of volatiles. This is debatable, however, because of destructive effects such as sputtering.

In the past, either sporadically or continuously, indigenous lunar gases, gases of cometary origin or other non-solar components may have been more abundant in the lunar atmosphere. Indigenous lunar gases undergo the same loss or re-trapping processes as the solar-wind gases. Thus there is a chance of studying the indigenous lunar atmosphere of past epochs by investigating the trapped gases. So far, indigenous lunar ^{40}Ar (resulting from ^{40}K decay on the Moon) has been unambiguously identified, but there is also evidence of non-solar nitrogen, carbon and xenon in the trapped gas.

It is interesting to assess the possible impact of manned activities on the lunar exosphere. It is of course difficult to define what would be an unacceptable contamination level. One possible criterion is to relate the magnitude of anthropogenic emissions to that of the naturally occurring situation. Due to the large mean free paths, one can assume that contaminants are redistributed over the exosphere as a whole. Elements contributed by solar wind are almost in a steady state. As an example, carbon is a major potential contaminant, as it is present at very low concentrations in lunar material. It represents 4×10^{-4} of solar-wind ions. The total flux received by the Moon is therefore 1.2×10^5 atoms cm^{-2} s^{-1}; i.e. 2.7×10^{29} atoms/yr or 5 t/yr. The carbon contribution of meteoroids is expected to lie in the same mass range. In conclusion, it seems safe to assume that if carbon emissions in gas form can be constrained at less than 5 t/yr, human activity would not significantly modify the lunar environment as far as this particular element is concerned.

1.6.2 Scientific objectives

The scientific goals related to the interaction of the Moon with its environment are as follows:

- Meteoroid and micro-meteoroid impacts, solar wind and cosmic rays are expected to dominate the evolution of Solar System bodies with neither atmosphere nor internal activity: all the asteroids, most satellites (with the exception of Io and Europa, for which internal processes are dominant, and Titan, with its thick atmosphere), and Mercury (which has a small but significant dipolar magnetic field). The dynamics of regolith formation and evolution therefore have widespread implications throughout the Solar System.

- Micro-meteoroid impacts and solar-wind implantation have a strong influence on the remote-sensing signatures of planetary surfaces, particularly in the visible and near-infrared ranges. In the framework of a new lunar exploration programme, the relationship between surface processes and remote sensing can be investigated in detail by combining the global coverage provided by a precursor mission (e.g. NASA's Lunar Observer), the detailed in-situ investigations provided by rovers and ground stations, and the analysis of selected samples brought back to Earth. It will then be possible to calibrate the remote-sensing signatures obtained from orbit using the actual chemical, mineralogical and petrological characteristics of surface material. Such results are essential for improving our interpretation of remote-sensing observations for the majority of the lunar surface (where no in-situ studies will be performed), as well as that of other regolith-covered bodies (asteroids, satellites, Mercury), for which extensive in-situ investigations or sample-return missions are not being considered for the foreseable future.

  The Moon can provide a useful platform from which to monitor the interplanetary medium and the Earth's geomagnetic tail over complete solar cycles. The chemical composition of the solar wind and its variation over the solar cycle can be investigated by deploying collectors with an area of several square metres from which implanted species can be retrieved. For a small fraction of its orbit (a few days), the Moon is embedded in the geomagnetic tail. During these passages, the geomagnetic tail can be studied from the near side of the Moon. Perturbations induced by the remnant magnetic fields are not negligible. Therefore, a dedicated satellite spending long periods in the geomagnetic tail (such as those of the International Solar-Terrestrial Physics Programme) will provide higher quality results. However, lunar-based observations can be considered as a useful complement, providing a monitoring capability over very long time periods.
- The lunar exosphere, its composition, as well as the distillation and loss processes, or the possible accumulation of volatiles in the polar regions, can be studied directly from lunar orbit with ion and neutral mass-spectrometers. It would be useful to monitor the increase in ion and neutral density linked to surface activities, so as to assess the magnitude of anthropogenic contributions to volatile concentrations.
- Most importantly, as demonstrated in the previous section, the Moon, when compared with other regolith-covered bodies, constitutes an ideal laboratory for tracing back the evolution of the interplanetary medium during the last 4 billion years. It is by far the easiest of these bodies to reach. The thickness of the regolith is about 5 – 10 m, which is much smaller than that expected on asteroids or satellites such as Phobos (a few hundred metres). This easily-sampled layer contains grains directly exposed at the surface throughout the period since surface material was emplaced (about 4.5 and 3.5 billion years for highlands and mare, respectively).

1.6.3 Strategy

The long-term evolution of the solar wind, solar flares and galactic cosmic rays has important implications for the evolution of the Solar System and its galactic environment. An important role of a renewed lunar exploration programme would be to provide deep cores reaching the lower levels of the regolith, where the oldest exposed grains are likely to lie. The deepest core sample to date (Apollo-17) reached a depth of 3.2 m. An increase to 5 or at most 10 m does not represent a major technological challenge. In fact, the most difficult problem is retracing the stratigraphic history of a core. In contrast to Antarctic cores or deep-sea sediment cores, which present a continuous and regular record, lunar cores are built up by a stochastic process, with both accumulation and retrieval episodes. Using cosmogenic isotopes, it is possible to obtain the exposure date of each layer with an accuracy of 10 to 20%. The detailed documentation of the surroundings of the core site should improve the reliability of these reconstitutions, by determining the size and age of nearby craters. Furthermore, a sound knowledge of regolith dynamics ensures the selection of favourable sites, where accumulation is expected to be more regular. An outstanding opportunity is provided by regions at the base of a hill, as lunar grains are subjected to a down-slope transport process induced by small meteoroid impacts.

When comparing the contributions of the Apollo and Luna programmes, it is quite clear that manned operations, particularly with geologically trained astronauts, provide by far the best means of obtaining deep cores at selected, well-documented sites, although the cost/benefit relationship may favour automated operations. The crew can also help to ascertain the pristine and undisturbed nature of a sample and its environment. This task is comparable with that of an archeologist establishing a complete record of the site and securing scientifically valuable material before the real-estate people, road builders and miners arrive!

1.7 A renewed lunar-science programme

The scientific goals for a new lunar exploration programme, which have been briefly outlined in the previous sections, can best be achieved within the framework of a multi-mission programme:

1.7.1 Lunar polar orbiter

The whole Apollo Programme, and probably the Soviet lunar exploration programme also (at least in its initial stages), was dedicated to the preparations for and implementation of a manned landing, and so landing-site selection was the main objective of remote-sensing investigations. As a result, only 14% of the lunar surface was observed from close range, with no observations at medium to high latitudes. This very incomplete coverage has severe implications for almost all aspects of lunar science, in particular our understanding of the Moon's global chemical composition and hence the origin of the Earth–Moon system. Moreover, there has not been a lunar orbiter in nearly 20 years, during which time remote-sensing techniques have vastly improved in terms of both sensitivity and spectral and spatial resolution. A first step in any scientifically meaningful lunar programme is, therefore, to obtain complete coverage of the lunar surface using modern remote-sensing techniques, which is best achieved from a polar orbit.

A lunar polar-orbiter is also essential to extend our gravity-field, magnetic-field and altimetry knowledge to the whole surface of the Moon, and to implement the great improvements in sensitivity that can be achieved with up-to-date instrumentation. As shown in the section on geophysics, our understanding of the internal structure of the Moon, in particular the thickness of the lunar crust as a function of latitude and longitude (near-side/far-side asymetry) would benfit strongly from such combined studies on a global scale.

Such a lunar polar orbiter has been studied by various agencies: ESA (Polo), USSR (Selene) and NASA (Lunar Observer). The launch of such a satellite is now being considered by NASA as the first mission within the framework of the manned Space Exploration Initiative. In the Russian planetary programme, Mars missions now seem to have higher priority. If these opportunities do not materialise, it may be interesting to reconsider an ESA mission of this type. A lunar polar orbiter is indeed one of the few remaining affordable planetary missions with strong scientific potential which could be carried out by ESA alone.

1.7.2 In-situ studies

The next step in lunar exploration should be an extensive in-situ science programme. Indeed, due to communications constraints, the six Apollo landing sites, as well as the three Luna landing sites, were all on the Moon's near side. In addition, for safety reasons, most were in the relatively flat maria. As a result, the Apollo and Luna surface stations were strongly biased towards mare close to the centre of the near side.

One of the highest priorities at this stage should therefore be the establishment of a global network of well-instrumented surface stations. One of the key objectives would be the global monitoring of lunar seismic activity, from which the size of a lunar core and the depth of other major discontinuities in the lunar interior could be determined. The surface stations could also record signals from active seismic experiments. The local chemical and mineralogical composition can also be assessed. These stations will extend our in-situ knowledge of the lunar surface to new regions, particularly the far side and the highlands, which are not well represented in the Apollo and Luna samples. In combination with the global remote-sensing coverage provided by a lunar polar-orbiter mission, the surface stations will vastly improve our understanding of the lunar chemistry and mineralogy on a global scale. Such stations could benefit strongly from experience with similar surface stations to be deployed on Mars in the framework of the Russian Mars-94 programme or the Marsnet proposal presently being studied by ESA.

A complementary approach is the use of lunar rovers. Owing to the relative smallness of the Moon, a mobility of several hundred kilometres would allow the observation of diverse geochemical provinces. Such rovers could deploy surface stations at selected locations, perform profiles across geochemical and geophysical boundaries, and implement sounding techniques.

Compared to the Mars rover, the lunar rover represents a much simpler technological challenge. The lunar terrain is relatively flat and predictable, and the communications-link round-trip time makes real-time operator intervention possible, allowing a semi-autonomous mode of operation, at least on the near side (and possibly on the far side with a relay satellite). It can be argued that a lunar rover would represent a good precursor to a Mars rover with a high level of autonomy. A second-generation lunar rover would be capable of carrying several astronauts in a closed environment for about two weeks.

1.7.3 Sample return

Despite the rapid developments taking place in in-situ instrumentation, laboratory techniques will remain ahead in most areas by very significant factors such as sensitivity and resolution. Although it is possible to benefit from instrument advances for sites already sampled, the distribution of past lunar sampling sites was strongly biased towards the near side and mare material. The major thrust of a renewed lunar sampling programme must therefore be to complement the Apollo and Luna samplings with material from various highland, far-side and polar sites. Another important goal is to obtain samples dedicated to specific objectives, such as deep cores in accumulation regions for monitoring the history of particle fluxes over billions of years.

1.7.4 Lunar base

The second stage of the proposed programme (surface stations and rovers) would benefit from lunar-base activity (refurbishment of rovers, manned maintenance of surface stations), but could be implemented at least in part as an autonomous mission. The scientific mission that would most benefit from the presence of man on the Moon would clearly be such a new lunar-sampling programme. Indeed, one of the key aspects for the interpretation of results from a sample is its documentation, i.e. the accurate and exhaustive description of the sampling site. Some samples, such as the proposed 5 − 10 m lunar cores extending to the base of the regolith, may require relatively heavy logistics. Furthermore, for such objectives as the history of the lunar environment, a sampling site close to a lunar base would not be a disadvantage. It should also be noted that the operation of a lunar base will require relatively frequent round trips from Earth during the construction and operational phases. The implementation of an extensive lunar-sample acquisition programme would be quite straightforward in this context. At the stage of an operational lunar base, it is possible to consider ambitious scenarios such as astronaut-driven vehicles undertaking a field geology programme.

1.8 Role of Europe

The European scientific community could be involved in most aspects of a lunar science programme, given its recognised excellence and expertise in several broad key areas:

- geosciences
- remote sensing
- in-situ techniques
- sample analysis.

1.8.1 Geosciences

The geosciences community is very strong in Europe. During the last ten years it has become increasingly involved in space missions for studying both the Earth and the planets. There is now an active community applying geophysical and geochemical techniques to planetary exploration. These include remote-sensing techniques for geodetics, gravimetry, magnetic-field measurements and spectroscopy. These orbital techniques can easily be transposed to a lunar polar orbiter. Other important techniques such as seismology and heat-flow measurements require surface stations. One cannot stress enough the important role that field geologists can play in the training of astronauts for a manned lunar programme, as was demonstrated by the much higher quality of the last Apollo sampling activities when compared with those from Apollo-11 and 12.

1.8.2 Remote sensing

The involvement of European scientists in planetary missions has increased dramatically over the last ten years, due to the launch of ESA's first interplanetary mission, Giotto, in 1986, as well as the opportunities provided by Russian and American missions. This has been demonstrated by the very large contributions of the European Principal- and Co-Investigators to the major planetary programmes of the 1990s: the Cassini mission to Saturn and Titan, a joint undertaking by ESA and NASA, the Russian Mars programme and, to a lesser extent, the Comet-Rendezvous Asteroid Flyby (CRAF) and NASA's Mars Observer (MO). Indeed, Europe's expertise now spans the full spectrum of remote-sensing techniques, gamma-ray and X-ray spectrometers, ultraviolet spectrometers, cameras, infrared spectral imagers, radar sounders and imagers.

1.8.3 In-situ techniques

The European scientific community is also at the forefront in the field of in-situ investigations, as witnessed by the major European contributions to the payload of the Huygens descent probe, the Phobos lander, and the in-situ modules of the ambitious Russian Mars-94/96 missions (balloons, small surface stations and mini-rover). The first spatialised high-resolution mass-spectrometry chemical analyser on NASA's CRAF mission (now cancelled) was also the responsibility of a European Principal Investigator. In summary, the European scientific community can now provide competitive proposals for virtually all aspects of a lunar remote-sensing/in-situ-analysis programme.

1.8.4 Sample analysis

The sample-analysis community in Europe has developed along two lines, guided by the strong interest in meteorites due to the rich European collections, collected over more than two centuries, and the lunar sample-analysis programme, which triggered a renewal of interest in extraterrestrial material. Indeed, it can be argued that this lunar science programme, with the strong drive for improving sensitivities induced by the small sample quantities, is in a large part responsible for the discovery of isotopic anomalies in meteorites, which required the isolation and analysis of very small inclusions of carbonaceous chondrites.

More than forty teams in Europe have been involved at one time or another in the sample-analysis programme. A strong community dealing with isotopic analysis, seleno-chronology and rare-gas studies has developed. Another strong field has been the study of radiation effects, from the solar wind to nuclear reactions induced by galactic cosmic rays. Most of these teams then applied their techniques to meteorites, and, more recently, were involved in the new collection techniques that have broadened the supply of extraterrestrial material: stratospheric sampling of small grains, the collection of new meteorites and micro-meteorites in both the Arctic and Antarctic polar regions, and the collection of micro-meteoritic material in Earth orbit. This convergence of lunar-sample and meteorite studies has been dramatically enhanced by the discovery in Antarctica of several meteorites that have proved to be of lunar origin. These objects also exhibit puzzling differences which emphasise the biased character of the lunar sampling sites. European scientists are playing a key role in all these activities, on a par with the American community and, for the Antarctic meteorites, the strong Japanese programme.

As the necessary scientific expertise is so clearly available in Europe, the only potential constraint on the European scientific community's involvement in a lunar exploration programme stems from the current limited availability of funding within the ESA Member States. However, Europe could already undertake a number of studies (e.g. lunar exosphere) in preparation for any possible future lunar mission, using the exploration strategy outlined in the previous section as a basis.

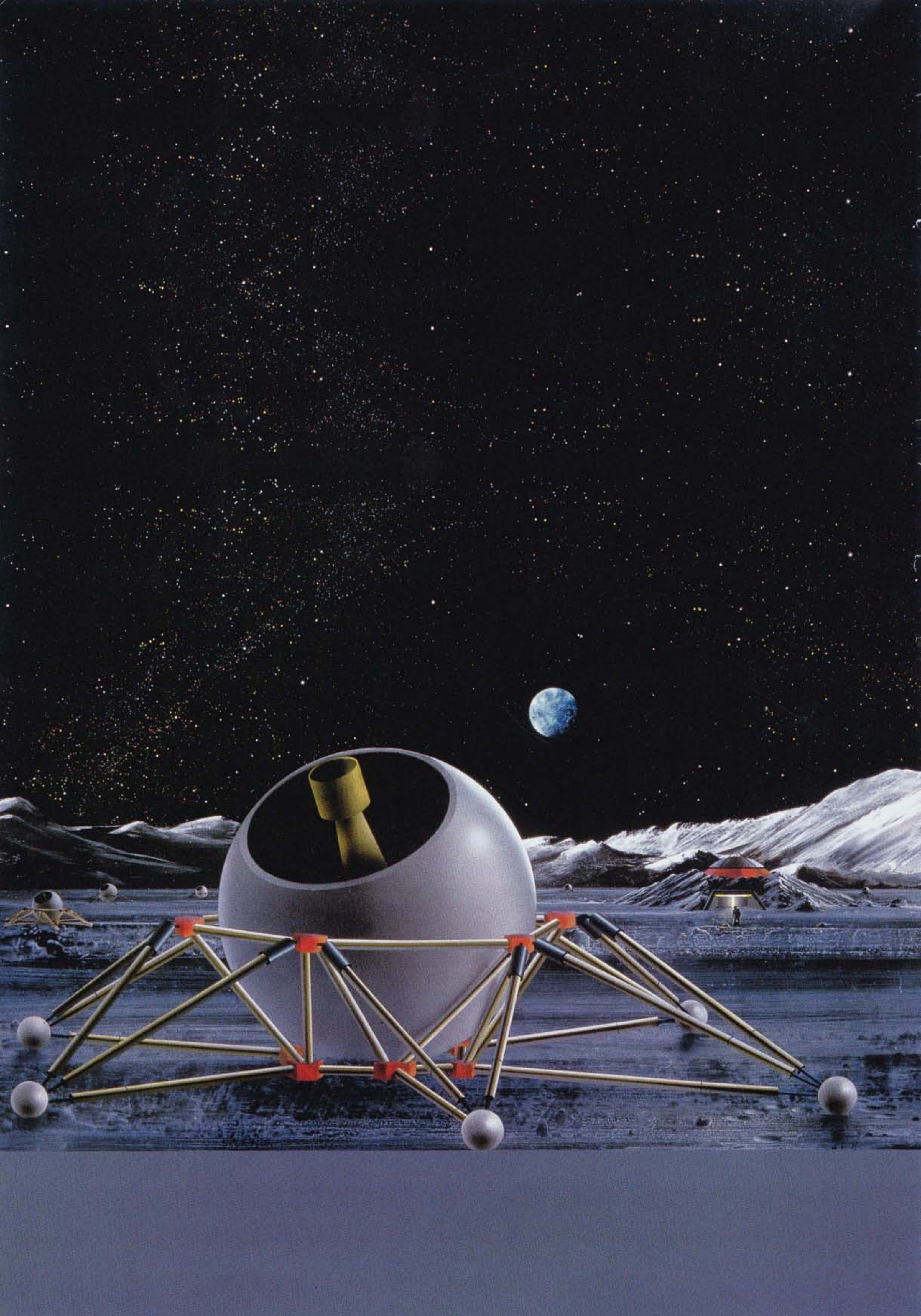
1.9 Conclusions and recommendations

A lunar polar orbiter is considered a mandatory precursor mission for any high-level lunar exploration programme. Such a programme should also include the implementation of surface elements such as surface stations and rovers. The lifetimes of these elements could be extended at a later stage, using the possibility of refurbishment provided by the presence of a manned lunar base. A sample-return programme would benefit greatly from the presence of trained geologists on the Moon, as careful site selection and documentation are critical for making the most of the subsequent laboratory investigations of samples.

The implementation of such an ambitious programme, extending well into the next century, is likely to require worldwide cooperation. Major decisions about lunar exploration are expected to be taken in the next few years. Within the framework of the Space Exploration Initiative (SEI), NASA is considering a sequence of missions starting with a lunar polar orbiter (Lunar Observer) and leading to a permanent lunar base. ESA and the Commonwealth of Independent States have also studied lunar polar orbiters – Polo and Selene, respectively – and Japan's ISAS is preparing a lunar mission (Lunar-A) with an orbiter and three penetrators. ESA should take steps to provide access to such missions to the European scientific community (data analysis, procurement of experiments, analysis of new samples). Furthermore, we strongly recommend that ESA participate fully in such an international effort.

A lunar polar orbiter would represent an attractive early European contribution. After completion of this first step of the programme, European-supplied surface stations and rovers would represent a very significant further contribution within reasonable budgetary constraints.

2. Astronomy and Astrophysics



Chapter 2. Astronomy and Astrophysics

2.1 Introduction

■ What makes the Moon a desirable astronomical site?
The Moon clearly offers enormous advantages for astronomical observation compared with Earth-based astronomy. These include:
- no atmosphere, no ionosphere, no wind, little corrosion, little or no secondary cosmic rays
- high stability, and low levels of seismicity
- no electromagnetic radiation from the Earth on the far side of the Moon
- very low, stable temperatures (70 K on the ground) and no light in polar craters
- an attitude in space known in terms of both position and libration
- very low rotation rate, allowing long-duration observations.

■ The Moon's drawbacks
- Micrometeorite impacts yielding micro-craters; a hundred such craters larger than 0.05 mm are formed per square metre per year.
- Dust pollution caused by vehicles and humans.
- Large temperature variations between lunar day and night at non-polar sites (from 100 K to 385 K at the surface).

Some of the advantages are also shared by facilities mounted on spacecraft. The main advantages of the Moon in this respect are the stability and large dimensions of the possible interferometric baselines, the potential for improved thermal stability, and the possibility of extending and servicing the instruments, whereas the disadvantages are minor or non-existent. However, the price to be paid is considerably higher than that for spacecraft. One must therefore decide whether observation from the Moon is really advantageous and the choice is not always clear-cut.

■ Protection against interference and other nuisances
It is essential that lunar astronomical sites be strongly protected from man-made nuisances of all kinds. The importance of this protection has already been underlined by international organisations such as the IUCAF, CCIR, IAU and ICSU. In particular, the far side must remain completely free from interference, particularly radio, with the exception of a few relay satellites. It may be that in the long term, alternative transmission systems such as optical fibres or optical links to satellites will have to be implemented. Also, all possible measures should be taken from the outset to avoid local radio and dust pollution in the vicinity of manned sites. Specifications to minimise this pollution should be defined and verified for any equipment sent to the Moon.

■ Guidelines for the selection of astronomy projects on the Moon
Selection of astronomical projects for deployment on the Moon should address the following questions:
- Are the envisaged scientific experiments really worth the effort?
- Is the Moon really necessary for these experiments, or are there alternatives?
- Is the project feasible with present technology or that of the foreseeable future?

Trying to answer these questions leads to the classification of the astronomy projects into three categories:
1. Premature or difficult-to-justify projects: Some of the projects, proposed at various times, appear to belong to this category, such as:

- A gravitational wave detection experiment. The apparent advantage of the Moon is its absence of seismicity; however it seems that the problem of suppressing the effect of Earth seismicity has been essentially solved in the preparatory studies for the European VIRGO project.
- A large (10–16 m class) diffraction-limited optical telescope. This project is already difficult on the ground, barely feasible on the Moon at present, and in any case extremely expensive.
- A very large Arecibo-type radio telescope. This appears to be marginally feasible, but is certainly not a first- or second-generation project. Moreover, the scientific return may not warrant the level of effort involved.
- Imaging (astrophysical) Very Long Baseline Interferometry (VLBI) between the Moon and the Earth. The fixed length of the baseline allows marginal imaging. It would be considerably more interesting to perform VLBI using Earth satellites in eccentric orbits. However, Moon–Earth VLBI is of very great interest for the astrometry of point sources because of the high level of knowledge and high stability of the Moon–Earth baseline, considered later in this report.

2. Projects that can be undertaken if there is an opportunity: Some of the projects contemplated at various stages raise obvious questions as to the necessity or advantages of observing from the Moon compared with observing from satellites, probes or even the Earth, but may still be worthy of discussion if an opportunity for them arises. Examples are:
   - Instruments for the detection of X-rays, gamma-rays, cosmic-rays and neutrinos (with the exception of large gamma-ray detectors in the 10–100 GeV energy range). This is discussed in the High-Energy Physics chapter of this report.
   - Small multi-purpose telescopes (UV to IR range), solar patrol instruments, interplanetary dust collectors, etc. A comparison with satellite equipment must be carried out, although it is rather unlikely that the result will be in favour of the Moon a priori.

3. Projects for which a lunar implementation appears most appropriate: Three types of instrument that seem worthy of detailed study have been selected:
   - Interferometers for UV to sub-millimetre wavelengths.
   - Instruments for astrometry, including Earth–Moon VLBI.
   - Radio telescopes for very low frequencies.

These instruments would be used not only for galactic and extragalactic astronomy, but also for the observation of Solar System objects. They are examined in some detail below.

2.2 Interferometry

2.2.1 Summary

The Lunar Interferometry Study Team (LIST) report:

(i) addresses the scientific case for optical and sub-millimetre interferometry from the Moon

(ii) identifies some target parameters for lunar interferometry

(iii) outlines a strategy for the development of a lunar interferometry programme, and
(iv) recommends a number of actions to implement that strategy.

The findings are preliminary and further in-depth study may alter some of our conclusions. However, it is clear at this stage that the scientific impact of a high-angular-resolution interferometer on the Moon will be enormous, providing in the order of 100 μarcsec resolution in the visible range for objects as faint as 23 rd magnitude, and 100 marcsec resolution in the sub-millimetre range (500 μm). This report provides more detail on the scientific case. It also appears at this stage that separate interferometers are required for the ultraviolet/optical/infrared and sub-millimetre regions.

Although the astrometry area is the topic of a separate report by the Astrometry Study Team, the possibility of astrometric measurements using interferometry with microarcsecond precision is also considered here.

Observations from Earth suffer from the limited transparency of the atmosphere and atmospheric turbulence ('seeing'). Both space- and Moon-based interferometers are free of these limitations, and the Moon has additional advantages over space by providing a stable base, a very long baseline (1 km), and expandability. On the other hand, the realistic time-scale of a lunar interferometer, the need to validate various technologies, and the great scientific benefit resulting from a more modest baseline ($\approx$ 100 m) all point to the need for one or more precursor space-based interferometers.

The Study Team's recommended strategy is therefore as follows:
- ESA should commence conceptual studies for interferometers, in readiness for their location on the Moon, in the UV/visible/IR and the sub-millimetre region.
- Considering (i) the scientific benefits, (ii) the technical readiness, and (iii) the realistic time-scale for a lunar interferometer, ESA should plan for at least one 'precursor' interferometer mission in space.
- ESA should identify key technologies of crucial importance to space-based interferometry in general, and lunar interferometry in particular.

Actions to implement this strategy have been identified and are summarised in Section 2.2.6.

2.2.2 Introduction

■ Earth-based high-angular-resolution interferometry

The Rayleigh criterion for the minimum resolvable separation α of two point sources by a diffraction-limited telescope of diameter D at wavelength λ is

$$\alpha = 1.22\frac{\lambda}{D}$$

This same formula applies to a sparse array of telescopes whose outputs are combined ('aperture synthesis'), D being the largest dimension of the array. Thus an interferometer whose largest baseline is, for instance, 1 km is capable of an angular resolution of 125 μarcsec in the visible region of the spectrum (λ = 500 nm) and 125 marcsec in the sub-millimetre region (500 μm or 600 GHz).

The presence of the atmosphere severely limits Earth-based observations in two ways. First, there is very considerable absorption in the ultraviolet, infrared and sub-millimetre regions of the spectrum, making observations impossible at many wavelengths. Secondly, where the atmosphere is transmitting, the optical effect of the turbulence (i.e. 'seeing') severely limits the angular resolution. Earth-based observations can only approach this limit if a means of compensating for atmospheric turbulence is implemented. In the case of filled-aperture optical telescopes, image-plane (speckle) and pupil-plane interferometry have been shown to yield diffraction-limited images, and the technique of adaptive optics promises diffraction-limited performance in the near-infrared for 8 m-class telescopes in the near future. Adaptive optics may be capable of providing diffraction-limited resolution in the visible. Indeed, the implementation of bright laser guide-star technologies in adaptive optics provides high hopes.

Long-baseline (sparse aperture) optical interferometry was pioneered by Michelson in the 1920s, then Hanbury Brown & Twiss (1960s) and Labeyrie (1970s). At present there are approximately twelve Earth-based optical long-baseline interferometry projects in progress or funded, including the European Southern Observatory (ESO) Very Large Telescope Interferometer (VLTI). Earth-based optical long-baseline interferometry has already yielded valuable astrometric results and many stellar diameters. However, the restrictions imposed by atmospheric turbulence are severe; one of the problems facing the Study Team has been to extrapolate the expected state of ground-based optical interferometry for the coming 20 to 40 years. This depends greatly on the extent to which corrections for atmospheric turbulence can be made.

Interferometry and aperture synthesis using heterodyne detection has been under continual development since its invention in the 1940s, and heterodyne interferometers at high frequencies are thus extensions of well-developed techniques. Heterodyne detection retains the phase of the incoming signal (with a corresponding fundamental loss of sensitivity), which allows substantial flexibility in processing and recording data. Along with two interferometers working at 30 THz (10 μm; CH Townes' ISI project), two ground-based interferometers working at sub-millimetre wavelengths will test the limits of ground-based interferometry in the next few years (a CSO/JMCT collaboration and the SAO's dedicated six-element array, both located on Mauna Kea, Hawaii). Poor to nonexistent atmospheric transmission and phase stability severely limit the sensitivity of infrared and sub-millimetre ground-based interferometers, especially at these wavelengths; only space or lunar interferometers will approach fundamental sensitivity limits.

■ Space- and lunar-based interferometry

Interferometry performed outside the Earth's atmosphere has two obvious potential advantages over ground-based observations:

- the wavelength window is totally transparent
- the turbulence, 'seeing', or atmospheric phase instability associated with ground-based observation is completely absent.

In the transparent regions of the atmosphere (i.e. visible and part of the near-infrared), the case for space- and/or lunar-based interferometry (as opposed to ground-based observation) rests largely on whether the absence of atmospheric 'seeing' provides an overwhelming advantage. This issue is discussed further in Section 2.2.4 (interferometry from the Moon provides a higher signal-to-noise ratio for the very faintest objects, but does not have a clear signal-to-noise advantage in general).

The advantage of space- and/or lunar-based telescopes operating at wavelengths at which the Earth's atmosphere is opaque, or of low transparency, is self-evident and provides a strong motivation for the construction of interferometers, particularly in the sub-millimetre region of the spectrum.

For observations made outside the Earth's atmosphere, the choice between space-based and lunar-based lies in the merits of:
- the mechanical stability of a lunar-based instrument
- the potential for improved thermal stability on the Moon.

A polar site on the Moon would probably provide excellent mechanical and thermal stability.

■ **Possible options for lunar interferometry**

Initial efforts have been concentrated on the concept of a dilute synthetic-aperture-array interferometer, with the largest baseline of the order of kilometres (50 − 100 μarcsec resolution in the visible and 5 − 10 marcsec in the sub-millimetre range), and a 'limiting magnitude' (which is not a straightforward quantity to define) fainter than $m_v = 20$ in the visible. Such an array should be capable of imaging as well as astrometry, and this implies a fairly large number of telescopes (≥ 6) of the order of 1 m diameter for visible observations and 5 m for sub-millimetre work. The scientific case outlined points strongly towards the kind of orders-of-magnitude increase in angular resolution over, for instance, the design specification of the Hubble Space Telescope, that such an interferometer would provide (i.e. a factor of 100 or more).

One possible concept, already proposed by A. Labeyrie, is the Lunar Optical Infrared Synthesis Array (LOISA), an array of fifteen to thirty-three 1.5 m-diameter telescopes. One particularly novel aspect of this concept is the mobility of the telescopes, all of which move continuously during observations in order to maintain path-length equalisation for the beams. This eliminates the need for optical delay lines (which may be a major problem for long baselines).

An alternative approach would be to consider a filled-aperture telescope, such as a 16 m-diameter-class 'conventional' telescope. This could be fabricated from lightweight elements forming a single diffraction-limited collector in the sub-millimetre range, which could be one element of a sub-millimetre interferometer array. In the visible range, this collector would form a highly aberrated image consisting of slowly varying 'speckles' as the collector changed its distortion whilst tracking an object. Interferometric data collection and analysis could restore a diffraction-limited image. The advantages compared with a ground-based filled aperture would be: (i) the long speckle lifetime, and (ii) the larger isoplanatic patch, in addition of course to the advantage of a clear observing window at all wavelengths. Such a filled-aperture instrument would be capable of forming object maps of high dynamic range in a very short time (minutes or hours) compared with a dilute array (hundreds of hours per object). The disadvantage is, of course, the reduced angular resolution compared with a dilute array ($D = 16$ m provides an angular resolution α of $\approx 7.5 \times 10^{-3}$ arcsec in the visible). This concept is not considered in detail here, but deserves further study. Its angular resolution is complementary to that provided by a very-long-baseline instrument, and in particular it provides excellent sampling of the lower angular frequencies.

A lunar interferometer is an ambitious and costly project that would require several stages of implementation. The nature of precursor mission(s) has been considered here only in outline form, pending a more detailed study.

Finally, even at this preliminary stage, it seems clear that two different interferometers would have to operate for the ultraviolet, visible and infrared on the one hand, and in the sub-millimetre range on the other.

2.2.3 Scientific goals of interferometry

In this section the scientific goals of optical and sub-millimetere interferometry are described. The feasibility of achieving the necessary resolution and sensitivity is not discussed. One can divide the scientific drivers for interferometry into two broad classes, astrometry and imaging, with somewhat different scientific requirements, and these will be discussed separately. Since sub-millimetre interferometry combines certain aspects of both classes, as well as addressing the scientific issues in a different way, this discipline will also be discussed separately.

It is of course impossible to cover all aspects of astronomy and astrophysics benefitting from interferometry, so certain fields in which there is a general consensus are presented here. A fuller description can be found in, for example, the Proceedings of the ESO and NOAO Conferences that were held in 1988, 1990 and 1991.

■ **Astrometry**

Astrometry of single objects

The impact on astrophysics cannot be over-emphasised. At a level of 50 μarcsec and a limiting magnitude of 15, it becomes possible to map the motions of the spiral arms of our Galaxy to a precision of better than 10% by observation of O and B stars, making it possible to separate the contribution of density waves to galactic motion. At a level of 10 μarcsec and a limiting magnitude of 18, one can determine the dynamics of the nearest open clusters, detect oscillations of Earth-like planets out to 20 parsec, determine the parallaxes of all Cepheids in our Galaxy with a precision of better than 10%, with a subsample of 50 Cepheids with precision better than 1%. By attempting a precision of $1-0.1$ μarcsec and reaching a sensitivity level of $V-mag=20-25$, the following goals are all achievable: (i) internal dynamics in the Magellanic Clouds, M31 and other local group galaxies; (ii) proper motions of high-z quasars and relative motion in quasar pairs; and (iii) parallaxes and proper motions of most galactic and globular clusters to a precision of better than 1%. It also becomes possible to determine these parameters for Cepheids and RR Lyrae stars, with significant precision, in the Magellanic Clouds and M31. The direct astrophysical goals thus include a significant improvement of the extragalactic distance scale, as well as a thorough mapping of the dynamics within our Galaxy, well-determined luminosities and masses for all types of astrophysical objects found within our Galaxy, as well as these parameters to a lesser precision for more exotic astrophysical objects at high z.

Astrometry of binary objects
For many types of stars, and for luminosity class-III in particular, only sketchy information about fundamental properties such as mass and radius is available. This means that it is impossible to determine the evolutionary phase, and thus to test our models for stellar evolution. The field of double and multiple stars is one area where higher angular resolution will provide a wealth of information on stellar physics. If orbital parameters for multiple stars can be determined with a precision an order of magnitude better than with the Hipparcos spacecraft, direct testing of the theoretical stellar models will be possible. Key issues to be addressed include assumptions of chemical composition, the presence of convective over-shooting and internal structure during the evolution of the main-sequence. Masses and radii for luminosity class-V can be obtained (particularly important in the lower half of the Hertzsprung-Russell diagram). The discovery of hitherto unseen companions to certain classes of stars, such as barium stars and blue stragglers, would definitely improve our understanding of these objects. Part of these goals can be realised with speckle interferometry carried out from the ground with 8 m-class telescopes, or with ground-based multi-telescope interferometry.

Astrometry and general relativity
It must be noted that at the $1-10$ μarcsec precision level, it becomes possible to detect different relativistic effects, such as frame dragging, relativistic precession and to perform light-deflection experiments to the second order. It thus becomes possible to use direct measurements as tests of relativistic metrics.

■ **Imaging**
Imaging of the surface and envelopes of stars
Many questions regarding the basic properties of normal main-sequence stars can be addressed when observations are collected at high resolution. The ultimate goal is to resolve granulation cells on the surface of other stars. Here is a field in which theory has proceeded rapidly during the last decade and where observations are needed in order to understand the deeper aspects of stellar evolution. To resolve the granulation patterns on even the nearest of stars demands a resolution of better than $5-10$ μarcsec together with an extremely high dynamic range, with consequently stringent demands on the parameters of the optical interferometer. However, supergranules probing deeper layers in the convection zone of stars are orders of magnitude larger and easier to resolve. Large convective cells, plages, stellar spots and flares are expected on the surface of many types of stars, and with many different characteristic sizes, from roughly 0.1 marcsec downward. One important parameter that determines observability is the contrast of these features relative to the rest of the stellar surface. These questions need further study, taking the relative technical capabilities into account.

Circumstellar envelopes of both hot and cold stars produced by stellar winds should also be studied at high resolution. The images of cold stellar envelopes, for example, will look different when imaged in and outside a molecular absorption band, or in an emission line such as the hydrogen lines or the NaD doublet, allowing a direct study of the vertical distribution of the various components. One can also study departures from spherical symmetry, which are believed to be quite common. For hot stars, imaging in resonance and other lines especially in the UV range will also allow study of the structure of the wind, but this is of lesser importance since these winds are much better understood than for cold stars. For all of these studies, a resolution of $0.1 - 1$ marcsec is already very good, since among the most interesting objects are giants and supergiants.

Imaging of accretion phenomena around single objects
The concept of accretion discs appears in the theories for many different types of such astronomical objects as black holes and other compact objects within the nuclei of active Galaxies (AGNs), as well as among binary field stars within our own Galaxy. In the field of early stellar evolution, accretion discs have been evoked to explain the transfer of angular momentum from the parent molecular cloud to the ambient medium, as well as being the mechanism by which the young stellar object and its possible planetary system is itself formed. So far, nobody has observed unequivocally an accretion disc. As for the different classes of objects, the size scales predicted for the discs are in the range of a few tenths of arcseconds (for a few close star formation regions such as Taurus or Rho Ophiuchus) to less than a microarcsecond for the central engine in Quasars and AGNs. As far as star formation is concerned, the discs themselves are probably deeply embedded within the parental cloud, hidden by up to hundreds of magnitudes of visual extinction, until the latest stages in the process. Thus they are only directly observable in the sub-millimetre wavelength range (see Section 2.3.4). In the sub-millimetre case there might be a problem of contrast, since the amount of material in the discs is supposedly only a small fraction of the total column density viewed by the detector. The expected luminosity peak in the central region might be enough, however, to separate the disc emission. There are other manifestations of the accretion process itself directly or indirectly observable in the ultraviolet, optical and near-infrared ranges. These phenomena include rifts in the parental cloud or smaller optical depths in the polar region of the discs, where the emission from the accretion shock and photosphere/chromosphere gives rise to reflection nebulosities. In the region in which it is believed that the accretion discs have their poles, ionised jets including shock phenomena of the Herbig-Haro type and associated molecular outflows, sometimes of high mass, are commonly observed. Other types of interactions also need to be studied at higher spatial resolution. There is mounting evidence that both on the stellar surface and further out, electric and magnetic phenomena play important roles. It is quite possible that plasma physics will enter this field when the observable scale sizes become small enough.

Imaging of interacting binaries
Binaries were mentioned earlier in the context of determination of basic stellar parameters such as mass and radius. New physical processes also take place, and are interesting on their own account. This is particularly true in the case of interacting binaries. This is another case where accretion discs play an important role and where observations of spatially very small structures are important. Interacting binaries belong to different classes, depending on the mass of the accreting star, such that a main-sequence object is of Algol type, a white dwarf is called a 'cataclysmic variable', whereas a neutron star or black-hole system goes under the name of an 'X-ray binary'. If the mass donor is a low-mass giant, the term 'symbiotic variable' is often used

regardless of the nature of the accreting star. The angular scales, taking true sizes and distances into account, are approximately 5×10^{-5} to 5×10^{-6} arcsec. The parameters that one would like to observe/determine include:

- distances: it is important to know the absolute size of the system
- the relative orbit: it provides the masses
- surface structure of stars and accretion discs, and
- optical observations of the radio jets detected by VLBI in X-ray binaries.

From observations of these parameters on the scales previously stated, one could determine the mass ratio and internal structure of the stars, as well as the viscosity of the material in the disc, and the tidal torque provided by the mass donor star. The size and form of the disc also vary with the mass-transfer rate through the disc.

The observational requirements can be stated as brightness in the range $V\approx17-20$, spatial resolution $10^{-3}-10^{-6}$ arcsec and a dynamic range of $100-1000$.

Imaging of active galaxies and quasars
In normal galaxies, the size scale of the nuclei is between 0.1 marcsec (M31) and 30 marcsec (giant ellipticals). In Active Galactic Nuclei (AGNs), the distribution of ionised gas and the interaction zone between the radio jet and this gas is within $1-0.01$ arcsec, whereas the accretion disc and Broad Line Region (BLR) is between 0.01 and 0.0001 arcsec in size.

The high intrinsic luminosity of quasars makes them interesting because they can be observed at very high z. Existing VLBI radio observations show a complex structure and super-luminal motion with an angular extent of a fraction of a milliarcsecond to a few milliarcseconds. These radio hot-spots may or may not be visible in the optical range, but the optical jets observed in M87 (giant elliptical active galaxy), 3C273 (apparent brightest quasar), and SS 433 (high-mass, relativistic binary object in our Galaxy) all suggest that at least some of the radio emission regions will be observable in the optical range. The spatial resolution needed will be a fraction of a milliarcsecond, as discerned from observations of the nuclear region of Centaurus-A and our own Galaxy. The optical spectra of quasars show broad emission lines. This inner broad-line region has never been imaged. Theories include accretion discs with bipolar jets. Different spectroscopic appearance could indicate different orientation of disc and jet relative to the observer. Imaging with high spatial resolution (less than a few tenths of a milliarcsecond) is necessary. The broad emission lines in quasars are rich in metals, even for objects at high red shifts. The chemical composition is approximately solar, which implies that there has been extensive and rapid stellar processing of the gas – presumably like the 'dusty star burst' discs observed in Seyfert nuclei. This disc would have a size of a few tenths of an arcsecond (a few kpc). The mechanism by which matter is transferred from disc to central engine is unknown and imaging on a scale of milliarcseconds will be needed in the ultraviolet and optical ranges in order to understand this phenomenon. The most interesting science would come from high-red-shift ($z\geq3$) objects, which are fainter than 17 th magnitude.

Imaging of gravitational lensing
Gravitational lenses observed with a spatial resolution of milli- or microarcseconds

would be expected to provide information on dark matter in the form of compact objects (e.g. black holes, brown dwarfs) in the Universe. It would also be possible to determine directly the size of the different emitting regions of quasars, AGNs, etc. and thus increase our understanding of mass distribution and the structure of the early Universe.

The observational requirements can be stated as brightness in the range $V \approx 17-20$, and spatial resolution of $10^{-3}-10^{-6}$ arcsec.

■ **Sub-millimetre interferometry**

The sub-millimetre range is dominated by the thermal emission of objects at temperatures between 3 and 100 K. The thermal emission from interstellar dust dominates the continuum, although one must also consider thermal emission from the surface of stars and planets and synchrotron emission from AGNs and quasars. The interstellar medium emits a very rich spectrum dominated by molecular lines and a few fine-structure lines in neutral regions including molecular clouds, and other fine-structure lines in ionised regions. All these lines offer crucial diagnostics of the physical conditions and chemistry in the interstellar medium, which nicely complement the information on dust given by the continuum. While the continuum is observed using preferably incoherent detection (direct interferometry), the lines are better observed using heterodyne techniques in the wavelength range 0.1 – 1 mm.

Solar System

Observations in the continuum and in bands of various molecules of the limbs of planets and satellites with atmospheres will greatly help in assessing their vertical distribution. This requires an angular resolution of better than 1 arcsec. In addition, high-resolution mapping of the surfaces in 100 – 300 μm emission bands of several geologically-abundant materials and volatiles such as NH_3 will be possible, even for small-diameter asteroids if the angular resolution is sufficient.

Circumstellar envelopes

Envelopes around very evolved stars emit a very rich sub-millimetre molecular spectrum. Some light molecules (hydrides) and molecules abundant in the terrestrial atmosphere can only be detected from space. Molecules such as HC_nN, SiO, OH or H_2O give maser emissions. The goal of space sub-millimetre missions will be to find new molecules and to investigate the physics of the excitation of many molecules, in particular for masers, by observation of a variety of transitions. An example concerning OH is the rotational triplet at 119 μm. These programmes will include observations of isotopic substitutions in order to clear up the optical-depth effects. The observations of those lines emitted by the lower envelope will benefit enormously from a high angular resolution, particularly for bipolar objects.

Interstellar matter and formation of stars and planetary systems

Recent millimetre observations have shown that interstellar clouds, and in particular molecular clouds, have an extremely clumpy structure down to the smallest scales accessible at present (about 1 arcsec for VLA observations in the lines of NH_3). This structure is presumably fractal and may be due to turbulence, but we do not know if (or where) the self-similarity breaks. It greatly affects the physics and chemistry of the medium. For example, recent observations with the KAO have been very surprising in that they have shown that the C^+ line at 157 μm coexists in molecular clouds with the neutral carbon lines at 370 and 610 μm and with a variety of lines of the CO molecule, the strongest of which are in the sub-millimetre range. This can only be interpreted as the existence of a clumpy, porous structure of the clouds which allow them to be permeated by UV radiation. It will be of the greatest interest to observe the small-scale structure directly in all of these lines. In general, we will probably understand well the physics and chemistry of interstellar clouds, of their interfaces with the ionised gas, and of interstellar shocks (e.g. as seen by the lines of H_2O) only when observations at very high linear resolution in the millimetre and sub-millimetre ranges are possible. Any angular resolution from a few arcseconds down to the minimum allowed by the sensitivity will be sufficient. Complementary observations of the continuum dust emission at the same resolution will also help tremendously; they also have to be carried out in the sub-millimetre domain, as dust is already optically thick at 100 μm or so.

Some of the interstellar clumps will eventually become stars after a very poorly understood collapse/fragmentation/coalescence phase. This phase can only be studied at sub-millimetre wavelengths because the optical thickness of the dust is too great in the infrared. Actually only stars born some time ago can be seen in the infrared range (*the shorter the wavelength, the older the star*). Observations in the continuum will provide the distribution of dust, and hence of matter in general, whereas observations in molecular lines are needed to measure the temperature and the velocity field. If one wishes to see anything but the grossest properties of star formation, an arcsecond to sub-arcsecond resolution is required even for the closest known star-formation regions such as the Taurus molecular clouds at 150 parsec. An interferometer is certainly required to study the structure of the proto-planetary discs believed to exist around newly-born stars, and in particular to detect the condensations that will eventually become giant planets. Although no such disc has yet been resolved with certainty, their size is believed to be smaller in general than 100 astronomical units, i.e. 0.6 arcsec at the distance of the Taurus clouds (the size of the remnant of such a disc around ß Pic is 400 AU). Another class of interesting and still puzzling features connected with star formation is the bipolar jets seen starting from very young stars: their origin can only be understood from observations at very high resolution. It is clear that detailed studies of the formation of stars and planetary systems will be a major objective of a large lunar sub-millimetre instrument.

Extragalactic astronomy

The sub-millimetre range is where the cooling of interstellar matter takes place: this is obvious for the dust, which is generally at temperatures between 10 and 30 K. For the diffuse gas and the molecular clouds, cooling is dominated by the 157 μm line of C^+, the lines of OI at 63 and 145μm and by the many lines of CO (the strongest of which are below 1 mm). For the ionised component, cooling takes place mainly via the O^{++} lines at 88 and 53 μm. Observation with a sub-millimetre interferometer of the emission of dust and of those lines will allow us to study in an extraordinarily high amount of detail the thermal balance and more generally the physics of the interstellar matter in nearby galaxies in relation to star formation: individual molecular clouds, interfaces and shocks will be resolved. It will also be possible to map the distribution of the very youngest stars which escape detection in the infrared for the reasons given above. In this case, angular resolution is a necessity in order to combat confusion. The regions of star formation in the nuclei of interacting galaxies, as well as the immediate environment of active galactic nuclei, all of which are regions often completely hidden by dust absorption even in the mid-infrared range, will also be selected objects for a high-resolution sub-millimetre instrument. It will also be able to map the residual dust that is supposed to give rise to the sub-millimetre emission bump seen in the spectra of a substantial fraction of AGNs and quasars. Even a resolution of a few arcseconds is of great interest, but it is likely that a much higher angular resolution will reveal very exciting phenomena. It is certainly too early to assess the possible contribution to cosmology of a large sub-millimetre instrument on the Moon, as the understanding of proto-galaxies and of young galaxies at high red shifts is still in its infancy: as for most other interferometry programmes, sub-millimetre observations are required with single antennas, which will be made with FIRST or from the Moon itself.

2.2.4 Identification of parameters

As a result of preliminary discussions, no definitive single concept for a lunar long-baseline interferometer has emerged. However, a number of design parameters are clear if the interferometer is to be capable of a significant step forward in imaging capability compared with Earth- or space-based systems. Some of these parameters are summarised in Table 2.1.

Table 2.1. Interferometer parameters

Wavelength Coverage	$0.1-100\ \mu m$
Maximum Baseline	$100\ m - >1\ km$
Spectral Resolution	$>10^4$
Number of Spectral Channels	$>10^3$
Measurement of Polarisation	$<1\%$
Dynamic Range	$>100{:}1$
Field-of-View (no. of res. elements)	$>100\times100$
Number of Telescopes	≥ 6
Sensitivity (visible)	$m_v \geq 23$
Sensitivity (sub-mm, lines)	1 K
Time to Record One Image	≤ 1 day

A key issue to be considered in more detail (for visible wavelengths) is the question of sensitivity, particularly the comparison of what should be achievable on the Moon with what *might* be achieved from Earth. Given the promise of adaptive optics, it has been assumed, for visible wavelengths, that partial wave-front correction can be implemented, if necessary using laser guide stars, giving an image point spread function in which the Strehl ratio is significant (≥ 0.4 at 500 nm, corresponding to ≥ 0.8 at $1\mu m$). In addition, it is assumed that images from separate telescopes can be referenced using a particular reference star (so-called 'Blind Operations Mode'). Given these assumptions, and allowing for the fact that an Earth-based interferometer may have large individual telescopes (≈ 8 m), Earth- and Moon-based interferometers yield similar signal-to-noise ratios for a measurement of the fringe visibility for brighter objects. A 'bright object' in this context is one which provides a sufficient number of detected photons per snapshot to estimate the fringe visibility in the total observing time (e.g. 1 h); for the Earth-based case, the exposure time of a snapshot is taken to be ≤ 1 s, whereas it is the full integration time (1 h) in the Moon-based case. For objects fainter than this, the estimate of fringe visibility deteriorates rapidly for the Earth-based interferometer, but more slowly for a Moon-based one, with the result that the Moon-based interferometer will be able to take measurements, albeit with a fairly low SNR, of fainter objects that cannot be measured from Earth. Figure 2.1 compares the Earth- and Moon-based cases for the parameters given below.

If the following parameters are used for a lunar-based interferometer:

- 0.1 system efficiency
- 3600 s integration time
- 100 nm band pass at V
- 0.05 value of power spectrum (resolved object)
- six-telescope array
- target SNR of 10

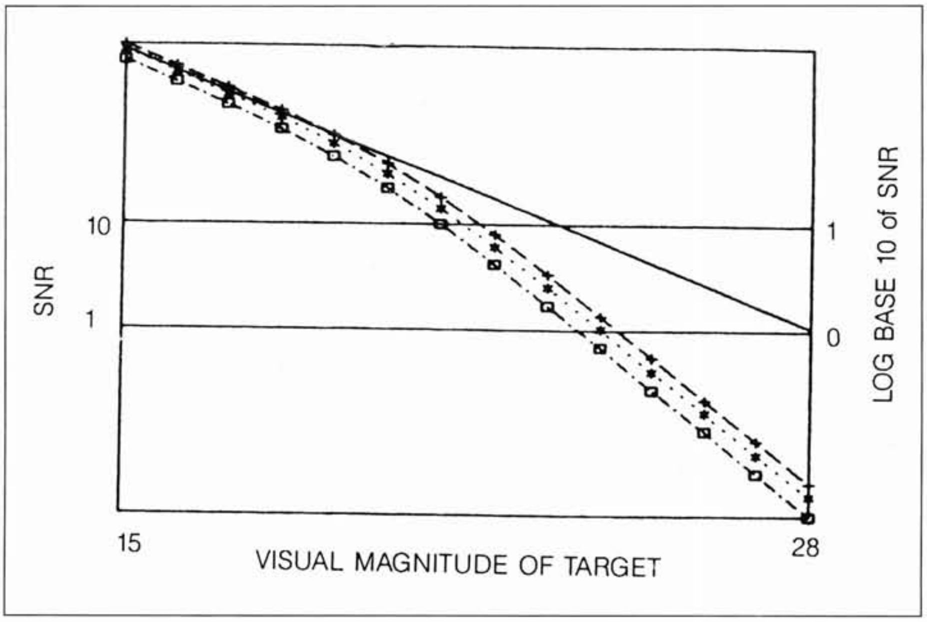


Figure 2.1 Signal-to-noise comparison for lunar and terrestrial interferometers

then the limiting magnitude is $V = +23$ for 2 m-diameter telescopes; this value is quoted in Table 2.1. It is necessary to carry out a more detailed study of sensitivity issues before firm estimates of the signal-to-noise ratio can be made.

2.2.5 Advantages of lunar interferometry

Table 2.2 summarises an initial comparison of some key issues for interferometry on Earth (using the VLT-I), in space, and on the Moon.

Table 2.2. Key issues for interferometry on Earth

Issue	Earth	Space	Moon	
			Equator	Pole
Atmospheric Effects				
Seeing: Diffraction-Limited Imaging	–	++	++	++
Seeing: Isoplanatic Path	–	++	++	++
Absorption/Emission	– –	++	++	++
Sensitivity	+	+	++	++
Environmental Effects				
Wind	–	+	+	+
Microseisms/Vibrations	–	?	+	+
Gravity	?	?	?	?
Dust	–	++	– –	–
Absence (Cyclic) Solar Heating	–	– –	– –	++
Passive, Radiation Cooling	–	+	+	++
Tidal Forces			–	+
Configuration Aspects				
Large (>1 km) Baselines	++	–	++	++
Declination δ Coverage	+	++	+	–
Extendability	++	–	++	++
Other				
Telescope Aperture	++	–	+/–	+/–
Capital Cost	++	–	– –	– –
Operational Cost/Maintenance	++	–	– –	– –
Main Advantages	Low cost Large diam. Long base	All λ Phased ops. No dust	All λ Phased ops. Long base	

A preliminary comparison led to the conclusions that in particular:

- The dominant advantage of any interferometer placed outside the Earth's atmosphere is the wavelength coverage. In the case of sub-millimetre wavelengths, this advantage alone justifies a space- or Moon-based interferometer.
- The Moon provides a much more stable base than is likely to be achieved in space. This base is expandable, so that one could start with a modest baseline and facility and allow it to grow in the light of experience and scientific goals.
- At visible and near-infrared wavelengths, where the atmospheric absorption is not such a major factor, the Moon has the advantage over the Earth that no special techniques are required to overcome atmospheric 'seeing'. The extent to which special Earth-based techniques such as adaptive optics and laser guide stars will be successful in twenty to forty years time is not clear at present although one may be optimistic in this respect.
- Several uncertainties regarding the lunar environment, such as the stability of the lunar surface layer, the amount of dust likely to be encountered and the feasibility of a polar site, need to be resolved.

2.3 Very-low-frequency astronomy

2.3.1 Introduction

This study is motivated by the fact that very-low-frequency (VLF) radio astronomy from Earth-based observatories ranges from impossible at frequencies less than 2 MHz (because of ionospheric absorption and reflection), to extremely difficult at frequencies of less than 100 MHz (because of man-made interference).

For the purposes of the study, a working definition of low frequency has been adopted as the range 30 MHz – 100 kHz (wavelength range 10 m – 3 km), the ionospheric cut-off frequency being typically 20 MHz by day and 10 MHz by night. However, even at frequencies as low as 2 MHz the ionosphere is sometimes partially transparent (Reber 1990), making terrestrial observations possible at such wavelengths from particular locations on Earth, e.g. Tasmania. Should metre-wavelength radio astronomy assume new importance in the future, efforts would be made to seek the best terrestrial sites before embarking on anything as expensive as a lunar observatory.

In the frequency range adopted, several different sources of radiation have been considered, including:

- the Galaxy: the galactic background and discrete galactic and extragalactic sources
- the planets: magnetospheric emission has been detected from Jupiter, Saturn, Uranus and Neptune
- the Earth's magnetosphere
- the Sun: solar bursts produced in the interplanetary medium at large distances from the Sun.

At the very lowest frequencies, a totally new study is being embarked upon, particularly for conventional radio astronomy. Thus an entirely new class of sources of, for instance, coherent emission which might have very steep spectra not extending to the higher frequencies, cannot be ruled out. Unexpected discoveries appear in all new wavelength regimes, and so why not at VLF?

In the course of this study, several recent documents addressing the same theme, or the general theme of low-frequency radio astronomy from space, have been identified. Where necessary, these documents have been summarised and referred to; no attempt has been made to repeat their arguments in detail. Instead an attempt has been made to highlight new areas where a lunar-based telescope could be usefully employed.

Finally, it should be added that an observatory on the far side of the Moon may ultimately become necessary, not only for VLF radio astronomy, but for all radio astronomy at decimetric and centimetric wavelengths unless the commercial users o those regions of the spectrum can be persuaded to comply more fully with the WARC regulations.

2.3.2 A VLF observatory on the Moon

The Moon has a number of natural advantages over the Earth as a site for a low-frequency radio telescope. First, it has no significant ionosphere, and so VLF radiation may reach the surface. Secondly, an observatory on the far side of the Moon will be well shielded from sources of interference from the Earth and will remain so, provided attempts are made from the outset to keep the local environment free from man-made interference.

Since a telescope is considered to be an array of antennas, the lunar surface has the advantage over, for instance, an orbiting array of telescopes, of providing a stable base.

It should also be noted, however, that the radiation environment of the Moon is far from benign. The absence of a substantial lunar magnetic field or atmosphere allows the solar wind, the magnetospheric plasma and cosmic radiation to penetrate directly to the lunar surface. Along its orbit, the Moon traverses the interplanetary medium and different regions of the Earth's magnetosphere. It is actually in the solar wind for about three-quarters of its orbit. In these different regions, the local electron density varies by large amounts – from less than 1 cm^{-3} in the geomagnetic cavity to more than 100 cm^{-3} in the magnetosheath. The background noise is enhanced, a broad line can appear near the local plasma frequency, and observations at frequencies greater than 100 kHz can be affected. The physics of an antenna in the vicinity of the lunar surface which is expected to be charged will need to be assessed, providing important parameters for the antenna design.

2.3.3 Limits to angular resolution at VLF set by interplanetary/interstellar scattering

The interplanetary medium causes strong scattering (rms phase modulation 1 rad) for a source outside the Solar System, in any direction, for frequencies below 15 MHz (Readhead et al. 1978). Source-blurring is then dominated by interplanetary scattering. Independent estimates of the magnitude of blurring obtained from interferometric observations at 38 MHz (Okoye & Hewish 1967), from IPS data (Readhead et al. 1978) and from spacecraft phase scintillation (Spangler & Armstrong 1990) are in reasonable agreement and lead to a value $\theta \approx 1°$ at an elongation of 90°

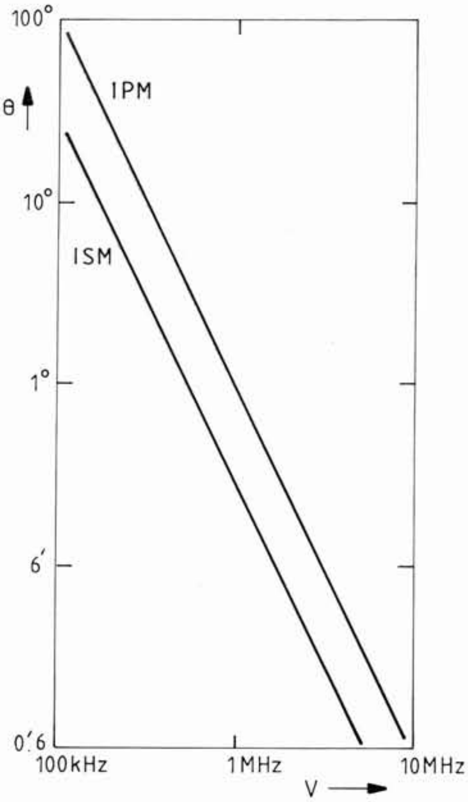


Figure 2.2 Magnitude of interstellar (ISM) and interplanetary (IPM) scattering in the 0.1 – 10 MHz range

at 1 MHz. This may be taken as a typical value of source blurring over the whole sky, including the anti-solar direction, for elongations exceeding 90°. Source blurring increases rapidly at smaller elongations. Day-to-day variations by a factor of 2 above and below this value are typically expected.

Interstellar scattering varies considerably for different lines of sight through the Galaxy. The best average value is probably that derived by Hajivassiliou (1991) from IPS observations. Except for directions close to the Galactic Centre, he obtained $\theta \approx 0.2°$ at 81.5 MHz, which scales to $\theta \approx 0.3°$ at 1 MHz. This agrees well with estimates by Cordes (1990) based upon pulsar scintillation data. The magnitudes of interstellar and interplanetary scattering for frequencies in the 0.1 – 10 MHz range are shown in Figure 2.2.

To calculate the blurring for radiation within the Solar System, such as kilometric wavelength radiation from Type II and Type III bursts, it is necessary to adopt a suitable model for the distribution of irregularities in plasma density along the line of sight. If $b(r)$ is the scattering power of the irregularities at a distance r from the Sun, and l is their physical scale, we have

$$\theta = \frac{\lambda\varphi}{l}$$

where

$$\varphi^2 \propto \int b\ (r)\ dz$$

Adopting the RKH model (Readhead et al. 1978) where $b(r) \propto r^{-4}$ and $l \propto r$, and normalising to the observed scattering at 90° as in Figure 2.2, we obtain for a source at infinity

$$\theta \approx (\nu/\mathrm{MHz})^{-2}\ (r/\mathrm{AU})^{-3}\ (\Delta z/\mathrm{AU})^{1/2}\ \mathrm{deg}$$

This formula assumes that b(r) and l are constant along a path Δz, where Δz is the equivalent length of the dominant portion of the scattering path according to the IPS weighting function given in RKH. This estimate is not sensitive to the assumed model and a Kolmogorov irregularity spectrum would give similar results.

A typical Type III burst radiates over frequencies ranging from 300 kHz at 0.1 AU to 30 kHz at 1 AU. Using the thin-screen approximation, which replaces the extended medium by a thin layer localised between the source and the observer, and making allowance for a geometrical factor depending on the ratio of the distances of the source and observer from the screen, we obtain the following estimates for a point source travelling directly towards the Earth:
- source at 0.1 AU, $\theta = 125°$
- source at 0.5 AU, $\theta = 370°$.

For such large values of theta, the source is of course not blurred out over the same angle. Under these conditions the angular diameter of the scattering disc is determined by the geometry of the scattering medium. The situation is roughly as shown in Figure 2.3, so that the point source is smeared over about 10° at 0.1 AU and 90° at

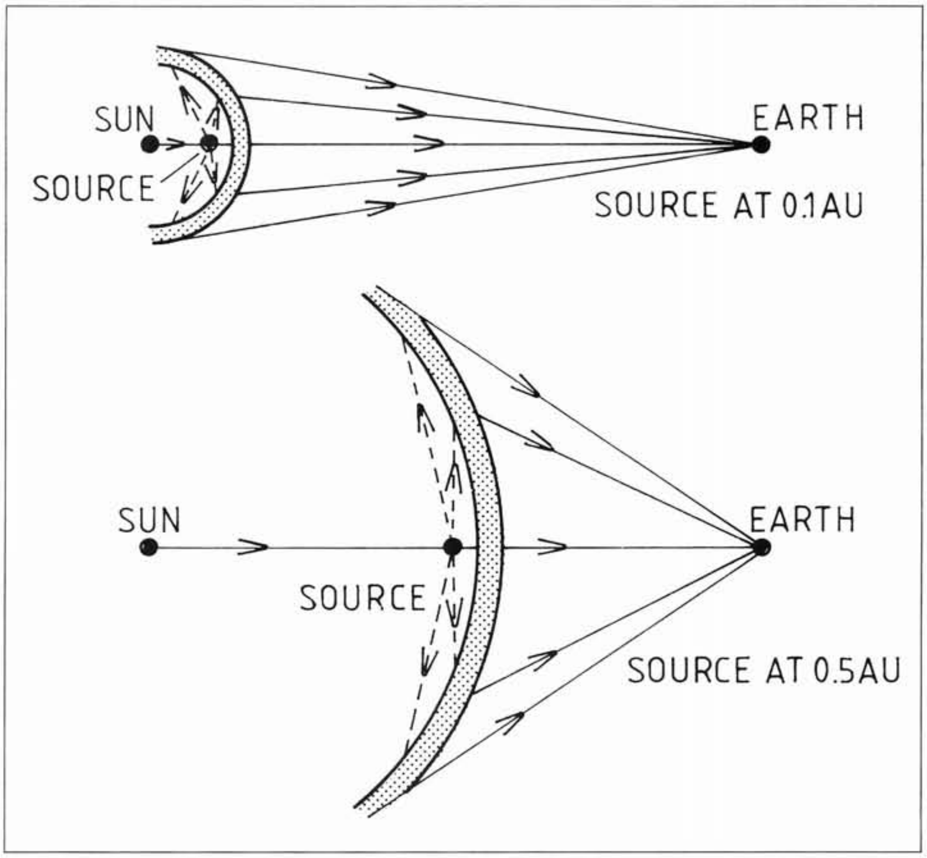


Figure 2.3 Scattering geometry for sources in the inner Solar System

0.5 AU. These estimates are in good agreement with observed values (Steinberg et al. 1984). When the source travels at an angle to the Sun−Earth line, scattering is greater due to the longer scattering path.

In addition, the strong variation in scattering with radial distance r will introduce a corresponding intensity gradient across the scattered source and the centroid of the distribution will be displaced from the true position of the source.

This simple analysis shows that scattering rules out the possibility of mapping and tracking solar kilometric sources with any confidence under normal conditions. When other complications are taken into account, the situation becomes even worse. Such factors include refraction and ducting in large-scale density structures, which must occur for sources radiating near the local plasma frequency, and gross variations in the scattering power transverse to the line of sight, which will normally exist even for sources moving directly towards the observer. These effects will also distort and displace the apparent source.

The remaining possibility is the mapping of auroral kilometric sources from a lunar orbit. In this case the typical path through the scattering plasma at 1 AU is 4×10^5 km $= 3 \times 10^{-3}$ AU. The frequency range is 45 to 150 kHz, giving values of $\theta \approx 27°$ to $2.5°$. Scattering alone does not rule out crude mapping, except at the lowest frequencies. Refraction and ducting now dominate and will produce substantial angular displacements. The best place for these observations is a near-Earth orbit, not a lunar site. These points should be borne in mind when reading the solar section.

2.3.4 Galactic and extra-galactic radio astronomy

As stated in the Introduction, a number of meetings on VLF astronomy from space have been held in the past few years. The Proceedings of these meetings contain much speculation about the astrophysical importance of the VLF regime. Some of the important areas of research are summarised here, but tempered with the results of the previous section on scattering.

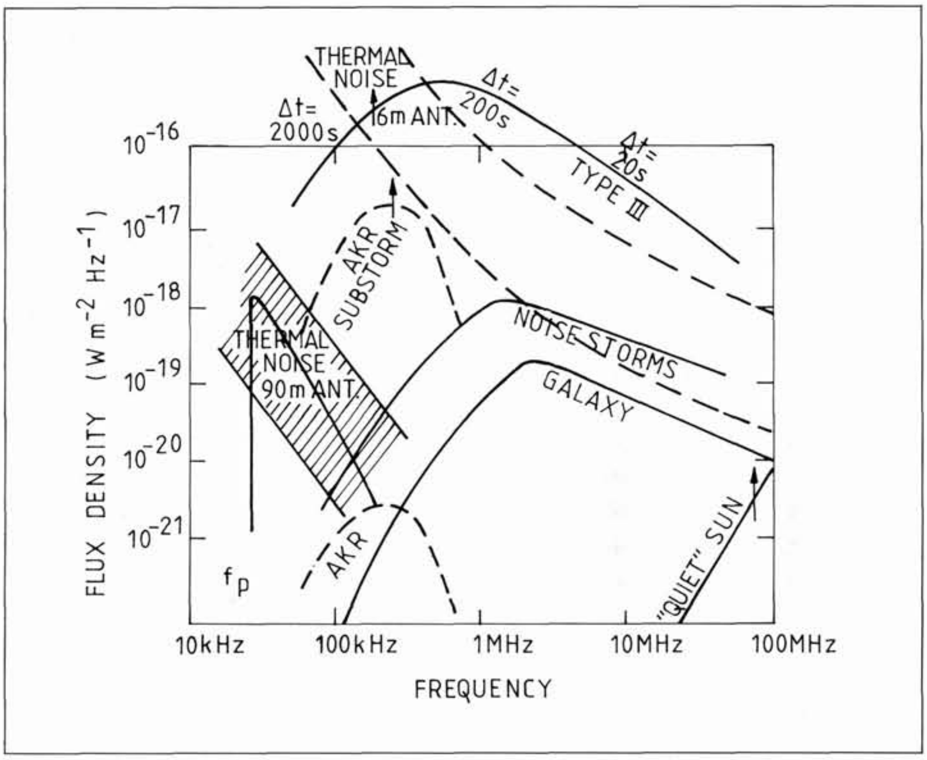


Figure 2.4 The flux density of a variety of sources in the 10 kHz to 100 MHz range, as measured by the 90 m tip-to-tip antenna on the ISEE-3 spacecraft

Radio astronomy at frequencies as low as 10 MHz is possible from the surface of the Earth. Ironically, however, in the context of this study, there has been little interest in this spectral region and even some of the instruments built to study low-frequency radio astronomy have been closed down because of lack of funding. This can only be seen as another manifestation of lack of community interest. Nevertheless, some stalwarts have persevered and the Cambridge group, under Dr. John Baldwin, recently performed a 38 MHz survey using a 5 km array (Rees 1991). Rees also speculates (1990) that such an array could be built for the 20 MHz band and suggests that Tasmania could be a possible site.

The scientific case for VLF radio astronomy will now be discussed. noting that the emission will be dominated by plasma effects and so these are examined first.

2.3.5 Plasma effects

Duric has discussed the effects of the mechanisms that modify source spectra at VLF. Such effects are both intrinsic to the source and due to the environment.

■ Free-free absorption

Plasma between the source and the observer can absorb radiation through free-free transitions. The optical depth of such a plasma is given by

$$\tau = T^{-1.35} \nu^{-2.1} \int n_e^2 \, dl$$

which is unity at a frequency

$$\nu \approx 0.3 T_e^{0.68} n_e l^{0.5} \text{ GHz}$$

where l is the path length in parsec and the other symbols have their customary significance.

For the Galaxy, we may assume $T_e = 10^4$ and $n_e \sim 0.1$ cm^{-3}, giving a turnover frequency of 1 MHz in lines of sight through the galactic plane and about 200 kHz perpendicular to the plane. Thus at 1 MHz extragalactic sources will be strongly attenuated when viewed through the plane. For standard HII regions with higher electron densities, even the galactic background emission will be attenuated at frequencies an order of magnitude greater than this.

■ **Suppression of radiation by a cold plasma**
According to Tsytovitch (1951) and Razin (1960), relativistic electrons embedded in a plasma will suffer suppression of their synchrotron radiation below a critical frequency given by

$$\nu_c \approx 20 n_e / B \text{ Hz}$$

For typical interstellar gas densities of 0.1 cm^{-3} and a transverse magnetic field $B \sim 10^{-6}$ G, $\nu_c \approx 10^5$ Hz. However, if the environments where particles radiate have higher electron densities or lower magnetic fields, this effect could be expected at higher frequencies. The effect could be searched for in galactic supernova remnants and in active galactic nuclei and would be recognised by the unique spectral shape. Measurement of ν_c and, for instance, an equi-partition calculation of B could yield the thermal gas density.

■ **Synchrotron self-absorption**
The importance of synchrotron self-absorption at low frequencies is that it is likely to affect virtually all extragalactic sources and not just those extremely compact sources that commonly show the effect at very high (GHz) frequencies. A source is opaque to its own synchrotron radiation when

$$\upsilon_s \approx 34 \, (S\nu_s \, / \, \vartheta^2) \, B^{\,0.2} \text{ MHz}$$

where B is in Gauss, ϑ is the source angular size in arcsec, and S is in Janskys.

At the higher frequencies, the self-absorbed spectrum rarely reaches the theoretical shape ($S \propto \vartheta^{2.5}$) because different components reach self-absorption at different frequencies. This dilution will be less at VLF, but must nevertheless be accounted for. Once the self-absorbed components have been identified, it may be necessary to measure their angular sizes at higher frequencies. When this is done the magnetic field may be derived.

2.3.6 Observing programmes at VLF
Based on the above discussion, some potentially interesting observing programmes for a VLF telescope or array can be proposed:

■ **A series of galactic surveys at frequencies between 30 MHz and 1 MHz**
At frequencies approaching 30 MHz, surveys concentrated on the plane would locate HII regions as absorption dips against the distributed galactic emission. An example of the importance of this work has been provided by Kassim (1990) in a description of studies with the Clark Lake array. As a simple example, imagine that the opaque HII region just fills the beam of the telescope, and then the intensity in that direction is the sum of the optically thick free-free emission from the HII region and the foreground emission from the Galaxy along that line of sight. Knowing the electron temperature of the HII region, from higher frequency observations and its distance, the mean synchrotron emissivity, E_n, in that direction can be derived.

E_n is particularly important to cosmic-ray physics because of the relationship between the cosmic rays and the distributed galactic synchrotron emission via their interaction with the galactic magnetic field. In particular, the synchrotron emission in the $1-30$ MHz region is related to cosmic rays in the 100 MeV$-$1 GeV energy range, and these cannot be measured directly near the Earth because of solar-wind modulation (Kassim 1990).

As well as having important applications in cosmic-ray physics, VLF surveys can be used to investigate the distribution of HII regions in the Galaxy, and they can also be used to resolve kinematic distance ambiguities because only foreground HII regions will appear in absorption.

Such surveys are also important at the lower frequencies and will be concentrated towards higher latitudes to measure any absorption due to the (poorly understood) warm ionised gas component first proposed from analysis of VLF observations with Reber's array (Hoyle & Ellis 1963). Such observations combined with optical observations of H_α at similar resolution (1°) could provide entirely new perspectives on the composition and structure of the interstellar medium and the distribution of synchrotron emissivity in the Galaxy (Reynolds 1990).

■ Observations of supernova remnants

Low-frequency observations of supernova remnants (SNRs) are of interest for studies of their intrinsic spectrum extended down in frequency and the information it can give about the emission mechanism. A peculiarity at low frequencies cited by Chevalier (1990) is the 38 MHz 'flare' observed in Cas A in the mid-1970s. Perhaps such events carry important clues for the understanding of radio emissions. Much more information is needed about phenomena such as this.

Absorption measurements are also important because of their bearing on the properties of the immediate neighbourhood of the remnants and the possible interaction of the SNR with its surroundings.

■ Are pulsar observations possible?

Time delays due to multipath propagation in the interstellar medium are of order

$$\delta t \approx \frac{z\theta^2}{c} \propto \nu^{-4}$$

For the least dispersed pulsars,

$$\delta t \approx 1 \text{ s at 10 MHz}$$
$$\delta t \approx 10^4 \text{ s at 1 MHz Cordes (1990)}$$

An independent check on these estimates can be made from the magnitude of interstellar scattering obtained by Hajivassiliou (1991). For a typical pulsar, this gives $\delta t \approx 3\times10^5$ s at 1 MHz. It is clear that pulsars could only be observed as continuous sources due to the large temporal smearing.

■ Refractive scintillation by the ISM

It has been suggested that VLF observations are important in the context of refractive scattering in the interstellar medium. However, at 1 MHz, interstellar scattering by small-scale irregularities in the ISM gives $\theta \approx 20'$. The time scale for refraction by large irregularities is $z\theta/v$, where v is the average speed of the interstellar wind. This gives a typical time scale of $10^3 - 10^4$ years for refractive intensity variations at 1 MHz, and so VLF is not useful for studying the ISM.

■ Low-frequency source surveys

Baldwin (1990) has addressed the topic of source surveys at low frequencies and whether many, or any new sources, may be expected to be seen. He concludes that any 'new' sources in a survey at, for instance, 3.8 MHz will be those weak sources seen in the present 38 MHz surveys with spectral indices greater than the mean, provided the sensitivity is similar at both frequencies. Of course, sources with dramatically steeper spectra might not be evident, but such a population would be very exciting. We already know of millisecond pulsars with very steep spectra; if another class of source with coherent emission were serendipitously discovered, a low-frequency observatory would be fully justified.

2.3.7 Scientific motivation for low-frequency studies of the planets

■ Planetary low-frequency radiations: overall properties

The Voyager interplanetary mission (Stone 1977) has demonstrated that all the highly-magnetised planets in the Solar System produce powerful, low-frequency radio emissions. They are the four giant planets Jupiter, Saturn, Uranus and Neptune, and the Earth. These radiations have several very specific properties in common:

- Spin modulation: they are strongly modulated by the planetary rotation, with different kinds of planetary longitude or local time dependence. The analysis of the modulation provides access to the rotation rate of planet's interior, as the radiation is beamed in a fixed direction with respect to the magnetic field.
- Polarisation: they are highly (100%) polarised, mainly circularly, with a sense consistent with an X-mode emission from both northern and southern hemispheres.
- Spectrum: their frequency ranges are different from one planet to another, reflecting the different planetary magnetic-field intensities. Their average power varies from 0.1 (Neptune) to 1000 times (Jupiter) that of the Earth (10^8 W). This represents only a part ($10^{-5} - 10^{-6}$) of the incident solar wind power.
- Fine structures: they display a rich morphology of fine structures in frequency and time, with instantaneous bandwidth as low as a few Hertz (measured in the case of the kilometric emission of the Earth) and time scales ranging from a fraction of a millisecond (short bursts of Jupiter, Uranus and Neptune emissions) to several hours.
- Beaming: when measured, the instantaneous beam was found in all cases to be narrow, and at a wide angle from the magnetic field in the source region (hollow cone). When averaged in time, it fills a much broader emission lobe.
- Brightness: their inferred brightness temperatures are higher than 10^{14} K (even assuming the source to be as large as the planetary disc) and may exceed 10^{21} K (Jovian decametric emission).
- Solar-wind relationship: several close relationships with solar-wind variations were demonstrated: the terrestrial emission is primarily associated with the magnetospheric substorm activity (Benson & Akasofu 1984); the Jovian kilometric and hectometric radio components are somewhat correlated with the solar-wind

velocity (Barrow 1985); in the case of Saturn, the emitted power is strikingly related to the pressure of the solar wind incident on its magnetosphere (Kaiser et al. 1984).

Owing to the observing limitations (e.g. direct measurements of the plasma parameters in the source are available only in the case of the Earth), all these properties have not been determined directly, nor independently for each planet.

Nevertheless, a common scenario at the origin of the planetary radio emissions, with variations depending on the specific magnetospheric context, can be drawn. These emissions originate from key regions of the planetary magnetospheres, near the boundary between closed and open magnetic-field lines; the radio sources are spread along active 'auroral' magnetic-field lines and the radiation is emitted close to the local electron gyro-frequency by conversion of free energy contained in unstable populations of energetic electrons (with characteristic energies of a few keV).

■ Study of the planetary low-frequency emissions from the Moon

In order to model the planetary radiations, a non-thermal, coherent generation process, the 'cyclotron maser' mechanism (Wu & Lee 1979; Le Quéau 1988) was developed during the last decade. It appears to possibly account for the production of electromagnetic radiation in a number of other astrophysical situations, where low-density, highly magnetised plasmas are to be considered (solar and stellar magnetospheres) (Winglee et al. 1986).

A number of basic characteristics of planetary radiations are still not understood, including the following:

- Where precisely are the radio sources (unanswered for Uranus and Neptune)?
- How exactly are the emissions modulated by the rotation (especially in the case of Saturn, since the radio source is fixed in local time)?
- How big are the active regions?
- What determines the amount of emitted power?
- What is the nature of the interaction with the solar wind?
- Can we really compare the Saturnian case with a 'simple' magnetic field, and the Uranian or Neptunian cases, involving a very complex multipolar magnetic field?
- Why is Jupiter's decametric emission elliptically polarised?
- Are the short- and the long-duration planetary bursts produced by the same mechanism?

In the next two decades, several planetary missions (Ulysses in 1992, Galileo in 1995–1996, Cassini in 2002–2006) will revisit Jupiter and Saturn, and probably answer several of these remaining questions. Spacecraft orbiting the Earth will re-observe the terrestrial radiation in situ.

Nevertheless, it will certainly be valuable, especially at frequencies below the ionospheric cut-off not observable from the ground, to continue to study the planetary radiations themselves.

■ Applications

The following examples show that a good understanding of the generation of radio emissions should provide a very powerful tool for the remote sensing of the magnetospheres from which they originate.

Remote sensing of the planetary auroral regions and magnetospheric activity

A theoretical model computation of the spectrum of the Saturnian emission was carried out (Galopeau et al. 1989) which showed that realistic results could be achieved by using a model with only one free parameter, the electron characteristic perpendicular energy. Identification of spectral drifting features in the dynamic spectrum of the Saturnian emission might thus serve to monitor the variations of the hot electron distribution in the source region (Zarka 1991).

Planetary magnetic-field structure

Most of the characteristics of the planetary radiations vary with the planetary spin, in particular the maximum emission frequency. This is also the case for Saturn despite its very regular magnetic field as measured by the magnetometers aboard Pioneer and Voyager spacecraft (Connerney et al. 1984). On the other hand, progress in theoretical modelling in principle allows the quantitative description of the emission spectrum (achieved in the case of Saturn; Galopeau et al. 1989). This calculation shows that the high part of the spectrum is steeply cut off for certain critical values of the plasma parameters in the source region, including the gyro-frequency, which is proportional to the intensity of the ambient magnetic field. This was used (Galopeau et al. 1991) for computing a correction to the existing Saturnian magnetic-field model, not detectable by the magnetometers.

Detailed description of the structure of the magnetic field very close to the planet's surface was thus demonstrated to be possible by using remote radio observations.

Planetary magnetic-field monitoring

The sidereal rotation periods of the giant planets were all determined by the analysis of the modulation of the radio emission (Kaiser & Desch 1984). A long-term monitoring of the modulation might reveal secular variations in the magnetic field and in the planetary dynamo regime.

Solar-wind variations

By using the close relationship between the ram pressure of the solar wind at the distance of the planet and the observed power of its radio emission, one might simultaneously achieve a multi-point measurement of some solar-wind parameters in the $1-10$ (or 20) AU range.

Planetary ionospheres

Voyager spacecraft detected thunderstorm lightning in the equatorial atmospheres of Saturn and Uranus (Zarka & Pedersen 1986). The low-frequency cut-off of the observed events by the planetary ionosphere allows indirect measurement of the maximum ionospheric electron density and its diurnal variations.

Search for planets in other stellar systems

In the range of the low frequencies, the planetary radiations appear to be quite different from intrinsic stellar emissions. This contrast is further enhanced by the spectral signature of these emissions, their high degree of polarisation, and their modulation by the planetary spin.

Despite the large interstellar distances, an electromagnetic radiation as intense as that produced by a planet such as Jupiter, in orbit around a nearby star, might be detectable with a large Moon-based radio telescope (Lecacheux 1990). With a 10×10 km antenna, the emission of Jupiter, located at 1 parsec from the Earth, might be readily detected in the $3-30$ MHz frequency range, with a signal-to-noise ratio of 5 and a detection bandwidth of 1 MHz, in less than 100 s.

On the other hand, a very intense Jupiter-like planetary radiation might be very unusual, leading to too optimistic an estimate of the probability of such a detection. It is worth noting that this method does not favour the large bodies, but could also permit the detection of Earth-like magnetised planets.

■ Instrument requirements

In order to achieve all of these objectives, a basically broadband instrument is needed covering the $0.1-10$ MHz frequency range (3 km to 30 m wavelength range), with broad instantaneous bandwidth (typically $\delta f/f \approx 1$) and high spectral-resolution ($\delta f/f \approx 10^{-2}$) capabilities. A polarisation-measurement capability is also required.

Table 2.3. Antenna temperature contributions of planetary radiations compared with sky temperature for a 10^6 m^2 collecting area

Source	Flux at 1 AU	Flux	T_A
Solar Type III	4×10^{-16}	4×10^{-16}	—
Earth TKR	5×10^{-21}	8×10^{-16}	—
Jupiter DAM	1×10^{-18}	6×10^{-20}	2×20^9
Saturn SKR	2×10^{-19}	2×10^{-21}	7×10^7
Saturn SED	7×10^{-22}	8×10^{-24}	3×10^5
Uranus UKR	5×10^{-23}	1×10^{-25}	4×10^3
Uranus UED	1×10^{-22}	3×10^{-25}	1×10^4
Neptune NKR	2×10^{-23}	2×10^{-28}	7×10^2
Sky Background	—	—	1×10^7

Table 2.4. Expected signal-to-noise ratios of different Solar System radio emissions for different antenna collecting areas

Source	$A_e=10^3$ m^2	$A_e=10^6$ m^2
Jupiter KOM	—	—
Jupiter HOM	S/N = 20	S/N = 2×10^4
Jupiter DAM	—	—
Saturn SKR	S/N = 0.7	S/N = 7×10^2
Saturn Lightnings	S/N = 3×10^{-3}	S/N = 3
Uranus UKR	—	S/N = 0.04
Uranus Lightnings	—	S/N = 0.1
Neptune NKR	—	S/N = 7×10^{-3}

The Decametre Array in Nançay, France (Boischot et al. 1980), which is devoted to daily monitoring of Jovian decametric activity, is an example of a large, low-cost, planetary-dedicated radio telescope. Its 144 helix spiral elementary antennas fill an aperture of 10^4 m^2. The effective area is 3000 m^2 at 20 MHz, in each sense of circular polarisation. The instantaneous bandwidth is 10−40 MHz and the useful tracking time is 8 h per day. Several specialised, digitised spectrographs (broadband coverage, high time and frequency resolution analyses) are operated simultaneously.

The physical size of the radio telescope is determined by the desired sensitivity. Taking into account the required time−bandwidth product for efficient spectral analysis (namely $\tau = 1$ s and $b = 10$ kHz), the detection threshold above the galactic background is given by the 'effective aperture' A_e of the telescope, which is nearly the physical aperture in the case of a filled-aperture, phased antenna array.

Man-made noise from the ground, and the enormous signal emanating from the Earth's natural radiation (most of the time), require that an instrument be located on the quiet, far side of the Moon. Nevertheless, the solar-burst activity (10% of the time during the day) must be coped with, when the Sun itself is not the observing target.

In any case, the large amount of real-time data and the need to reduce the interference level caused by unwanted signals (from the sky, or locally from the lunar base), require powerful, real-time data processing.

2.3.8 Scientific motivation for solar-terrestrial studies

■ Solar radio sources at low frequency

Radio sources of solar origin are the dominant feature of radio emissions at low frequencies (Bougeret, 1990). Solar radio sources are sporadic and very frequent during the maximum phase of the solar cycle; they may be very intense (more than 40 dB above the background) and they can be localised almost anywhere in the interplanetary medium. They represent a unique means of detecting perturbations originating from the solar activity and of tracking their progression through the interplanetary medium and particularly into the Earth's environment.

■ The heliosphere

The heliosphere represents the solar corona and the interplanetary medium − the solar wind. The heliosphere at large is the medium that surrounds the Sun and is contained within the magnetic cavity created by the Sun's magnetic field and delimited by the hypothetical bow shock created by the interaction of the solar wind with the interstellar medium. This medium represents a unique laboratory, offering more extreme conditions than can be simulated on Earth. The solar wind, a steady flow of charged and neutral particles, interacts with bodies in the interplanetary medium, namely the Earth, the Moon, planets, and comets. It is traversed by energetic particles accelerated in the lower layers of the solar atmosphere; it conveys or supports disturbances that are related to the solar activity. This medium interacts in many ways with the terrestrial environment; many phenomena in the biosphere eventually depend on these interactions. A global understanding of the cause and effect relationships between the physical processes that link different regions of the Sun−Earth dynamic environment is one of the primary aims in solar-terrestrial physics.

■ Radio physics of the Sun

At low frequencies, solar radio sources are essentially sporadic and highly variable. At high frequencies (e.g. 200 MHz), a full burst may last less than 1 s, while the duration of a typical single burst is 2000 s at 100 kHz. Near the phase of solar maximum, some bursts may be very frequent and occur in 'storms'. These storms may be present for long periods (weeks or months) and determine the level of background noise for other observations. The intensity of the solar sources can be thousands of times that of the background noise (essentially the Galaxy). The frequency of radio emission depends strongly on the distance from the Sun and the source range increases with the distance from the Sun, and can be several million kilometres. Radiation is produced by the non-thermal mechanisms of plasma radiation and gyro-synchrotron radiation. Above about 10 MHz, radiation from the quiet solar corona and from coronal condensation can be observed. It is produced by the thermal mechanism called Bremsstrahlung.

Contributions and limitations of the radio technique

The solar atmosphere – solar corona and solar wind – is a unique plasma laboratory. The corona is a fully ionised, magnetised plasma at one or two million degrees Kelvin and which can be highly turbulent.

Radio observations yield information about transient phenomena and structures:

(i) Transient phenomena
- Acceleration and propagation of relativistic particles. Observations at high frequencies (above 100 MHz) can yield information very close to the acceleration source. At low frequencies, the observations trace the propagation of the particles through the interplanetary medium up to the Earth's orbit and beyond.
- Shock waves and more generally magneto-hydrodynamic (MHD) waves. These are frequently associated with Coronal Mass Ejections (CMEs). The low-frequency observations enable the tracking of these waves between the upper layers of the solar corona out to 1 AU and sometimes beyond.

(ii) Coronal and interplanetary structures
- The quiet Sun: its appearance changes from frequency to frequency. Its diameter increases with frequency and is hardly detectable at frequencies below a few tens of MHz because the contrast with the bright galactic background becomes less and less. Hot, bright regions can be observed, superimposed on the quiet Sun. At frequencies greater than about 10 MHz, they are associated with active regions. At decametric wavelengths, they are associated with coronal condensations lying above filaments.
- Coronal holes can be observed. They sometimes appear as bright regions, and at other times as depressions in the brightness, depending on the optical thickness of the source along the line of sight to the observer.
- Magnetic-field arch systems sometimes appear as bright regions. At other times large-scale magnetic field arches can be traced by brief radio emissions produced by sub-relativistic electrons that propagate along them.
- Streamers can also be visualised by means of the radio trace left along them by sub-relativistic electrons accelerated near their root on the Sun.
- Long-lived radio emissions inside coronal structures can yield information about the characteristic modes of these structures.

Radiation transfer

Propagation effects that occur between the source region and the observer can be important. The physics at play is that of electromagnetic waves in a magnetised plasma. It has several aspects:

(i) In the source region, the dispersion equation has to be solved. Turbulence on a small scale can have dominant effects. The local conditions are usually unknown and critical hypotheses often have to be made. Major progress could be made by in-situ measurement of local conditions using a space probe and the simultaneous observation of the radio emission.

(ii) During the flight of the electromagnetic waves from the source to the observer, the radiation can be refracted and scattered. This can affect the source size and structure, its apparent location and its polarisation. These effects can depend upon the position of the observer relative to the source and its location with reference to the Sun, because the observer is inside the medium where the radiation is produced.

(iii) The terrestrial ionosphere can have dominant effects (refraction and distortion) at decametric wavelengths and sizeable effects at metric wavelengths. Computer codes help to model and sometimes correct for the propagation effects. The propagation effects, in turn, can be used to deduce properties of the medium that is traversed by the electromagnetic waves.

The frequency/distance scale

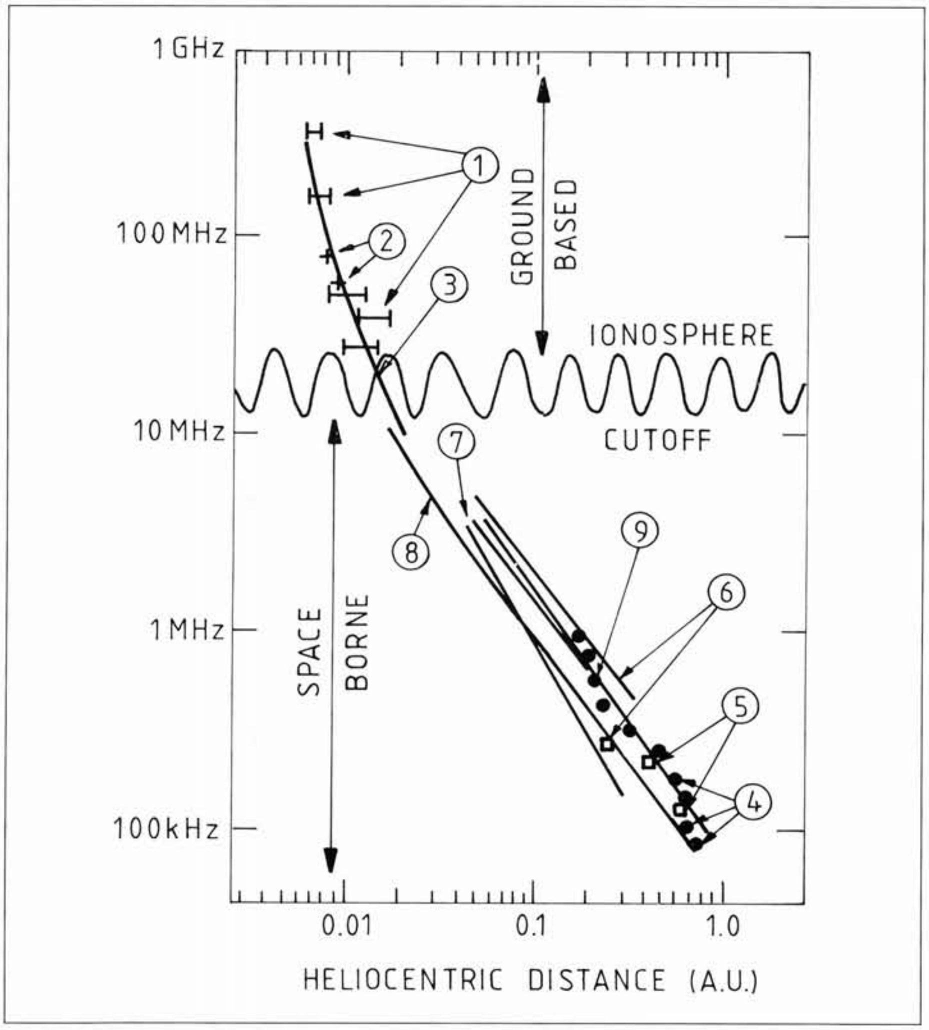
Figure 2.5 shows the distribution of the observed radio frequency as a function of distance from the Sun. This was obtained using different techniques and methods. The frequency of the radio emission depends strongly on the distance from the Sun: the lower the frequency of the radiation, the further from the Sun's surface it is emitted, for two reasons:

- The density N_e and thus the plasma frequency $f_p \approx 9\sqrt{N_e}$ decrease with altitude.
- Most radiation mechanisms produce radiation at or near the plasma frequency and its harmonic. This dependence is essential to solar radio astronomy and is of very special interest for spectral observations. Solar radio spectral observations not only provide information about physical processes, but also a distance scale for the phenomena being observed. Figure 2.5 illustrates this dependence, each frequency band being characteristic of a given distance range from the Sun.

Spaceborne observations at low frequency

Observations at low frequencies have to be made from above the terrestrial ionosphere. They have given rise to new observational techniques. Several instruments have been launched since the early days of space research and a radio package is almost always part of the major mission payloads (Alouette, IMP, Helios, RAE, Voyager, ISEE, and for future programmes: Ulysses, Geotail, Wind, Polar, Cluster, CRAF and Cassini). Usually the instruments are spectrographs. There have been a few attempts to carry out VLBI observations between two spacecraft, for instance to study the Auroral Kilometric Radiation (AKR) from the Earth's polar arcs. There are plans to use the four Cluster spacecraft to analyse the two-dimensional structure of terrestrial sources. There is, however, a big observational gap from 20 MHz to 1 MHz. This is especially worrying because it corresponds precisely to the range of distances from the Sun where the solar wind is accelerated – a range where several important radio phenomena also take place. This is actually a difficult domain to explore: it is close to the ionospheric

Figure 2.5 Distribution of observed radio frequency as a function of distance from the Sun



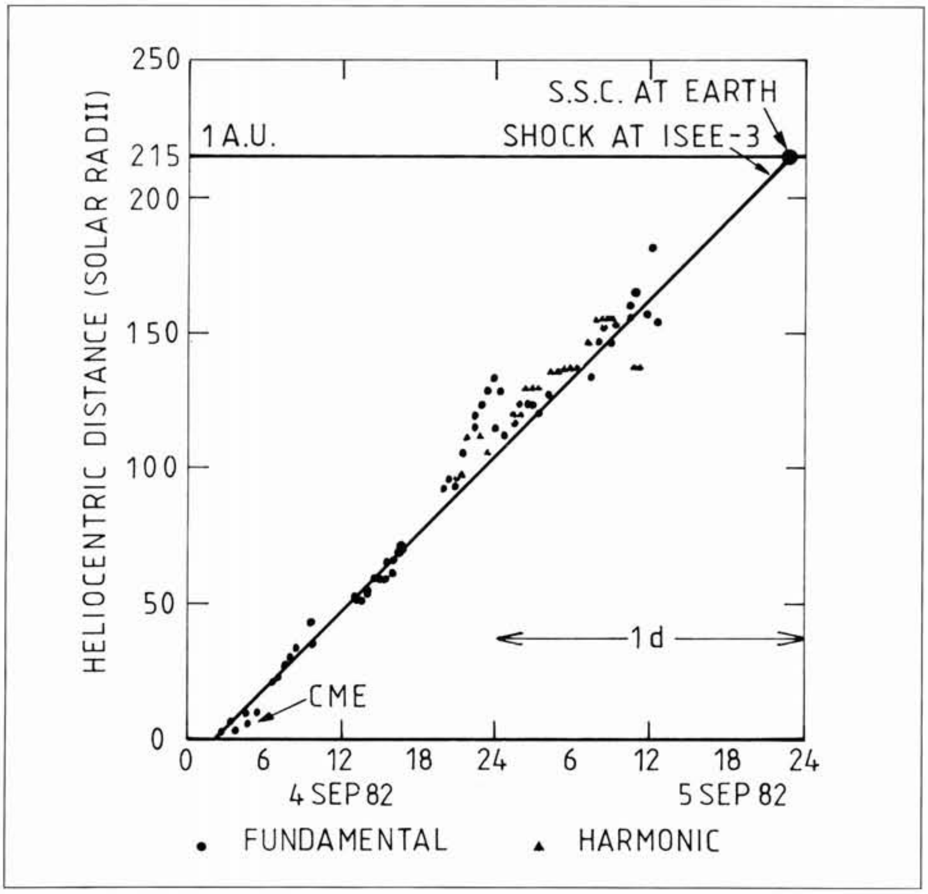


Figure 2.6 Progression of a large shock wave from the Sun to the Earth

cut-off and the observer has to be at a great distance from the Earth to reduce the level of terrestrial interferences. This unobserved range will be covered by an instrument on the Wind spacecraft. Some instruments have used the spin of the spacecraft to obtain the source direction from the modulation of the intensity during a rotation. The source size can also be estimated. A technique that combines the observations from dipoles in the spin plane and along the spin axis also yields the degree of circular polarisation. This technique will be used for the first time on Ulysses, and it will also be used on the Wind spacecraft.

■ Highlights of solar-terrestrial physics at low frequencies

Shock waves in the heliosphere

Collision shocks are interesting and important for numerous reasons. They are the simplest configuration in which a macroscopic flow is regulated by microscopic dissipation, a problem common to many plasma processes. Collisionless shock waves are therefore of basic plasma-physical interest. Coronal Mass Ejections (CMEs) are frequently associated with the major shock waves. A detailed analysis of the association of the radio emission with the mass ejection yields information on the shock formation and evolution.

The observation of the so-called Type-II radio events as tracers of shock waves is one of the major successes of solar radio astronomy: the Type-II burst is often the only indicator of a shock in the corona.

Shock waves in the heliosphere are among the most dramatic and energetic phenomena of solar origin; they may have extensive effects in the heliosphere. They produce the geomagnetic Storm Sudden Commencements (SSCs), the proton showers, some aurorae, and a wide variety of phenomena at the Earth.

The observation of Type-II radiation has a key role because it proves the existence of a shock; it is believed that a shock is needed to accelerate electrons to about 10 keV, where distribution becomes unstable to the production of Langmuir waves which are converted into radio waves. Thus, Type-II emission provides a unique means of tracking shocks remotely.

One usually distinguishes between two classes of shocks which exhibit different characteristics:

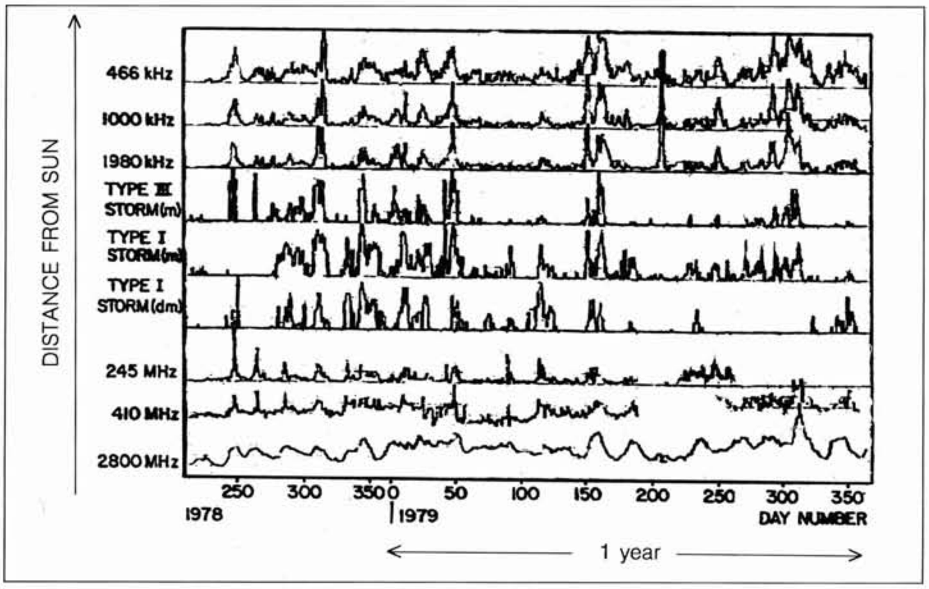
- The 'blast waves' are flare-related, starting at high frequency in the low corona (typically 250 MHz or 1.1 $R_\odot$). They usually die out at $2-3$ $R_\odot$ and can hardly be observed at low frequencies.
- The 'piston-driven' shocks are usually part of the 'big flare syndrome', being 'pushed' by large CMEs, and they start typically at $2-3$ $R_\odot$.

An example is given in Figure 2.6, which shows the progression of a large shock wave from the Sun to the Earth.

Large-scale co-rotating structures

Figure 2.7 shows the history of solar radio emission at a variety of frequencies. The flux density is well correlated from the transition region up to at least 0.5 AU.

Figure 2.7 History of solar radio emission at several frequencies



The successive peaks describe radio 'storms' that may last up to 10 d or more, i.e. approximately half a solar rotation.

They correspond to an almost continuous radiation emitted from a structure rooted on the Sun, near active regions. The combination of the solar-wind expansion with the solar rotation gives the structure the classical shape of the Archimedean spiral. Such structures can be seen at radio frequencies rotating from east to west as the Sun rotates. For instance, it can be predicted when such an electron stream is going to hit a given spacecraft or a terrestrial environment. All this information directly yields the velocity of the solar wind at different distances from the Sun and allows the interplanetary magnetic-fields structures associated with active regions and slow solar-wind streams to be mapped.

Mapping of the interplanetary magnetic field

Using the direction-finding capability of the radio instrument on ISEE-3 and a density model, it has been possible to derive the trajectories of thousands of individual electron events (Type-III bursts). Figure 2.8 gives a view from the Earth of all these trajectories which trace magnetic-field lines. This remote-sensing capability provides a unique means of analysing the structure of the interplanetary medium. Large deviations from the shape of an Archimedean spiral are frequently observed; some of these are caused by propagation effects (or proximity effects) at the lower frequencies; others can be attributed to variations in the solar-wind velocity along the electron-beam trajectory. Such deviations from the simple Parker model are generally due to the inhomogeneous conditions near the Sun which give rise to plasma streams of different velocities.

Fast and slow streams will collide, and at the boundary between them a shock wave will build up and large gradients will appear, favouring the appearance of instabilities of various kinds. Cut-offs and discontinuities are actually observed in Type-III burst trajectories, which can yield the location and magnitude of the perturbations.

Prediction/monitoring of solar flares and radiation effects

Prediction and monitoring of solar flares are important (vital) in the context of the lunar base because of the potential hazards they present to man and equipment on the Moon (Hildner 1990). Low-frequency observations, and particularly imaging, can provide unique information on the evolution of major disturbances in the interplanetary medium. A better knowledge of the structure of the interplanetary medium, which can be gained by large-scale mapping, would also improve our understanding of the propagation of streams of energetic particles through the interplanetary medium.

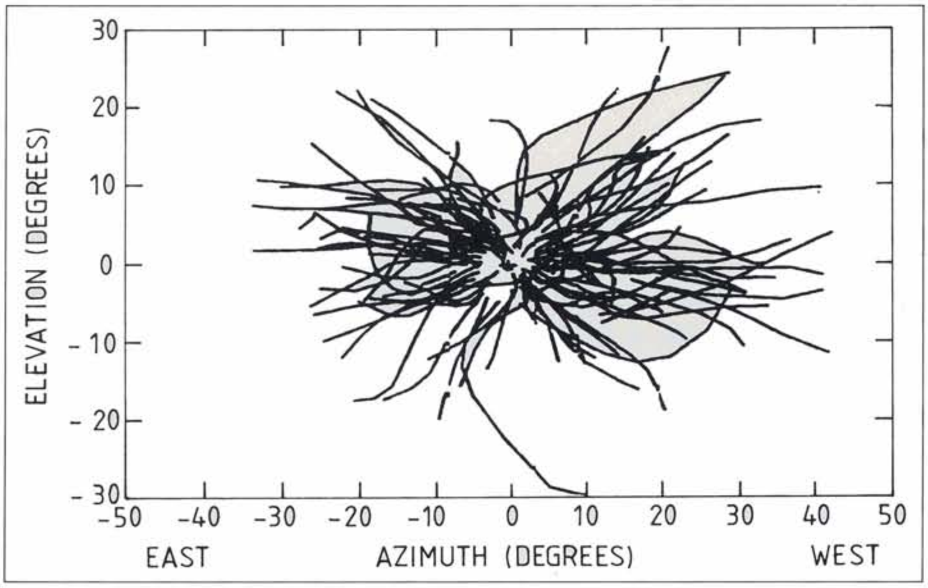


Figure 2.8 Trajectories of individual electron events as measured by the ISEE-3 satellite, and viewed from Earth

■ The geospace

The Sun is a variable star. It has an extended atmosphere – the solar corona – which generates the 'solar wind'. The interaction of the structured, time-varying solar wind with the Earth's magnetic field creates a whole range of effects – aurorae, geomagnetic storms, disruptions in short-wave radio communications, power surges in long transmission lines – which are collectively called solar-terrestrial phenomena, and which occur in a region known as 'geospace'. This environment is shown in Figure 2.9.

The Earth emits two different kinds of low-frequency radio radiation. The first is the non-thermal continuum, or NTC (Gurnett 1974); it is slowly varying and its spectral index is negative. It is observable in the tail down to the lobe's plasma frequency of 5 kHz. At frequencies lower than the plasma frequency of the magnetosheath or of the bow shock, the NTC cannot escape into the interplanetary (IP) medium and remains trapped in the tail cavity. At higher frequencies, it escapes into the IP medium. The second type of radiation is the Auroral Kilometric Radiation, or AKR (Gurnett 1975), which is much more intense than the NTC; it is very spiky and its spectrum has a positive spectral index. It is sometimes observed in the tail down to 30 kHz.

Most of the NTC is radiated at the plasmapause, i.e. about 4 R_E from Earth. AKR is emitted by very small sources in the auroral regions at geocentric distances not larger than again 4 R_E. These regions were determined from low-orbiting spacecraft, some of which crossed the source region.

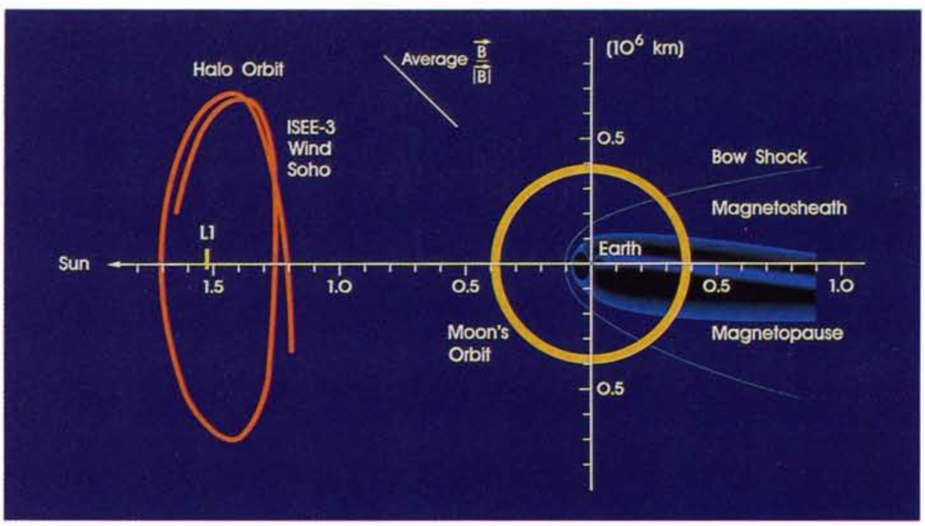


Figure 2.9 Interaction between the solar wind and the Earth's environment, as viewed from the north ecliptic pole

When observed onboard a properly instrumented spacecraft flying near the lunar orbit or farther from Earth, a number of propagation phenomena are observed: for instance, seen from within the solar wind in the 03:00 to 06:00 local time sector, the AKR source can be seen at 56 kHz as far as 10 to 20 R_E tailward from Earth (Steinberg et al. 1989). It has been shown that this effect is due to refraction at the magnetopause, inside the magnetosheath and at the bow shock. Some random scattering certainly takes place on random density inhomogeneities.

At present, only the direction of the centroid of the AKR or NTC image can be determined from ISEE-3 data, plus a very rough estimation of its size. These data already show that variations in the solar-wind density produce changes in the apparent direction of the source. But some changes in the shape of the shock or of the magnetopause could also produce some of the observed changes. This possibility is being explored at the University of California at San Diego (by Dr B.V. Jackson) and at the Paris Observatory in Meudon using ISEE-3 data. If one could measure not only the direction of the source image centroid, but also the distribution of the radio brightness over that image, one could obtain a view of the magnetosphere as a whole and follow the variations of its shape with the solar-wind conditions: plasma flow and magnetic field. This would require some angular resolving power at low frequencies; the only other way to obtain that kind of information would be to use a large number of free-flyers, some of which might eventually cross the boundaries between magnetospheric regions at different places almost simultaneously.

2.3.9 The location and use of a lunar observatory

- Mapping of the Earth's magnetosphere obviously requires a low-latitude site in full view of Earth.
- If the low-frequency observations are to be used to assess the hazards caused by solar flares, continuous monitoring of the solar environment would be required, which means three equatorial sites.
- The ultimate site for VLF (and maybe all radio astronomy) will be on the lunar far side. It has to be realised that this is a unique site in the inner Solar System for these kinds of studies, since only here will shielding from interference be possible.

2.3.10 Conclusions

A lunar observatory would be an ideal base for solar-terrestrial radio physics studies for the following reasons:

- Low-frequency radio sources which pervade the interplanetary medium and geospace can only be observed from above the Earth's ionosphere.

- Large arrays (several kilometres long) are required to achieve sufficient resolution:
  - Due to the scattering limit towards the lower frequencies, solar observations satisfying imaging of the interplanetary medium can be achieved down to about 300 kHz only, which would cover distances from the Sun up to about 0.5 AU.
  - Imaging of the Earth's magnetosphere is a new field that could provide a global view of large-scale Sun−Earth phenomena and could help us to monitor these relationships.

- The Moon is close to the Earth and traverses key regions of geospace.

- The Moon is far enough from Earth to be free of the Earth's noises (man-made).

- The evolution of the solar-flare activity throughout interplanetary space can be monitored by imaging at low frequencies. This may prove to be an essential contribution to the observation of Sun-related phenomena at low frequencies in the context of the lunar base.

2.4 Astrometry

2.4.1 Introduction

The last ten to fifteen years of this century are seeing a considerable resurgence of interest in astrometry, with scientific prospects becoming ever more evident as significant instrument advances become feasible. In both the radio and optical wavelength ranges, milliarcsecond astrometry has become a reality.

In the radio range, the advent of VLBI has led not only to morphological studies at the milliarcsecond level and below, but also to the development of a radio reference system, accurate to the level of a few milliarcseconds over the entire sky. The first proper motions of radio stars are now becoming measurable, and direct distances will follow from the VLBI trigonometric parallaxes.

The Hipparcos satellite is currently measuring the positions, parallaxes and annual proper motions of 120 000 stars at the 2 marcsec level in the optical region. It has already gathered sufficient data for the construction of a catalogue of positions accurate to about 4 marcsec. HST and other programmes should lead to the interconnection of the radio and optical reference systems at the milliarcsecond level, both relative and absolute, with respect to a quasi-inertial framework. Optical interferometry promises to open up the possibility of determining the positions of bright stars at the 10−100 μarcsec level in the near future.

On the ground, meridian circles are likely to have clearly-defined astrometric goals over the coming decades. In particular, they will play an important role in extending the Hipparcos reference frame to fainter magnitudes, and enhancing the stellar reference frame density over that achievable by Hipparcos (about three stars per square degree) and the Tycho experiment (ten to twenty stars per square degree). But with the fundamental limitations of the atmosphere, future advances in accuracy, whether in terms of positions, proper motions or parallaxes, will almost certainly result from space experiments. Even the maintenance of the Hipparcos positional reference system, which will deteriorate through the effects of the uncertainties in the individual proper motions, will require further space measurements if significant degradations over the coming decades are to be avoided.

There are already several space proposals related to astrometry which are under study to improve on the Hipparcos results: the US Points proposal, considered for the Space Station, and four distinct proposals under study in Russia, following on to a greater or lesser extent from the principles of Hipparcos. As on the ground, such projects can be divided into narrow-field or wide-field astrometry, with different goals and different techniques necessary to achieve them.

Consideration of the lunar environment has led the Astrometry Sub-Committee to conclude that the Moon is a very attractive location from which to carry out astrometric observations. It offers an absence of an atmosphere (the most important limitation on the advance of astrometry from the Earth), a stable platform (seismologically inactive), slow rotation (which may be advantageously exploited), low gravity (compared with the Earth), and perhaps simpler stray-light baffling than a free-flying satellite. On the other hand, unless the astrometric instrument is located at the lunar equator, a full-sky survey (which is important for the global reference frame) would not be possible. Furthermore, it may only be possible to implement certain types of astrometric measurements (such as the Astrolabe) away from the lunar poles.

However, even these attractive features do not mean that the next major advance in space astrometry will depend completely upon the lunar observatory capability: a free-flying mission (with the modest optical baselines that are likely to be demanded by the immediate successors to Hipparcos) will still benefit markedly from the all-sky visibility, gravity-free and constant thermal environment that can be exploited for such a mission. The situation is therefore somewhat different from the case of, for example, long-baseline optical interferometry. This is likely to prove very demanding to implement for free-flying satellites, and where full-sky visibility is perhaps not mandatory, so that the Moon would seem to offer very considerable advantages over the technological demands of a free-flying mission.

The situation can be broadly summarised by saying that a lunar-based observatory would undoubtedly offer a very attractive environment for future astrometric observations, and should be exploited, although a free-flyer is likely to offer distinct advantages over any 'bare-bones' lunar instrument.

2.4.2 The astrometric goals for the future

The question of what should be the next step in space astrometry can be approached in two ways:

(i) It may be argued that any advance over existing capabilities is scientifically important. Thus, for example, if a lunar observatory existed, it would be worthwhile undertaking experiments aiming at rather 'modest' advances, e.g. an improvement by a factor of 5 in the proper motions for all 120 000 Hipparcos stars (even, perhaps, without new information on parallaxes, or without the addition of further stars). Were such a facility to exist, its merits should certainly be carefully studied (such advances might be expected of the space astrometry missions currently under study by the Russian groups). Alternatively, one might consider a significant improvement in the precision of parallaxes and proper motions for a large number of stars distributed in discrete areas over the sphere. Some accuracies for existing or proposed space astrometry projects are given in Table 2.5, and some current or projected astrometric accuracies (from the ground, or from ground + space missions) are given in Table 2.6.

Table 2.5. Space astrometry projects (performances as given by the authors)

Parameter	Hip-1	AIST	Lomonosov	Regatta-Astro	Hip-2
	1975	1987	1990	1990	
Proposal by	ESA/Hϕg	Pulkovo	Sternberg	Moscow	Pulkovo
Approval	1980	—	—	—	—
Mission Launch	1989	>1995	>1995	1995	>1997
Duration (yr)	3	3?	3?	2—3	3
Magnitude Limit, Vis.	12.5	13—14	13.0	13—14	13.5
Accuracy (mas)	2	2	2	<10	2
Telescope(s) (cm)	29	2×35	100	6×20	40

Table 2.6. Stellar reference frames — present and future (1 mass = 0.001 arcsec)

Catalogue	Sky Coverage	No. Stars	Positions at Epochs 1990 (mas)	2000 (mas)	1900 (mas)	Prop. Mot.
FK5 basic	N+S	1400	40	40	40	0.7
IRS	N+S	40000	300	300	—	5
SAOC	N+S	250000	1500	1500	500	20
PPM all stars	N	181000	270	270	120	4.3
PPM all stars	S	145000	260	260	150	5.5
PPM high prec.	N	31000	120	120	100	2.4
Hip-1 3 years	N+S	120000	2	20	200	2
Tycho	N+S	>400000	30	—	—	(4)
Hip-2 3 years	N+S	500000	20	2	200	2
Hip-2−Hip-1	N+S	120000	2	2	30	0.3

Table 2.6 gives the sky coverage (north/south), the number of stars and the approximate random errors (standard deviation) of a stellar position or proper motion in the catalogue. Each catalogue is affected in addition by various kinds of systematic errors depending on alpha, delta, magnitude, etc.

The PPM by Roeser & Bastian (1989) is a new catalogue covering the northern sky and including AGK3, the Astrographic Catalogue, etc. A similar catalogue for the southern sky has been provisionally published in 1990.

A Hipparcos catalogue is expected to have smaller systematic errors than ground-based catalogues. Therefore, the rather large systematic errors of ground-based positions from SAOC or PPM at, for example, epoch 1900 can be determined by comparison with a Hipparcos catalogue and can then be partly removed.

The Tycho proper motions may be obtained with a precision of 4 marcsec/yr using older ground-based observations.

Hipparcos-2 (Hip2) is a project being considered by the Russians for launch in about 1999. The 120 000 stars in common with the scope of Hip1 would measure very accurate proper motions from the positions at two epochs.

(ii) An alternative approach to the scientific problem is to determine what precision is needed to achieve certain scientific goals. It can be argued, for example, that extending the horizon of accurate trigonometric distance estimates significantly, or accurately determining the motions of stars at distances of the Galactic centre, calls for parallaxes and annual proper motions at the 0.01 marcsec (10 μarcsec) level for stars as faint as 16 mag to cover the brightest quasars directly, as well as more distant populations of the H-R diagram. Absolute parallaxes and system-free proper motions at this level would call for a Hipparcos-type global measurement system, but with two orders of magnitude improvement in accuracy. Such a system might well be technologically feasible with a 1 m-class aperture, albeit highly demanding, although it is difficult to envisage such a measurement system being designed for anything other than a free-flying satellite, where all-sky visibility, the absence of gravity, and a quiescent thermal environment would be achievable. An assessment of the possible scientific return as a function of achievable astrometric accuracy is given in Table 2.7.

It is important to note that at the $1-10$ μarcsec level and below, a variety of general relativistic effects (frame dragging, relativistic precession) become significant, so that global astrometric missions at this accuracy will be capable of direct metric measurements as tests of general gravitational theories.

Table 2.7. Accuracies for Astrometry Breakthroughs (OC = Open Clusters, GC = Globular Clusters)

Accuracy	Parallaxes	Annual Proper Motions
50 micro-arcsec	All stars <400 pc to better than 2%	Stellar gravitational lenses at <100 pc to <10%
V limiting: 15	Red giants and C stars <10%	Motion of spiral arms from OB stars up to 5 kpc <10%.
		Detection of density waves' contribution to motions
10 micro-arcsec	100 Open Clusters (OC) <1%	Dynamics in nearest OC
	All OC <10%	Oscillation of Earth-like planets at 20 pc <10%
	20 Globular Clusters (GC) <10%	
	50 Cepheids <1%	
V limiting: 17—18	All Cepheids <10%	10—20 micro-arcsec is about the upper limit on the relative
	40 RR Lyrae <1%	motion of two QSOs 30 arcmin apart
	OB Stars and Supergiants up to 5 kpc <10%	Global optical astrometry
	Radii of Suns at 100 pc <2%	
1 micro-arcsec	All OC <1%	Internal dynamics in LMC and SMC
	Stereoscopy of nearest OC	Tests on proper motions of QSOs
	20 GC <1%	
	100 GC <10%	
V limiting: 21	All Galactic Cepheids <1%	Relative motions of angularly close QSOs
	LMC, SMC Cepheids <1%	
	All Galactic RR Lyrae <5%	
	LMC RR Lyrae <10%	
0.1 micro-arcsec	100 GC <1%	Internal dynamics in local group galaxies and M31
	All Galactic GC <10%	Tests on proper motions of distant QSOs in QSO pairs
	Cepheids in M31 <10%	
V limiting: 26	RR Lyrae in LMC <1%	
	RR Lyrae in M31 <10%	

2.4.3 Astrometry using VLBI with an Earth–Moon baseline

The system of positions of extragalactic radio objects will soon become the basis of the definition of a fixed celestial reference frame to which all motions of stars or objects of the Solar System can be referred. This is being now worked out by the International Astronomical Union. When a lunar base is established, therefore, quasars and other point-like extragalactic sources will be the primary fiducial points for all astrometry.

The baseline is of the order of 4×10^{11} mm. It will be measurable to 2–3 mm accuracy using lunar laser telemetry, with two or three wavelengths to eliminate the refraction (such an instrument is already at an advanced planning stage). This means that clock synchronisation is possible by direct continuous observation, and that it can achieve the precision of present hydrogen masers (10^{-13} or 10^{-14}). This will permit us to know a priori the position of the central fringe to a few wave periods, so that the quantity of data to be collected will be reasonably small. Technical efforts should then be devoted to obtaining a significant sampling of the radio wave. With the advent of picosecond electronics, this should allow us to work at $\lambda < 1$ cm with a sampling of 10–20 points per wavelength. This would give angular precisions better than 1 μarcsec. The astrometric returns would be primarily a reference frame defined to better than 1 μarcsec. We would gain access to the proper motion of some galaxies and quasars. It would also be possible to monitor the Earth–Moon distance to a fraction of a millimetre. Such a sensitivity on lunar orbit and libration would lead to large domains for new interpretation of the dynamics of the Earth–Moon system. Parallaxes of distant radio stars might also be obtained to 1 μarcsec accuracy, giving a 1% accuracy in distance measurements at the distance of the centre of the Galaxy.

2.4.4 Lunar possibilities and constraints

It has already been noted that the slow rotation of the Moon would be advantageous for certain types of scanning measurements. However, the simplest type of scanning system, with the scanning of a particular declination strip at the lunar siderial rate, could yield positions to very faint magnitude limits (25–26 mag for a 2 m-class telescope equipped with a CCD detector), and thus be suitable for a sky survey, but would not yield parallaxes or proper motions, even if the declination strips were re-scanned after 5–10 years.

Programmes aiming for parallaxes and proper motions would ultimately have to be linked to the local vertical, and would therefore call for accurate monitoring of the lunar motion. Currently, lunar ranging yields the Moon's motion to an accuracy of roughly 2 cm, whereas 2 mm might be a reasonable expectation within 5–10 years. Global astrometry, as noted previously, would require the instrument to be located at the lunar equator (whereas a meridian-circle system could be located at the pole). High stability would be required about the vertical axis, while the relationship of the star positions to the vertical could be supplied via an Astrolabe-type principle. Whether one or two viewing directions would need to be observed simultaneously would depend on the amplitude of the harmonics of the lunar rotation.

2.5 Planetary astronomy

2.5.1 Introduction

Ground-based planetary astronomy has become more and more attractive to planetologists for at least two reasons. First, the major exploratory missions have raised a large number of questions that need to be answered in order to understand the basic processes in the Solar System. It is obviously impossible to design and perform numerous ad hoc missions addressing topics too small to be economically justified.

Secondly, ground-based instrumentation is becoming increasingly sophisticated and so the possibility of transferring at least part of the activity to a lunar base is very appealing.

Major questions that could be studied directly from the Earth–Moon system are reviewed below, and the advantages and disadvantages of a lunar base are examined.

2.5.2 Time-variable phenomena

A number of delicate problems in planetary sciences are related to phenomena with time scales in the order of days or even less, such as volcanic events and the evolution of atmospheric features. Other phenomena such as the mutual orbital events of satellites or the behaviour of ring particles occur on time scales of months or years. Though suitable for an initial recognition of these phenomena, space missions generally do not allow prolonged observations, because of the orbital constraints of the missions and/or because of conflicts between the observing instruments onboard.

Table 2.8 provides the linear and angular diameters of the objects. Issues requiring extensive observation include:

- *Venus*: time-variable features in the atmosphere of Venus; sulphuric-acid clouds as evidence for volcanic activity.
- *Mars*: time-variable features in the Martian atmosphere; atmosphere/surface interactions, such as polar-cap formation and dynamics; orbital dynamics of Phobos and Deimos.
- *Asteroids*: discovery and recovery of asteroids; asteroidal surface activity, in connection with asteroidal dust streams; Earth-crossing asteroids.
- *Giant planets*: atmosphere dynamics, in particular time-variable phenomena in clouds; Jupiter's Great Red Spot; magnetospheric phenomena and electroglow systems; discovery of small or distant satellites.
- *Satellites of outer planets*: Io and Triton volcanism; time-variable phenomena on the surfaces of Europa and Ganymede; Titan's atmosphere and atmosphere/surface interactions.
- *Rings*: ring dynamics; evidence of accretion collisional disruption phenomena in the ring systems of giant planets; structure and dynamics of Neptune's rings; time-variable phenomena such as spokes and knots; orbital evolution of shepherding satellites.
- *Pluto*: orbital dynamics of the Pluto–Charon system; surface features of the two bodies.
- *Comets*: discovery and recovery of comets; orbital dynamics of comets, especially near the Sun; activity of comets as a function of distance; detection of complex molecules in cometary emission; surface features such as active regions, dust covers; splitting phenomena.

2.5.3 Other studies

Many other types of astronomical observations that would be of great interest for planetary studies would be possible with instrumentation placed on the far side of the Moon, due to the optimum visibility and clean environment there. They include:

- Detailed observation of planetary surfaces, especially Mercury, Mars and the satellites of the outer planets, and the surface features of the major asteroids.
- Chemistry and physical properties of interplanetary dust particles.
- Detection and study of circumstellar discs around nearby stars.
- Detection of emissions from the Oort Cloud or other cometary reservoirs.
- Precise measurements of the solar wind.
- Orbital dynamics in general.

2.5.4 Usefulness of a lunar observatory

As for any other astronomical observation, a lunar base would significantly help planetary astronomy. As a matter of fact, the possibility of using large telescopes in a very clean environment allows recognition of very faint objects, or observation of planetary surfaces requiring a very good resolving power.

Observing instruments should be placed on the far side, near the lunar equator, and a system of relay satellites around the Moon would then ensure a continuous link with the Earth.

Telescopes may be remotely operated, so that a human presence would be needed only for maintainance purposes or the exchanging of instrumentation.

European scientists would certainly make great use of such facilities. This has been demonstrated by the interest in the Orbital Planetary Telescope (OPT) mission proposal, and by its very active presence in research domains such as asteroid and comet sciences, planetary surface geology and mineralogy, and atmospheric studies.

Like other observatories on the Moon, planetary facilities would not produce pollution per se. However, even the sporadic landing of manned spacecraft could contribute substantially to the alteration of the content and structure of the lunar atmosphere. In principle, a twin lunar base, with one facility near the equator and another on the far side, may cover all the domains of planetary sciences.

Current technology is already sufficient to carry out very good observations. It is hoped, for example, to extend and complete the statistics of asteroids down to the 18 th and 19 th magnitude (it is now complete down to the 12 th and 13 th). It is easy to forecast that much higher level studies will be possible in the near furure with the rapid developments in focal-plane instruments.

2.6 Role of Europe

2.6.1 Interferometry

Considerable expertise is available in Europe in ground-based interferometry, both in the optical and in the millimetre domain. Key technologies for lunar interferometry include the development of detectors for interferometric use, mixers for heterodyne interferometry, feed antennas and electronic correlators for submillimetre

Table 2.8. Linear and angular diameters of major bodies

Bodies	Linear ϕ (km)	Angular ϕ (arcsec)
Mercury	4,878	6.7 (4.6—12.5)*
Venus	12,102	16.6 (9.6—60)*
Earth	12,756	6,844
Moon	3,476	—
Mars	6,787	18 (14—24)*
Phobos	27×21×19	0.06
Deimos	15×12×11	0.04
Ceres	1,025	0.80
Vesta	555	0.56
Pallas	538	0.42
15 asteroids	>250	0.17
Jupiter	142,800	47
Amalthea	240	0.08
Io	3,640	1.2
Europa	3,130	1.0
Ganymede	5,280	1.7
Callisto	4,840	1.6
Saturn	120,660	19.5
Mimas	340	0.05
Enceladus	500	0.08
Tethys	1,060	0.17
Dione	1,120	0.18
Rhea	1,530	0.25
Titan	5,150	0.8
Hyperion	220	0.03
Iapetus	1,460	0.24
Uranus	52,400	4.0
Miranda	472	0.04
Ariel	1,158	0.09
Umbriel	1,172	0.09
Titania	1,600	0.12
Oberon	1,550	0.12
Neptune	50,000	2.4
Triton	3,200	0.15
Nereid	~200	0.01
Pluto	2,400	0.11
Charon	~800	0.03

* Extreme variations in angular ϕ are given in parentheses for the terrestrial planets

wavelengths, large deployable antennas and structures, optical beam combination (transport, delay lines and stabilisation), fringe measurement and image and precision metrology.

2.6.2 Astrometry

There is a long tradition of astrometry in Europe, further augmented today by the Hipparcos programme. All phases of a lunar astrometry project would benefit from the expertise that already exists or will be developed during the next decade in order to exploit the Hipparcos data.

2.7 Conclusions and Recommendations

2.7.1 Interferometry

The scientific case for increased angular resolution in the study of the Universe is very strong, and it is clear from our preliminary study that the Moon offers the almost ideal site for interferometry, while there does not appear to be any technological barrier to the installation of interferometric facilities on the Moon. We have no reason to doubt that the Moon is the 'ultimate location' for interferometry.

The logical strategy is to build on the current Earth-based technology in long-baseline interferometry towards an intermediate-baseline (100 m) space based optical interferometer (which has its own scientific justifications), and ultimately construct a very-long-baseline (1 km) lunar instrument. The intermediate step could be a short-cut for the sub-millimetre instrument as experience will be gained in the framework of the First Cornerstone mission.

The recommended actions are:
- An in-depth design study for a 100 m-class, space-based, imaging, optical interferometer that provides a substantial increase in capability over Earth-based systems. The specifications for this system could be established by the Lunar Interferometry Study Team (LIST) and would take into consideration the recommendations of the Space Interferometry Study Team (SIST), modified appropriately in view of the long-term lunar objectives.
- A conceptual study of a very-long-baseline (>100 m) interferometer, to be carried out by the LIST.
- The exploration of a variety of routes to encourage the further development of key technologies for lunar interferometry such as: (i) in-house studies at ESTEC; (ii) collaboration with ESO, IRAM or other European organisations with significant relevant expertise; and (iii) Technological Research Programmes (TRPs), including the specific technologies used in ground-based interferometry which are of crucial importance in space and on the Moon (e.g. detectors).
- A close involvement in the tracking of the technical progress of the TRP Optical Synthesis Technologies by the LIST, the specialist expertise of the LIST members being beneficial to the successful completion of this project.
- A major ESA initiative beyond the present TRPs is essential if Europe is to play an appropriate role in space interferometry and in the lunar initiative.

2.7.2 Very-low-frequency astronomy

A lunar observatory would be an ideal base for solar-terrestrial radio-physics for the following reasons:
- Low-frequency radio sources that trace the interplanetary medium and geospace can only be observed from above the Earth's ionosphere.
- Large arrays (several kilometres long) are required to provide angular resolution.
- The Moon is close to the Earth and traverses key regions of geospace.
- The far side of the Moon is free from Earth's interferences, either man-made or natural.
- Monitoring of the evolution of solar-flare activity throughout interplanetary space can be achieved by imaging at low frequencies. This may turn out to be an essential contribution of the observation of Sun-related phenomena at low frequencies in the context of the lunar base.
- A large VLF telescope on the Moon may be the most unambiguous way to detect planets beyond our own Solar System.

A telescope based on the design of the Nançay Decametre Array may be the most suitable lunar installation for achieving all of the above. In order to exploit it fully, it should be located on the far side of the Moon. Another possibility in an early phase is that small Surveyor-sized spacecraft could be soft-landed on the Moon, from which simple wire antennas could be ejected to large distances to provide spatial resolution at VLF. This concept requires study as there are several inherent problems, mainly in the construction and deployment of the (simple) antennas, and the deployment and autonomy aspects of operating on the far side of the Moon.

A spacecraft of this type would be relatively simple and could be deployed regardless of the status of the lunar base. It appears quite feasible to combine this VLF experiment with the far-side lunar-science surface stations discussed in Chapter 1.

2.7.3 Astrometry

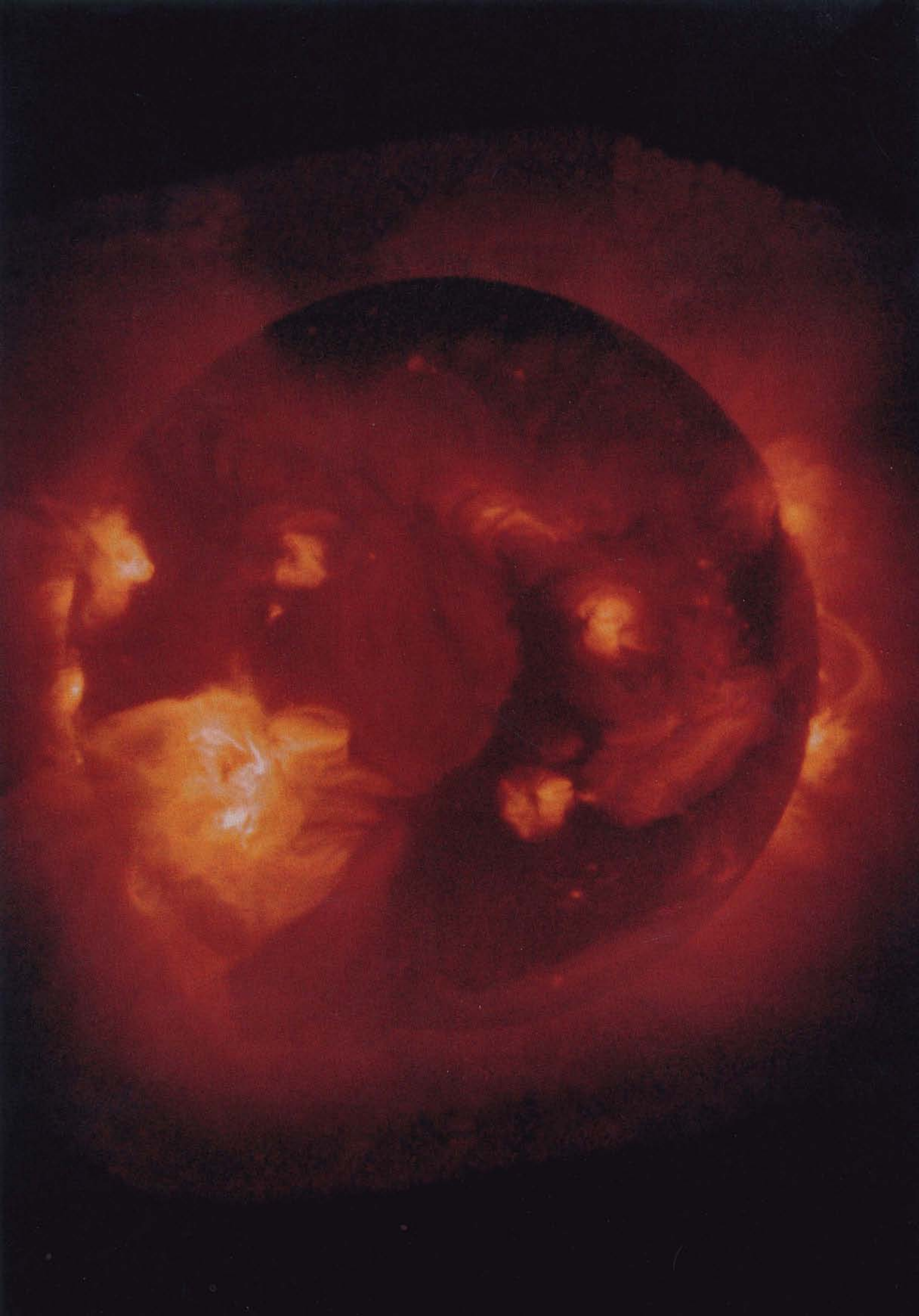
A 1 m-class Hipparcos-type orbiting observatory could achieve 0.01 milliarcsec accuracy and and V-limiting magnitude of 16. Such a spacecraft mission would have a precision two orders of magnitude better than Hipparcos and would clearly be feasible within a few years. However, such a satellite would be limited to only a subset of the scientific goals outlined above, and it would have to be of the Hipparcos type, i.e. an instrument with a split field of view which depends on its own smooth motion to relate the positions of the observed stars. On the other hand, a small astrometric interferometer could achieve the same levels of accuracy.

When determining parallaxes and proper motions, the relative observations will ultimately have to be linked to a reference frame. The International Astronomical Union is presently developing such a reference system based on extragalactic radio objects. This makes Moon–Earth VLBI very promising in order to improve on the reference frame. The possibility of measuring the Earth–Moon distance to an accuracy of a few millimetres and the current advent of picosecond electronics promise to provide angular precisions much better than 1 μarcsec. This results in the possibilities of achieving:
- a reference frame to better than 1 μarcsec (per year)
- access to proper motions of some galaxies and quasars.
- monitoring of Earth–Moon distance to better than a fraction of a millimetre
- parallaxes of radio stars giving 1% accuracy at the distance of the Galactic Centre.

It seems likely that the most ambitious goals of astrometry for the next century can only be achieved through interferometry. The dynamics of both our Galaxy as a unit and of such far-removed objects as quasars are within the scope of 1 μarcsec astrometry and, it must be said, are extremely important areas of study.

Further study on how to implement these techniques is therefore recommended, and it is suggested that this be done under the auspices of the Lunar Interferometry Study Team (LIST).

3. Physics of the Sun



Chapter 3. Physics of the Sun

3.1 Solar physics

3.1.1 Introduction

In addition to sheer scientific curiosity, the Sun is studied because it is the only star whose surface can be spatially resolved. This enables the study in great detail of the subtle magnetic structuring and evolution of its atmosphere, the heating of the corona and, by means of helioseismology, the structure and dynamics of the interior. Solar physics is important for basic sciences such as magneto-hydrodynamics, plasma physics, and atomic and particle physics. In addition, the Sun forms the basic reference for the theory of stellar structure, stellar evolution and stellar activity. Owing to this solar−stellar connection, increasingly detailed questions will be asked of solar physicists. To enable them to provide answers, a spatial resolution and a data continuity is required that it will be very difficult to procure from satellites. It is predicted that a lunar base will be important for solar physics in two respects:

- It will provide a strong incentive to obtain continuous observations of global solar behaviour over a very long period (decades).
- It will provide a natural environment in which to deploy the large structures (typically 10−100 m) that are necessary for high-resolution studies of the solar surface.

3.1.2 Goals for solar physics in the 21st Century

Due to the complexity of the physics of the solar plasma, new discoveries in solar physics have been made almost exclusively by observation. Therefore, new observational facilities are essential for solving the major remaining questions. Progress in solar physics has generally been the result of increased resolution (spatial, temporal, spectral, signal-to-noise ratio and polarimetric accuracy), coupled with a large spatial, temporal and spectral coverage. Of these, spatial resolution is arguably the most important and technically the most difficult item. Numerous global aspects of the Sun, in particular solar activity, are the result of small-scale processes. For example, the causes of the global solar luminosity variations can only be determined if the luminosity within small magnetic flux tubes is accurately known. It is therefore necessary to understand the small-scale solar structures thoroughly before a satisfactory physical description of many global properties of the Sun and other cool stars can be arrived at. Table 3.1 summarises the typical range of instrument parameters that will be relevant for optical solar physics early in the next century. It is emphasised that these requirements cannot and need not all be met at the same time; trade-offs are discussed later. The most important requirements are reviewed as follows:

Table 3.1. Performance of solar optical instrumentation (10^2—10^4 nm) required in the 21st Century

	Resolution	Coverage
Imaging (spatial)	10^0—10^{-2} arcsec	10^2—10^4 arcsec
Imaging (temporal)	10—10^{-1} s	10^4—10^9 s
Spectrometry	10^4—10^6 ($\lambda/\Delta\lambda$)	10^2—10^4 nm
S/N	10^2—10^5	10^2—10^4 nm
Polarimetric Accuracy	$(S/N)^{-1}$	10^2—10^4 nm

■ **Spatial resolution: down to 0.01″ (7 km on the Sun)**
Near the turn of the century, programmes such as OSL, HESP, LEST, THEMIS, P/OF and others will have pushed the spatial resolution to about 0.1″ at wavelengths ranging from the visible to X-rays. However, the structures that matter on the Sun are expected to be at least an order of magnitude smaller. The horizontal mean free path of photons at the level where the vertical optical depth is unity is of the order of 10 km. Recent observations (by HRTS) suggest that the transition region has a fine structure almost certainly smaller than 100 km and possibly as small as 10 km in size. Magnetic reconnection also takes place on this scale: the typical thickness $B / \nabla B$ of a magnetic current sheet is believed to be 1 – 10 km. X-ray Bremsstrahlung-emitting regions are expected to be approximately 10 – 100 km in size. On the other hand, there is at present no (direct or indirect) observational evidence or physical argument which suggests that observable structures smaller than about 0.01″ exist.

The following is a brief selection of unresolved problems that could be solved if observations with a spatial resolution of 0.01″ were available. The first of these concerns photospheric and chromospheric magnetic fields. The internal structure of magnetic flux tubes and their interaction with the granulation is unknown (e.g. how thick is the boundary current sheet and what dissipative processes take place there?). Another important question is what wave modes exist within flux tubes, how are they excited and dissipated, and how much energy do they transport? If we knew their importance relative to acoustic waves, we would be able to determine the heating mechanism(s) of the chromosphere. Some magnetic energy will be hidden in tangled form on very small spatial scales, but how much is unknown, as are the consequences for the outer atmospheric layers.

A second set of basic open questions involves the higher layers of the solar atmosphere. The nature and the origin of the small-scale structure of the transition zone, its relation to spicules, Hα fibrils and the many chromospheric and transition-region phenomena is not understood. Owing to insufficient spatial resolution, we still do not know the mechanism(s) responsible for a solar flare – the archetype of a magnetic instability. As long as the solar-flare phenomenon is not understood, rescaling to, for example, flare stars and flares in accretion discs remains speculative. Is there a continuous scale down to 'micro- and nano-flares', and what are the relevant magnetic instabilities along this 'spectrum'? To understand coronal heating, it is important to know the relative importance of small-scale reconnection (micro- and nano-flares?) and wave heating (dissipation of Alfvén waves).

One of the major challenges facing any advanced solar observatory will therefore be to improve the spatial resolution to 0.01″ .

■ **Temporal coverage: uninterrupted viewing for as long as possible (ideally decades)**
We must emphasise here the importance of long and uninterrupted data sets for solar physics, covering the entire disc, or at least a good fraction of it, at a reduced spatial resolution (1″ – 0.1″). Like many stars, the Sun is a magnetic oscillator. The nature of this solar cycle with a period of 22 years is poorly understood. Monitoring programmes recording sunspots, magnetograms, luminosity variations (as a function of wavelength band), variations in the differential rotation, statistical properties of the

granulation, etc. extending over many cycles are essential. Such records are now beginning to provide important clues to the nature of the solar dynamo. Helio-seismology needs a long and continuous baseline to secure a frequency resolution that is sufficient for inversion. Changes in the oscillation frequencies attributed to the solar cycle have recently been detected. Eventually, this will provide a very powerful diagnostic tool for the solar dynamo. An extensive database is also required for understanding and predicting the occurrence of flares and the concomitant hazardous radiation.

Truly uninterrupted viewing is only possible by having several observing platforms at various longitudes in the ecliptic at 1 AU. That would allow complete coverage of the sphere (important for the life history of active regions and hence for predicting and alerting, and for helio-seismology), and also stereo viewing of surface structures, determination of directivity of emissions, etc. The importance for solar physics is recognised in ESA's Long-Term Plan 'Space Science: Horizon 2000', where it is referred to as the Synoptic Array Programme.

■ **Temporal resolution: 0.1 s at optical and UV wavelengths, 0.01 s at radio frequencies**

In the photosphere, the sound and Alfvén travel times across a 0.01″ structure are 1 s and 0.5 s (for kG fields), respectively. In the chromosphere and transition region, the Alfvén travel time decreases steadily. Very fast variations occur in solar flares. The shortest (temporally resolved) observed features correspond to 0.01–0.1 s at radio frequencies and 0.1–1 s in X-rays. Of course, many observations (such as helio-seismology) do not require such a high temporal resolution.

■ **Spectral resolution: $\lambda/\Delta\lambda \geq 10^6$, corresponding to a velocity resolution better than 300 m/s**

This number follows from elementary considerations: The typical velocity of the granular flow field is 1 km/s; line half-widths of photospheric lines are typically 2–3 km/s. The minimum required to resolve the line profiles is 0.6 km/s (asymmetries, turbulence, Zeeman broadening, etc.). Note that line asymmetries begin to degrade at a resolution worse than 0.3 km/s; polarimetry tightens the requirements on $\lambda/\Delta\lambda$ by a factor of 2–3; Observations of solar oscillations of low l (≤ 4) require an accuracy of better than 1 m/s.

■ **Spatial coverage: 2000″ (full disc), or at least 200″ (active-region scale)**

Monitoring and prediction (e.g. of flares) and observations of global solar oscillations require full-disc observations. The magnetic-field morphology and evolution (field lines emerging at one surface position may connect to distant solar-surface features) also needs to be studied at this scale, as does the connection between small and large scales (e.g. granulation and super-granulation).

■ **Wavelength coverage: from <0.1 nm to approximately 1 km**

This is required for the coverage of different height ranges. The corona emits mainly γ-rays (flares), X-rays, EUV and radio waves. The transition region is best observed in the EUV and the UV, while the photosphere and chromosphere may be studied in the UV, the visible and the IR. The spatial resolution is proportional to $1/\lambda$ for a fixed telescope size (short wavelengths are an advantage). Zeeman sensitivity of spectral lines scales as λ, which implies that long wavelengths are an advantage. For Planck continua, the intensity contrast corresponding to a given temperature difference decreases rapidly with λ. Finally, very long radio wavelengths (approximately 1 km) can only be observed from the far side of the Moon. They represent one of the last uncharted frontiers of the solar/heliospheric electromagnetic spectrum.

- **S/N ratio: $10^2 - 10^5$, depending on the application**

A high signal-to-noise (S/N) ratio is required for accurate temperature, velocity and magnetic-field determinations from spectral line profiles. Variations in the local and global luminosity can be arbitrarily small. The current accuracy is of the order of 10^{-4} of the continuum intensity. Signal amplitudes from intrinsically weak magnetic fields (ubiquitous turbulent field, intra-network fields) are less than 1% of the continuum intensity. The rapid decrease in the modulation transfer function of a telescope with increasing spatial frequency means that the contrast of the smaller spatial structures is greatly reduced. To obtain reliable information on them, a high S/N is required. Structures smaller than the horizontal photon mean free path at vertical optical depth $\tau = 1$ may be seen in light coming from larger depths, but at greatly reduced contrast so that a high S/N is required. This may allow 0.01″ to be achieved in the IR ($1-2$ μm).

- **Polarimetric accuracy: same as the expected S/N ratio**

High polarimetric accuracy is required to measure the full magnetic vector via the Zeeman or the Hanle effects realiably, and implies low, or at least constant, instrument (de-)polarisation. A polarimetric accuracy much below the S/N ratio is a waste of resources, but very high accuracies may not always be technically feasible.

3.1.3 Solar physics: from the Moon or from a free-flyer?

It is evident from the previous section that solar physics in the 21st Century will need large structures to achieve 0.01″ spatial resolution as well as one or a set of modest telescopes (1 m class or less) for uninterrupted viewing at reduced spatial resolution. In principle, this may all be done from free-flyers, but a lunar base could be of considerable help.

During the early phase of a lunar base, solar monitoring station(s) will be necessary with the objective of predicting solar flares and radiation hazards. These observations could conceivably be made by a free-flyer in Low Earth Orbit (LEO), but crew safety is such an important item that a lunar-base programme is unlikely to accept to have to rely on remote satellites. Rapid serviceability will be necessary for security reasons. Thus a lunar base needs solar physics. Solar physics on the other hand would greatly benefit from lunar monitoring stations, because these provide high-quality data with a continuity that is very difficult to procure in practice from a free-flyer. Hence there is a clear synergy between the interests of solar physics and those of a lunar base. Against this background, it would be very cost-effective to design the solar monitoring stations such that they are able to serve the general needs of solar physics in this area, rather than only those of a lunar base (predicting and alerting). This would require only a relatively minor additional effort. An advanced monitoring programme should be able to predict the appearance of active regions on the east limb of the Sun, and to follow the complete life history of active regions. This will provide a strong incentive for a Synoptic Array Programme, which is also of great interest for solar physics.

During the later, more advanced, stages of a lunar base, one could augment one monitoring station with one or a cluster of 1 m-class normal-incidence telescopes, for 0.01″ observations in the EUV (length: tens of metres). Likewise, one might consider deploying a dedicated solar optical interferometer (diameter 10−15 m), an externally occulted coronograph, and X- and gamma-ray imaging collimators (typical size 100 m). Here the advantages of the Moon include: high vacuum, low gravity, servicing, slow rotation (observing sequences of about 10 days are ideal for high-resolution studies), and a stable platform with a very well known position. The latter is important as it may provide a solution for the pointing reference problem, which is a very serious one for any such instrument on a free-flyer: the solar limb is defined to about 0.1″−0.2″ and is very ragged and variable on a scale of 0.01″. As an example of the advantages of serviceability, the presence of man makes it possible to adjust or replace multi-layered coatings. One could even consider manufacturing these under the excellent lunar vacuum, using the solar spectrum as a source for control during fabrication. The scientific return could be greatly enhanced by the possibility of servicing and simple logistics.

The structures needed for these observations can in principle also be deployed in space, assuming that large transportation systems become available, but it remains to be demonstrated that the pointing reference problem can be solved and that such large and complex instruments can ever run in a fully automatic mode.

It is quite possible that the evolutionary approach offered by a lunar-base concept is ultimately more cost-effective − more scientific return for invested money − than an in-orbit facility. A free-flyer in LEO would also permit an evolutionary approach, but there are severe thermal problems and the observing efficiency is low ($\leq 30\%$). In a High Earth Orbit (HEO), one could observe in a more efficient way than from any single lunar location being considered for a base, but refurbishment missions would not be possible; a whole new mission would be needed if one wanted to change instruments or take advantage of new results. At a lunar base, the observatory can evolve at an affordable rate, and one may improve or change the type of instrument and take full advantage of new inventions, discoveries and developments. Large telescopes on Earth are said to be moving towards fully automated operation, but in reality they will always need a small resident staff. Likewise, we foresee that the resident staff of a lunar base would include personnel trained in servicing and astronomy. The data would be transmitted to Earth for analysis.

In summary, although small solar telescopes need not be deployed on the Moon per se, we see definite scientific and programmatic advantages to putting such instruments there in conjunction with, or in preparation for, a lunar base. We anticipate that the programmatic synergy between predicting and alerting for a lunar base and solar physics will lead to continued interest in ESA's Synoptic Array Programme. We also see significant advantages in operating large (10−100 m) facilities for solar physics on the Moon from a lunar base. It is clear that solar physics requires large and advanced observing facilities (as detailed in Section 3.1.5) and that the more ambitious the effort is, the more likely it is to benefit from a lunar deployment.

A critical question is what pointing performance these high-resolution instruments can attain from the stable and known reference provided by the lunar surface, as compared to a free-flyer. This may render a lunar deployment necessary.

3.1.4 Scientific objectives – phasing and requirements

We can divide the scientific topics that a lunar-based solar observatory may address into four broad categories:

1. Monitoring, predicting, alerting.
2. Helio-seismology.

These first two require small instruments based on existing technology. They may thus be realised at an early stage in the lunar programme. They also exhibit a programmatic synergy: the interests of solar physics and those of a lunar base have many common elements. Continuity of observation requires two or three monitoring stations distributed near the lunar equator, or possibly a single station at the lunar pole. The monitoring stations are suitable for robotic deployment, and remote operation and data handling. These are aspects of general interest, which leads us to suggest that ESA study deploying a solar telescope on the Moon as a soft lander – well in advance and maybe even independent of a lunar-base programme. Ideally, the lunar monitoring stations should evolve into a node of a Synoptic Array Programme.

3. Physics of the corona.
4. Physics of the lower atmosphere (photosphere, chromosphere and transition region).

These topics require the deployment of large structures (10 – 100 m) on the Moon, which fits more naturally into a later stage of the development of a lunar base. Serviceability and evolutionary aspects make a location near a lunar base preferable, say at a distance of the order of 10 km.

The various objectives and requirements are considered in somewhat greater detail below.

Monitoring, predicting and alerting

- Monitoring the solar luminosity; both spatially unresolved (an important input for models predicting the evolution of the Earth's atmosphere) and spatially resolved monitoring is required.
- Monitoring solar activity, in particular the pre-flare evolution of active regions (important for flare prediction). This aspect is vital for the protection of astronauts working on the lunar surface from energetic particles and photons.
- Study of evolution effects, including those related to the solar cycle (also of interest for the solar – stellar connection).
- Helio-seismology based on measurements of luminosity variations and on velocity (i.e. spectral) measurements.

The technical requirements are:

- Continuity of observation.
- Coverage of the entire solar disc at a resolution of typically 0.5″ in the UV, visible

and IR; polarimetry in individual spectral lines, preferably in the IR (important for flare prediction). This implies a cluster of two or three small telescopes (typically ~ 25 cm) at each location, augmented by non-imaging X- and γ-ray and energetic particle monitors.

Helio-seismology
- Global structure of the solar interior (sound speed, rotation, helium abundance, etc.).
- Subsurface structure of magnetic and possibly of convective features (such as sunspots and giant cells).
- Evolution of the internal solar structure (solar-cycle effects).

Only the visible and IR spectral bands are necessary. The spatial resolution should be $0.1'' - 0.4''$ to be able to detect modes with $l \sim 10^3$ (Soho will achieve $1.4''$ at best). High spectral resolution (but only in one or two spectral lines) and/or high photometric accuracy (for the measurement of intensity fluctuations caused by the global oscillations) is required. Continuity of observations is crucial and coverage of the entire disc is of great advantage. The implementation of these requirements is relatively simple and comes down to an upgrading of the monitoring station(s).

Physics of the corona
- Identify the heating mechanisms of the corona (magnetic dissipation through nano-flares versus wave heating).
- Identify the source of the slow solar-wind component and the acceleration mechanism of the fast component.
- Understand the causes of reconnection and instability of coronal magnetic fields which result in flares, coronal mass ejections, prominence eruptions, etc.
- Identify the acceleration mechanisms of energetic particles.

An angular resolution of at least $0.1''$ is needed to achieve significant advances in our understanding of what we will have learned from the Soho mission, combined with a temporal resolution of $\leq 1-2$ s. In the UV and visible, a $\lambda/\Delta\lambda \approx 5 \times 10^4$ is needed, and $\lambda/\Delta\lambda \approx 10^4$ in the EUV. In the UV and visible, an occulting disc (i.e. a coronograph) is used. As a first step one might even employ the lunar limb to produce an eclipse near sunrise and sunset, to study the corona out to 1 AU. X-ray imaging with pinhole and imaging collimators is essential (an X-ray collimator with holes of $50 \mu m$ requires 100 m to achieve $0.1''$). Radio observations can be carried out from Earth, except at frequencies below 30 MHz. Observations at these low frequencies are important to study coronal mass ejections and shock waves in interplanetary space beyond $\sim 60 R_{\odot}$.

Physics of the lower atmosphere
- Understanding the physics of convection, magnetoconvection and turbulence.
- Emergence, evolution and decay of magnetic flux and the internal structure of magnetic features (sunspots, pores, magnetic elements, intra-network fields, a possible turbulent background field).
- Structure and energetics of the chromosphere and transition region (e.g. role of CO clouds, canopies, spicules, hot and cool loops, various mechanical heating mechanisms, etc.).

– Excitation, propagation and dissipation of sound and magneto-hydrodynamic waves.

It is crucial to achieve a spatial resolution considerably better than expected from OSL or LEST, i.e. better than 0.1″ . As argued in Section 3.1.2, the lunar observatory must have an angular resolution of typically 0.01″ at 500 nm. The required wavelength coverage ranges from the EUV (10 nm) to the IR (10 μm) with a spectral resolving power of 10^5 in the EUV and 5×10^5 in the UV, visible and the infrared. A polarimetric capability is necessary. A few instrument concepts for achieving these goals are discussed in the next section.

3.1.5 Potential elements of an advanced lunar observatory for solar physics

To illustrate what kind of instruments for solar physics could be installed on the Moon during the later mature phase of a lunar base programme, we have analysed a few selected concepts.

▪ A solar optical telescope (near-UV to near-IR)

An angular resolution of 0.01″ at 500 nm requires a baseline of 10 m in the main optics of the telescope. A temporal resolution of < 1 s necessitates a relatively densely filled entrance aperture. This in turn leads to a large collecting area, which poses substantial problems in handling the collected solar energy flux.

A number of possible configurations of solar telescopes are described in Annex 3. The preferred option is a Michelson interferometer configuration. It may consist of twelve Gregorian telescopes, each of 1.6 m diameter, mounted on the periphery of a

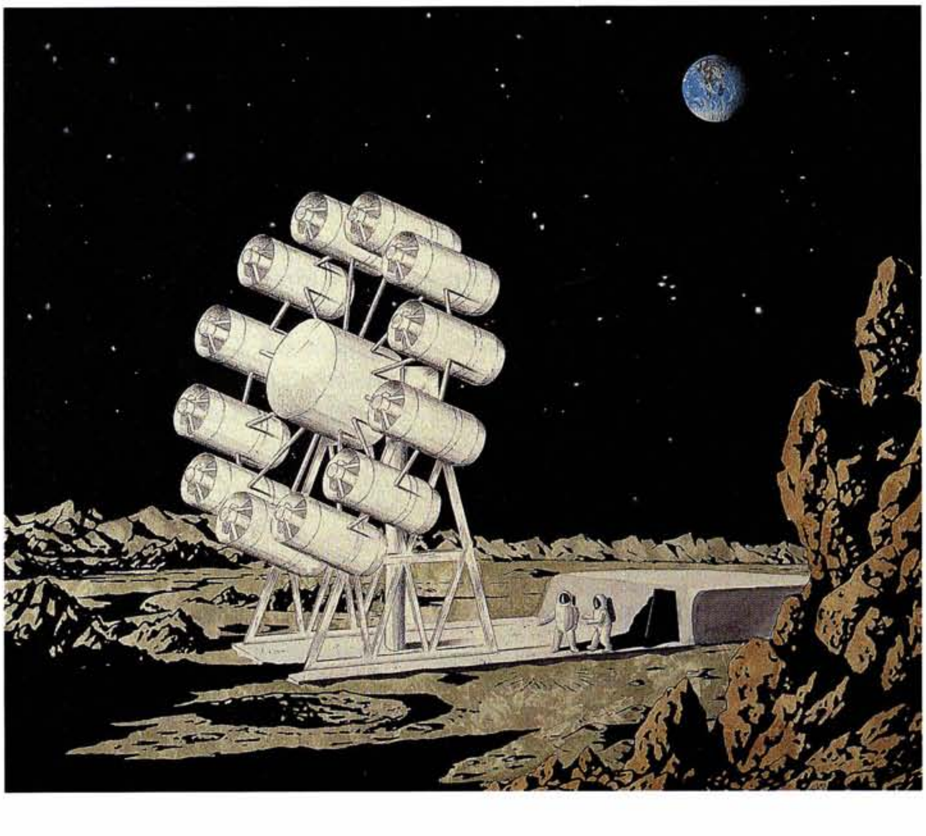


Figure 3.1 A solar Michelson interferometer (cross-section 10 m), deployed on the lunar surface

10 m circle (Fig. 3.1). The symmetrical mounting minimises the instrumental polarisation. Sunlight from each element is directed to the centre, where it is collected in the beam combiner station – itself a telescope of about 1 m diameter. A cross-section of one of the interferometer arms is shown in Figure 3.2. There is a solar image with interference fringes in the focal plane of the beam combiner. Suitable data analysis removes the fringes and restores the original intensity distribution. The array elements are arranged such that all Fourier components in the image can be instantaneously recorded, which is important for the study of fast phenomena at the highest angular resolution. Active control elements, such as rapid guiders and delay lines, relax the stiffness requirement from the fraction-of-a-wavelength level, and allow the structure to be relatively floppy. Other constraints apply, such as the precise remapping of the output pupil to the input pupil within a scale factor, to ensure a sufficiently large field of view.

The light is redirected vertically downwards, through a protecting tube, to a laboratory below the lunar surface. The post-focus instruments reside here, so that the sensitive equipment can be protected from radiation. A section of the laboratory can be pressurised for human access to the equipment. The purpose of the tower, apart from facilitating telescope pointing, is to protect the optics from contamination by lunar surface material. One may eleminate the subsurface laboratory by placing the post-focus instruments in a bay behind the main structure, but the instruments must then be quite autonomous as human access would be infrequent.

The available solar flux does not allow all requirements in Table 3.1 to be met simultaneously at the highest level. Assuming only photon noise, the signal-to-noise ratio at 500 nm in one pixel at the angular resolution limit of the telescope is given by

$$S/N \simeq 2.65 \times 10^6 \sqrt{\epsilon \, \Delta t \, / \, (\lambda/\Delta\lambda)} \qquad (1)$$

where ϵ is the overall system efficiency (including the filling factor of the primary),

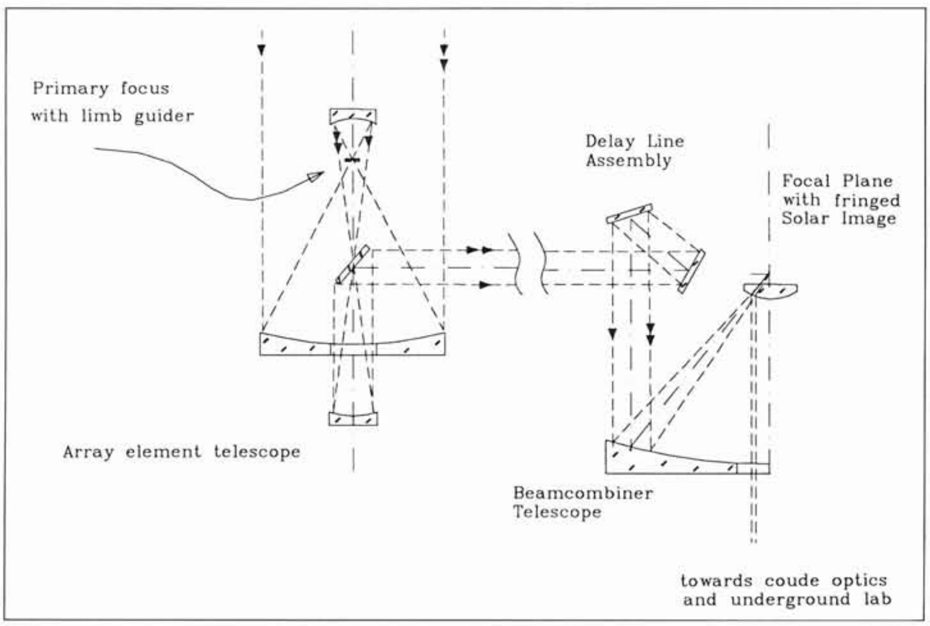


Figure 3.2 Optical design of a solar Michelson interferometer (cross-section through one interferometer arm)

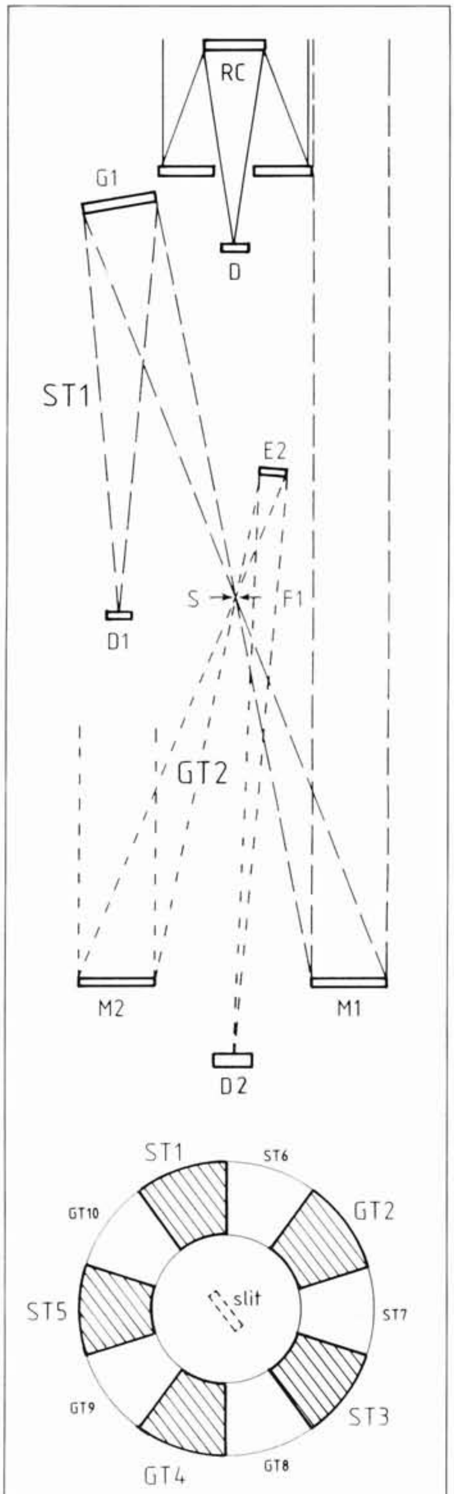


Figure 3.3 Schematic layout of a lunar solar EUV telescope

Table 3.2. Signal-to-noise, temporal and spectral resolution at maximum angular resolution, for $\lambda=500$ nm and $\epsilon=0.1$

S/N	Δt	$\lambda/\Delta\lambda$
10^4	0.1 s	7×10^2
10^3	1 s	7×10^5

Δt the time resolution in seconds, and $\lambda/\Delta\lambda$ the spectral resolution. Two typical examples in Table 3.2 illustrate that reasonable trade-offs are possible. Since the telescope will inevitably have a very large collecting area, it can be profitably used for other purposes also. One suggestion that comes to mind is to dedicate it to study the solar–stellar connection during the lunar night (stellar seismology, activity, convection, etc.).

Study and development of the necessary technologies have already been initiated by ESA in the context of the SIMURIS programme, and the Technological Research Programme (TRP) started in conjunction with the report of the Space Interferometry Study Team.

■ A solar-EUV telescope

The solar emission line spectrum is very rich in the range 0.1–30 nm. Imaging and spectroscopic observations in this domain are very interesting for the diagnosis of the chromospheric and coronal plasma. The multilayer coating technique makes normal incidence instruments reasonably efficient in this wavelength range. However, the number of optical surfaces must be kept to a minimum since reflectivities reach at best 0.5. There are a whole variety of possible instrument concepts, but since folded optics must be avoided they all tend to be very large. This makes deployment on a stable platform such as the lunar surface very attractive, even if not required in view of the pointing reference problem. The instrument described here consists of a circular array of segmented telescopes and spectrographs with only two reflecting components in each combination (Fig. 3.3). It is in fact a combination of three types of instrument:

1. A wide-angle telescope with ten selectable bandpasses, which is a simple extrapolation of the EIT on Soho scaled up by a factor of 8. This could monitor alternatingly up to ten emission lines at a resolution of 0.3″ covering the entire solar disc.
2. Up to five high-resolution telescopes in a Gregorian mount with a spatial resolution of 0.04″.
3. Up to three spectrometers using one mirror and a single grating to provide a wavelength resolution of 0.3 pm.

The wide-angle Ritchey-Chretien telescope (RC) has a diameter of 1 m and is located at the front. An array of ten primary mirror segments (M) covers a circular area with internal and external diameters of 1 m and 2 m, respectively, and feeds an array of a maximum of five telescopes (GT) and/or spectrometers (ST). These all observe the same area of 0.2″×13′ from the solar image. Each spectrometer or telescope uses one 36° segment of the parabolic primary mirror, which has a 10 m focal length, focussing the primary image on one common slit (S). Telescope GT2 and spectrometer ST1 shown in Figure 3.3 lie in planes subtending an angle of 72°. The multi-layer coated elliptic secondary mirrors in the Gregorian telescopes enlarge the prime focus image and ensure a resolution of 0.04″. Multilayer-coated toroidal gratings (G) with 10 m focal distances are used in each spectrometer, which provides simultaneously a spectrum of all 0.2″×0.2″ pixels from the 0.2″×13′ field of view. The spectral resolution can reach 0.3 pm at the 30.4 nm line of HeII, making Doppler measurements of velocities as low as 3 km/s possible. An impression of the sensitivity is given in Table 3.3. Note that due to restrictions imposed by the available solar flux and by geometry, the pixels are generally larger than the resolution limit.

Selection of optical configurations optimised for specific science programmes would necessitate astronaut activity for such tasks as exchange of gratings (G), mirror sections (M) and/or secondaries (E). Astronaut access during operations is thus an essential feature. The entire structure is huge, about 30 m long at a diameter of 2.5 m. It must be light in weight despite the 1/6 th gravity compared with the Earth; it may be assembled on the Moon from elements no more that, for example, 2 m in length. The alignment of the optical elements requires no external reference and is continuously maintained by active control through internal metrology. A simple equatorial mount with a horizontal axis (assuming a location near the equator) could be used for solar tracking.

Table 3.3. Typical photon rates of the Solar EUV Telescope at the highest resolution

RC	1—100	ph s^{-1} $pixel^{-1}$
ST	1—10	ph s^{-1} $pixel^{-1}$
GT	≈ 1	ph s^{-1} $pixel^{-1}$

3.2 Solar wind

3.2.1 Scientific background

Current hydrodynamic and kinetic models satisfactorily describe the slow solar-wind expansion, but fail to explain the density, temperature and bulk velocity in high-speed streams originating in coronal holes. Additional energy and momentum transfer is required in the outer corona to explain the existence of such high-speed streams. The source of this additional energy has not yet been identified. The heating of the solar corona is another long-standing problem which needs a powerful energy source in addition to waves of photospheric origin.

The origin of 'Strahl electrons' often observed in the solar-wind plasma is also an unsolved problem in solar-wind research. The distribution of solar wind beyond the ecliptic plane is a pending problem which fortunately will be investigated with the complement of scientific instruments forming the payload of the Ulysses mission.

The many other unsolved problems in solar-wind physics include the differences observed in the chemical composition of the solar wind and of the photosphere. Why is the ratio of iron to oxygen ions four times larger in the solar wind than in the photosphere? Why is this the case for all refractory elements with a first ionisation potential smaller than 9 eV?

Because of the lack of high-time-resolution plasma measurements in the solar wind (limited by the spacecraft's spin rate), the fine structure of 'discontinuity surfaces' separating two adjacent solar-wind plasma regions has not yet been observed with high enough resolution. Stabilised interplanetary platforms (or plasma detectors on the Moon's surface) can provide such high-resolution plasma measurements.

3.2.2 Scientific objectives

There are several techniques for observing the solar-wind plasma, its magnetic field, and the waves it carries out of the solar corona. The in-situ method of observation using space vehicles is the most direct and straightforward one. It is by using this technique that the solar wind was discovered in the early 1960s. Radio-scintillation and radio-propagation observations from ground-based facilities have also contributed to enhance our understanding of the distribution of solar-wind plasma in the heliosphere. Another original technique has also been used, consisting of deploying a metallic foil on the Moon's surface during an Apollo mission, and determining in the laboratory the concentrations of solar-wind ions which have been trapped inside this foil (Geiss et al. 1970).

■ **In-situ observations of the solar wind**

Many different physical variables and properties of the solar-wind plasma can be measured by this direct method. A wide range of instruments installed onboard interplanetary space vehicles (e.g. Helios, Ulysses) or Earth-orbiting spacecraft (e.g. ISEE-1 to 3) have been developed for this purpose. Several of the nine instruments comprising the payload of Ulysses (launched in October 1990) are dedicated to observation of the solar wind, not only in, but also out of, the ecliptic plane. The interplanetary magnetic-field intensity can be measured with a high time resolution (one B-vector every 0.1 s with an error of less than 0.01 nT); the full three-dimensional velocity distribution of solar-wind protons, electrons, and heavy ions can now be measured in great detail, but with a time resolution that is unfortunately limited by the spin rate of the spacecraft (5 rpm in the case of Ulysses). For spin-stabilised spacecraft (or from the Moon), however, a much higher time resolution could be achieved (e.g. one 3-D velocity distribution every 1 or 2 s); solar-wind ion composition spectrometers can determine the elemental and ionisation state of all ions from hydrogen to iron; these measurements enable us to determine the mean speed and temperatures of these different ions; not only the thermal solar-wind particles, whose energy is low ($< 10-35$ keV), are currently observed, but also other types of particle detectors measure the composition of energetic particles emitted by the Sun, especially during solar-flare events; cosmic-ray particle detectors are dedicated to the observation of the elemental and isotopic composition of galactic cosmic rays and solar energetic particles for energies up to 600 MeV. In addition to this variety of particle detectors (each able to cover only a limited energy spectrum), electrical antennas onboard spacecraft measure the power spectrum of solar radio waves and plasma waves propagating in the solar wind and in the heliosphere; solar-flare X-ray and cosmic γ-ray bursts are also being observed during this interplanetary mission. Indirectly related to solar-wind study, cosmic dust (micro-meteoroids) detectors complete the Ulysses scientific payload.

Although this list of in-situ instruments for studying the solar wind is not exhaustive, it illustrates the variety of physical variables that can be and are currently being measured from an orbiting spacecraft. Adapted and improved versions of these same instruments can be placed on the Moon's surface.

■ **Radio-propagation observations**

When radio waves emitted from a spacecraft antenna (e.g. Ulysses telemetry waves) or by a radio-galaxy traverse of the solar-wind plasma, they are: (i) scattered by small-scale plasma density irregularities; and (ii) they also experience a Doppler effect depending on the integrated electron column density. These additional techniques of observation provide the means to measure the density, turbulence and bulk velocity of the solar-wind plasma close to the Sun.

3.2.3 Strategy

Radio antennas can be installed on the Moon's surface to observe these waves in the same way as they are installed on the Earth's surface. From the past experience of the University of Bern group, it appears obvious that the Moon is a good place to observe the solar-wind composition using foil experiments as during the Apollo mission, or using ion and electron spectrometers, like those forming the Ulysses payload. The advantage of the lunar platform is that it is not spinning very quickly, and that high-resolution plasma measurements can be performed more easily from such a location than from a rapidly spinning spacecraft. It should be pointed out, however, that three-axis-stabilised automated spacecraft placed at the Lagrangian points of the Earth – Sun system would be able to monitor the solar-wind plasma more continuously and routinely than from a man-tended lunar base. Indeed, the incidence of solar-wind particles on the Moon's surface changes with the lunar day. Therefore solar-wind observations would have to be discontinued every lunar night. Furthermore, solar-wind observations will also always be interrupted when the Moon runs into the Earth's magnetotail. On the other hand, these occasions present great opportunities for magnetospheric physicists to study plasma distribution and phenomena occurring in the distant-magnetotail region, but it seems that dedicated orbital missions with free-flying spacecraft offer less expensive and more flexible opportunities for these kinds of studies (i.e. magnetotail and solar-wind plasma physics).

Solar-wind composition measurements are of great importance for studies of:
- Structure and dynamics of the chromosphere and corona.
- Solar composition, in particular isotope abundance.

Isotopic abundances of the noble gases, nitrogen and carbon in the Sun and in the solar wind carry unique information on solar history. They are also very much needed for understanding the fractionation processes that caused the variety of isotopic abundances in planetary atmospheres, because solar matter is safely assumed to be isotopically equivalent to the matter in the nebula from which the Sun and the planets formed 4.6 billion years ago.

At the present time, the solar wind's composition is studied with mass spectrometers on unmanned space probes. However, even the most advanced instruments presently under development have limited capabilities for precision isotope-abundance measurements. In fact, the best solar-wind isotopic data available so far resulted from the Apollo solar-wind collection experiments. If lunar landings are resumed, this technique could be re-introduced from the lunar surface or lunar orbit and could yield extended isotopic-abundance results for the noble gases, nitrogen and other volatile elements. Isotopic data on precisely these elements are needed for comparison with the trapped solar-wind gases in lunar material and with isotopic compositions found in planetary atmospheres. From these comparisons, hard information on solar history and on the evolution of planetary atmospheres can be derived.

An attempt should be made with foils exposed during the lunar night to collect interstellar gas atoms and to measure the densities and isotopic composition of helium and neon in this gas. These data would be of great astrophysical and cosmological significance, especially if they can be directly compared with the corresponding solar-wind data.

The foil collection technique is relatively simple and therefore it lends itself to early application at a lunar landing site or base. The radio-astronomy community studying radio scintillation generated by inhomogeneities in the coronal plasma density may be interested in a lunar-based radio telescope with which the solar wind distribution can be studied.

The physical analysis of spacecraft telemetry signals has proved to be a useful means of determining the total column electron content along a ray traversing the solar wind, the ionosphere, and/or the plasmasphere. A receiver antenna on the surface of the Moon may be an original solution which would be preferable to that of using receivers on the Earth's surface. Indeed, in the latter case the radio signals to be analysed must always traverse the ionosphere where significant distortions are imposed on the electromagnetic waves by the high ionospheric plasma density. A specialist in this field should be consulted and the possibility of studying other plasma regions (e.g. the plasmasphere) in addition to the solar-wind plasma should also be examined.

3.3 Role of Europe

3.3.1 Solar physics

There is a large (~ 600) and highly competent solar-physics community in Europe, involved in almost all aspects of observational and theoretical solar physics. Europe is in the forefront of several instrument-development efforts that are important in the present context:

- The most important European programme is ESA's Soho spacecraft mission.
- The Large Earth-based Solar Telescope (LEST).
- Optical interferometry (ESO's VLT; ESA's technological research; the SIMURIS project on Space Station).
- The French polarisation-free telescope THEMIS.
- The Kiepenheuer Institute Solar Spectrograph (KISS) for the Orbiting Solar Laboratory (OSL) of NASA.
- The Canary Islands solar telescopes constructed by Germany, Sweden and Spain.

This list is not exhaustive, but does include some of the major projects that are either currently being developed or have just become operational. It shows that the European solar-physics community has ample experience in designing, constructing and managing large, technologically complex programmes. There is also strong interpretative and theoretical support for these efforts. There are a few areas in which European solar physicists are presently less active, such as solar-flare prediction, instrumental high-*l* solar velocity seismology, and deployment and alignment of large structures in space, but this may well be a temporary hiatus. Without doubt the European community could take a leading role in any aspect of the development and exploitation of a solar observatory on the Moon.

3.3.2 Solar wind

There is strong expertise in Europe in all aspects of solar-wind science and applications to coronal physics, solar history and plasma processes, and there is an especially strong base in the observation and interpretation of solar-wind corpuscular radiation, stretching back to the Apollo lunar observations.

3.4 Conclusions and recommendations

3.4.1 Solar physics

There is a consensus that solar physics in the 21st Century will need observational facilities providing: (i) a high spatial resolution (0.01″) combined with extensive spectral coverage (X-ray/radio), and (ii) long and uninterrupted data sets at a reduced spatial resolution. This requires observations from outside the Earth's atmosphere.

Uninterrupted viewing can be provided by small telescopes ($\leq$ 1m) in a Sun-synchronous low Earth orbit (LEO), in a high Earth orbit (HEO), and from the Moon. Location on the lunar surface has the advantage of programmatic synergy with a lunar-base programme. We anticipate that the joint interests of a lunar base (predicting and alerting) and of solar physics will lead to continued interest in ESA's Synoptic Array Programme.

High spatial resolution requires the deployment of large structures (10 – 100 m) and this appears to be difficult to realise from a free-flyer, because of the impossibility of any servicing, which precludes an evolutionary approach. The pointing reference is a very serious problem on a free-flyer and may render a lunar deployment necessary. For a free-flyer in LEO, there are the additional thermal and a low-observing-efficiency problems. It appears that it may be more cost-effective to deploy these structures on the lunar surface in conjunction with a mature lunar base. The immediate advantages are accessibility, evolutionary approach, gravity, slow rotation (resulting in ideal observing sequences for high-resolution studies), and a stable platform with a very precisely known position, thus easing the pointing-reference problem.

There is ample experience within the European solar-physics community in designing, constructing and managing large technologically complex programmes, and this experience could profitably contribute to the development of solar observatories on the Moon.

We recommend that a study of a solar observatory that can be developed via an evolutionary approach be initiated. The observatory would be erected first near the lunar base (smaller instruments, monitoring activities) and later at somewhat greater distances (10 km) from the base (interferometers, big structures, 10 – 100 m). Moreover, a number of specific problems need to be studied:

- The problem of the pointing reference on the solar disc, with an accuracy of 0.01″, to determine whether this is more easily achievable from the Moon or from a free-flyer.
- The effect of the lunar environment and man on optics and telescopes.
- The problems of transmission and management of data rates in excess of 1 Gbit/s.
- The deployment of a small solar telescope as a soft lander in preparation for a lunar-base programme.

We recommend further that ESA continues to support the Synoptic Array Programme described in Space Science: Horizon 2000, and continues to study the SIMURIS mission, and that the results of the ESA Technological Research Programme on optical interferometry be studied in the context of a solar-physics observatory in space or on the Moon.

3.4.2 Solar wind

It is proposed that ESA should initiate – as a specific European contribution to the planning of future lunar exploration – a study that aims both at identifying the most suitable lunar samples for solar-history investigations, and at defining requirements and strategies for their collection.

Recent spectroscopic observations of sodium and potassium atoms in the lunar atmosphere have led to renewed interest in the study of the Moon's atmosphere. A multi-disciplinary study to examine the contamination of the lunar atmosphere by various natural processes and/or by future lunar space activities would be most appropriate also, within the framework of a European initiative to possibly build scientific stations on the surface of the Moon.

4. High-Energy Physics


15 BILLION YEARS
5 BILLION YEARS
1 BILLION YEARS
MATTER DOMINATED ERA
10^{13} sec.
PHOTON EPOCH
10^{2} sec.
LEPTON EPOCH
10^{-10} sec.
The desert
QUARK EPOCH
RADIATION DOMINATED ERA
10^{-34} sec.
10^{-43} sec.
Big Bang

Chapter 4. High-Energy Physics

4.1 Introduction

The Earth is continuously bombarded by high-energy cosmic particles. Experiments involving high-energy physics originated with such cosmic rays, discovered in 1913 by V. Hess. They are usually associated with showers of particles produced in the atmosphere by energetic primaries. This typically corresponds to a long series of cascading processes (Grieder 1989). For a long time, cosmic rays provided the only available source of particles with energies in excess of 1 GeV. From the thirties to the fifties, they were the highly fruitful hunting grounds for new particles. Experimentation with high-energy particles moved to accelerators in the mid-fifties, but cosmic rays have remained the object of some considerable research. Indeed, the mechanisms at the origin of cosmic-ray primaries are still important areas of study, the more so as specific properties of the sources still require important studies, which look particularly interesting since specific new features, such as localised sources of very high-energy particles, have been discovered (Grieder 1989; Samorski & Stamm 1983). The physics of such intense 'point' sources, in the TeV range and above, offers fascinating challenges.

At the same time, advances with detectors have pushed the exploration of the high-energy part of the spectrum far beyond the maximum energy available with present accelerators. Cosmic-ray particles of up to 10^{12} GeV have been seen (Fichtel & Linsley 1985). This corresponds to a beam energy in the order of 10^6 GeV for a proton−proton collider. This is two orders of magnitude greater than the potential of the machines predicted to be available only at the end of this decade, whereas with 10^6 GeV cosmic rays the present colliders are sufficient. Fluxes are, however, low. The integral incident flux, including incident particles above a particular energy E (GeV), corresponds to a rate of $(E/10^{10})^{-2}\text{km}^{-2}$ per year. For 10^8 GeV, for instance, this implies about one event per square kilometre per hour, which is rather low. With the Superconducting Super Collider (SSC), collisions at the same centre-of-mass energy will occur at a rate of 10^8 per second. Low fluxes notwithstanding, such extremely high-energy particles remain very valuable tools for a first exploration of high-energy phenomena at still higher energies than those that will be available with the coming generation of colliders. Operation of the CERN p-pbar collider, which over the past decade has provided a centre-of-mass collision energy of 630 GeV, is now drawing to a close. Now the Fermilab collider gives 1.8 TeV p-pbar collisions. The SSC will generate 40 TeV p-p collisions and the CERN Large Hadron Collider (LHC) will give 16 TeV p-p collisions. The energy of the coming generation of colliders corresponds to the $10^8 - 10^9$ range for incoming cosmic-ray protons.

In the recent past, cosmic-ray results have been quite useful for predicting what typical events at the present colliders would look like (Horgan & Jacob 1980). This corresponded to primary energies of typically $10^5 - 10^6$ GeV. There are even a few peculiar events reported in this energy range which still await an interpretation, such as the Centauro events (Lattes et al. 1980). Nothing like them was found in p-pbar collisions. In retrospect, considering the wealth of results from the CERN p-pbar collider (Jacob, 1981, 1987), one may say that cosmic rays provided only a small glimpse of what was found.

The links between astrophysics and particle physics have become numerous and fruitful. The observation of the neutrino burst from SN 1987A, with all the information which it provided both on supernova dynamics and neutrino properties, has been a particularly dramatic example of such a multi-faceted link (Schramm & Truran 1990). More generally, many questions associated with cosmology are addressed in terms of particle physics and huge underground detectors are searching for cosmic neutrinos and for effects due to hitherto unknown particles (Grieder 1989; Schramm & Truran 1990; Rich et al. 1987).

Today's signal quickly becomes the background of tomorrow. There are experiments for which the cosmic-ray background raises serious problems. This is the case particularly for the cosmic neutrino background from which we cannot screen ourselves by going deep underground. The present limit on the proton lifetime is such that this neutrino background presents a serious problem (Pati et al. 1986). It has to be dealt with better than at present in order to reach far beyond the present limit, which is of the order of 10^{32} years. It is with these questions in mind, all bearing witness to the thriving nature of this field of research, that we can approach experimentation on the Moon.

It is clear that experimentation on the Moon would offer some specific advantages, and even great opportunities. If a lunar base were available, and if it were possible to deliver bulky equipment at an acceptable cost, particle and cosmic-ray physicists clearly would love to use it. However, as we shall see, most of the challenging questions correspond to present limitations which are due to the rate of the primaries far more than to atmospheric screening. The conclusion is therefore that even if the Moon can offer direct observation of primaries and much better background conditions, most of the problems can probably be more efficiently addressed on Earth. The case becomes overwhelming when one takes into account the fact that there is a factor of 4 to 5 orders of magnitude in delivery costs, when it comes to the sophisticated parts of the detector which have to be constructed in the home laboratory. This is also a period when very high-energy cosmic-ray research is changing gear (Grieder 1989; Schramm & Truran 1990; Rich et al. 1987). Ambitious detectors originating from particle-physics research are now being substituted for the relatively simple devices long used in cosmic-ray studies. The prevailing attitude is that a high-return investment can be made in the development of large sophisticated Earth-based detectors, and that diverting a significant fraction of the available funding to much less ambitious work based on the Moon would not seem a promising approach. Whereas one meets enthusiastic individuals when discussing Moon-based cosmic-ray studies, a fair assessment is that the overwhelming majority are for the extension of Earth-based studies which have just entered a new era. The typical time scale for large equipment on the Earth is of the order of ten years. It is only a factor of two down compared with the likely time scale for the first permanent scientific exploitation of a Moon base. It is not therefore natural to try to fully disassociate the physics from the funding.

This having been said, one can still address the question with an open mind, since we are in fact considering long-range plans. The purpose of this chapter is to survey topical questions in cosmic-ray research, trying to be general in view of the time scale involved for any project, but also specifying those questions where research on the Moon would eventually offer specific advantages. The question of proton decay (Pati et al. 1986), for which the Moon could indeed present a unique opportunity, but for which the bulky nature of the detector needed and its underground installation would involve a still longer time scale, will be addressed in a separate chapter. We have concentrated on physics issues and not on experimental techniques which could be used on the Moon. It is clear that when such experimental possibilities arise, the available detectors will be far more ingenious and powerful than those that can be conceived of today. There are three main questions to be addressed:

(i) What is the nature and origin of very high-energy cosmic rays? This is an astrophysics question. The key problem in addressing it experimentally is rate, and the answer therefore lies in the size and instrumentation of the detectors. The Earth has advantages compared with the Moon, unless there are challenging questions associated with the composition of the primary flux at very high energy. There are no burning questions now. Some may emanate, however, from the results of satellite studies carried out first at lower primary energies.

(ii) Are there stable hitherto unknown particles? Acknowledged objects, often of unknown nature since there is more dark matter than bright matter, correspond to less than 20% of the critical density, which many theorists would like to find. This is both an astrophysics and a particle-physics question, since one would then have to understand the nature and production of such new objects and this may lead us to the high-energy conditions of the Big Bang. Here again, one should first wait for more information from Earth-based and satellite studies. There is no reason why such remnants should not be within the Galaxy. They should then be even more abundant at lower primary energies. In this case, the Moon (or satellites) has advantages compared with the Earth if cross sections are appreciable, since direct access is needed. This is the case particularly for mono-energetic gamma rays, which could signal the pair annihilation of hitherto unknown particles. Searches in Moon rocks, such as looking for possible evidence for monopoles, will continue on the sidelines.

(iii) Are there surprises in store at extreme energies ? This is a high-energy-physics question. We expect some, but we need pre-conceived ideas in order to assess the potential of specific experimental conditions. As an example, we may consider scenarios in which the electro-weak interaction could behave spectacularly at very high energy. We shall later discuss one in particular in connection with Cygnus X-3 (Samorski & Stamm 1983; Domokos et al. 1989). There is a possibility that a strong interaction behaviour, with geometrical absorption, could affect the cloud of weak gauge bosons carried along by quarks and leptons (Ringwald 1990). This has to do with the infrared behaviour of non-Abelian gauge theories at high temperature. High W densities in very high-energy collisions could provide the specific conditions required. The energy threshold, connected to the W mass, is expected to be of the order of a few TeV in the centre-of-mass. The cross section, also related to the W mass, would be of the order of a nanobarn to a microbarn. Such events would be spectacular, with an abundant production of gauge bosons, Higgs particles and leptons, with transverse momenta of tens of GeV, related to the short-range nature of the interaction (Ringwald 1990). A high-energy behaviour of that type is speculative, but such events are worth looking for. They could be missed with the coming generation of colliders because of their high-energy threshold. For the foreseeable future, cosmic rays possibly offer the only means of looking.

It should be noted that in all of these scenarios, the present one being a particularly striking example, quarks and leptons are affected in the same way since it is the electro-weak interaction which displays totally new features at very high energy. Neutrinos should therefore show such a spectacular behaviour, and for them it would be the only important interaction when traversing dense matter. We expect cosmic neutrinos to be abundant and, even with their cross sections, the atmosphere would still remain almost transparent. Indeed, 1000 g/cm^2 corresponds to one mean free path for a cross section of 3 millibarn. In such a case, experimentation on the Earth wins. The new challenging and speculative questions will certainly be different twenty years from now, after we have explored the TeV range at the constituent level. One may however venture the prediction that anything really peculiar at extreme energy, where cosmic rays could reach beyond the existing colliders of that time, albeit with very low fluxes, should involve a new important and short-range interaction of the neutrino. The Earth would then remain the winner. Purely high-energy physics on the Moon brings us back mainly to proton decay, which we shall discuss later. The electro-weak theory is simple but relies on a rather complicated vacuum which behaves as a medium and shows similarities with a superconductor. The Glashow-Salam-Weinberg theory of the electro-weak interaction can be considered a relativistic and non-Abelian extension of the Landau-Ginsburg theory of superconductivity. The critical temperature would then be of the order of 200 GeV. Peculiar effects should occur when the collision energy much exceeds such a value and, more specifically, when it is greater than 1 TeV at the constituent level. This is what sets the scale for the energy of the SCC (LHC). Neutrino interactions at very high energy are therefore a promising hunting ground.

The main conclusion is already clear from this extensive introduction. The Moon will certainly be used as a base for cosmic-ray studies and some particle-physics studies when it is accessible at an acceptable cost. At present, however, Earth-based and satellite equipment offer the best tools with which to address the challenging questions in these fields of research. There is no pressing need to conduct any major experiment on the Moon in view of the promising Earth-based projects which one wishes to pursue first. However, some limited experiments will probably be carried out at an early stage, in the wake of other projects for which a Moon base offers some unique opportunities (e.g. gamma rays in the 10 to 100 GeV range which are too soft for atmospheric penetration and may be too hard for satellite studies).

4.2 A survey of the cosmic-ray spectrum

Cosmic rays correspond to stable particles entering the atmosphere at very high energy. The majority correspond to protons. However, up to 1 TeV, where a lot of data exists, the fraction of medium and heavy nuclei increases with primary energy. The proton abundance up to 1 TeV is of the order of 40%, with about 20% for alpha particles, and 10% for iron. Above 1 GeV, all rates fall almost as sharply with energy, and primary protons thus retain an overwhelming role when one considers the incident rates in terms of the energy per nucleon. The very high energy spectrum is shown in Figure 4.1. The integral rate per square metre hour and steradian is given as a function of the energy of the primary, E, in GeV. To a first approximation, the spectrum falls an inverse power, E^{-x}, but shows two kinks. The first, which is quite noticeable, corresponds to a change in the exponent x from 1.7 to 2.1. The second, which is not as clear, is associated with a change in the exponent from 2.1 to 1.8. While the first kink could also involve a change of composition, both kinks seem to correspond to the effects of the magnetic galactic field, trapping lower energy particles and permitting the penetration into the Galaxy of extremely energetic particles only. If we consider a primary of 10^6 GeV for example, we are at the level of one event per day for a surface of 10 m^2. This is TeV physics in terms of colliders, which is where we are today. The composition is known up to 10^5 GeV from detectors at various depths in the atmosphere (Burnett et al. 1987). The composition changes, but does not show any wild variation with energy. Up to 10^6 GeV, cosmic-ray events can be studied with emulsion stacks of reasonable sizes. Above that energy, rates are such that one has to rely mainly on extensive air showers, collecting some of the many secondaries associated with the initial collision in the atmosphere and the following cascades. They may cover an extensive area. At sea level, a proton of 10^6 GeV impinging on the top of the atmosphere results in typically 10^5 particles, scattered over several thousands of square metres. Up to 10^9 GeV, cosmic rays appear to originate almost isotropically from space.

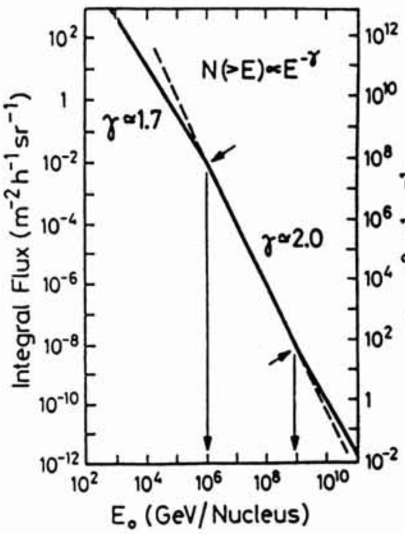


Figure 4.1 The integral cosmic-ray flux

As the recorded energy increases beyond 10^{10} GeV, some maximum may appear in the direction of the 'nearby' Virgo cluster, but this is still being debated. Hadron primaries produce a large number of mesons, which gives rise to muons before they are absorbed. Showers are therefore muon-rich. There are, however, showers which are muon-poor which are naturally associated with incident photons. A photon will indeed originate a cascade where, at each step of photon fragmentation, the relative abundance of charged particles corresponds typically to the inverse square of the mass, as implied by the Bethe-Heitler formula. In such photon showers, higher mass particles are mainly photo-produced. In quantitative terms, the relative probability for the formation of a muon and electron pairs is of the order of 10^{-5}. The relative probability of pion photo-production and electron-pair formation in the atmosphere is of the order of 10^{-3}. Photons are particularly interesting since they are not deviated by the galactic magnetic fields like charged particles. Their direction reflects that of the source, whilst the directional origin of even very-high-energy charged particles, in the 1000 TeV range, is smeared away. Photon primaries are therefore the particles to use when looking for point sources.

One of the prominent findings was the very discovery of such point sources. The most intense are typically galactic X-ray binaries (Samorski & Stamm 1983; Rich et. al. 1987; Domokos et al. 1990). The various sources which have been identified include Cygnus X-3, which is about 10 kpc away, and Hercules X-1, at about 5 kpc. Such sources produce very-high-energy photons over a wide energy range (from approximately $1-10^7$ GeV). The photon flux at 1 TeV corresponds to 5×10^{-11} cm^{-2} s^{-1} for Cygnus X-3, and to 3×10^{-11} for Hercules X-1. Such very high-energy photons also originate from the Crab Nebula, but its relevant luminosity is a hundred times lower than that of Cygnus X-3. These sources of very-high-energy photons are fantastic tools with which to study some of the mechanisms at the origin of cosmic rays. Photons can originate from electrons or from π^0 decay. In the latter case, charged pions should also give neutrinos and the observation of such very-high-energy neutrinos becomes a great challenge.

The dependence on the production conditions of the neutrino-to-photon ratio makes it a very important parameter to measure. An a-priori reasonable estimate gives fifty neutrino events per year from Cygnus X-3 for a 1 Megaton detector, but this assumes a steady source. To be more specific, one may venture a rate of 10^{-10} cm^{-2} s^{-1} for this neutrino flux at 1 TeV. Present detectors can reach only 10^{-7}. Generally speaking, the neutrino-to-hadron ratio in the high-energy part of the cosmic-ray spectrum is a very important parameter. It would tell us how pions can survive in the production zone and therefore allows us to probe the density. The rise of the neutrino cross section with energy, due to the short range of the interaction, makes the detection of such very energetic neutrinos easier than for those of lower energies. The cross-section indeed increases by five orders of magnitude between 10 and 10^9 GeV. The rise of the cross section partly compensates for the expected fall of the incident rate. The study of such neutrinos is currently attracting much attention and ambitious projects are being developed (Grieder 1989; Schramm & Truran 1990; Rich et al. 1987) or considered. One important problem with point sources such as Cygnus X-3 is their apparently erratic character. Fluxes are such that observations with present detectors have to be considered over several years and different data have to be combined (Rich et al. 1987; Domokos et al. 1990). This weakens the strength of the conclusions which can be drawn; hence the caveat when quoting a neutrino rate over one year. Nevertheless, in order to illustrate the tantalising features of this domain of physics, it is worth presenting an apparent puzzle and a possible interpretation (Domokos et al. 1990).

The only known stable and uncharged primaries that can be associated with a point source are photons and neutrinos. Let us consider such particles at the $1-10^4$ TeV level. Photons interact violently in the atmosphere; neutrinos hardly do so and they just pass through. The registered events should therefore correspond to photons. These photons should have been partly absorbed in the cosmic radiation microwave background. The threshold for electron−positron production in such photon−photon collisions leads us to the 1000 TeV range. Since the distance to the source is known, one can calculate the effect of this absorption on the integral spectrum and, assuming the production spectrum to be smooth in the first place, one can expect a relative depletion in the 10^3 TeV range. As shown in Figure 4.2, nothing like that is seen. The data cannot be considered to be fully conclusive. Yet, one is inclined to assume that an unabsorbed component is also at work. If it is due to neutrinos, one would have to conclude that the neutrino cross section is about six orders of magnitude bigger than it should be in the standard model. The neutrino energy threshold corresponding to such a new effect can be estimated from the requisite increase at 10^3 TeV. It is then of the order of 1 TeV in the centre-of-mass. This would be new physics, as expected, however, in composite models of quarks and leptons where neutrinos would have strong interaction cross sections at a sufficiently high energy, thus probing their target at very short distance. This is an exciting challenge. The cross section would have to be of the order of a millibarn to compensate for the unseen, and expected, depletion (Domokos et al. 1990). This cross section is much larger than the one expected from a peculiar infrared behaviour in electro-weak theory (Ringwald 1990), to which we have already referred.

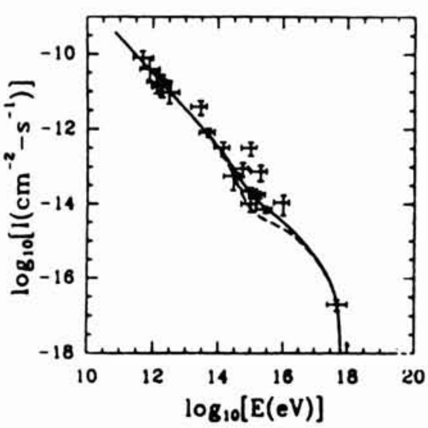


Figure 4.2 Integral spectrum from Cygnus X-3 with calculated spectrum associated with fits to the data

Most of the neutrinos would easily penetrate the atmosphere. Yet some will interact and a welcome consequence is that this would also explain another puzzling fact, which is that the extensive air showers from point sources seem to contain too many muons for their assumed photon origin. This observation is still, however, under dispute (Grieder 1989). It is still at the level of hypothesis, but this illustrates very well the tantalising need for more data on the very-high-energy neutral particles received from such point sources.

It is clear from this survey of cosmic-ray physics that there is a strong demand for far more extensive data than that currently available. The field is actually just opening up to new vistas as very ambitious detectors are developed. Much can be done on Earth, but could experimentation on the Moon eliminate the shortcomings of ground- or satellite-based experimentation?

4.3 Moon versus Earth operations

The Moon has no atmosphere, it has a very low intrinsic magnetic field, and it is well outside the radiation belts. As such, it is a very good observation site for the low-energy component of the cosmic-ray spectrum (GeV range). Its composition, measuring its antiparticle and heavy-element content, is of particular interest. This is, however, an energy domain where the global incident flux is large and much could be done with dedicated satellite research, which typically represents an order-of-magnitude cost saving compared with an installation on the Moon, weight for weight. Such relatively light equipment could however be given a 'lift' to the Moon with another payload. With no extra launch involved, the Moon would indeed offer an advantageous location. Whilst there are important unanswered questions in this energy range, the special interest of the Moon definitely lies in the high-energy part of the spectrum where fluxes are very low. One then needs large and steady areas, for which the Moon would be an ideal platform. One also needs detectors with calorimetric properties, which require a large amount of material. This material could largely be found on the Moon itself, since a good fraction of the mass is associated with absorbers in the detector. This represents an important advantage over satellite studies, which could even extend down to the GeV region for photons. Photons with energies up to 100 GeV are indeed too soft to penetrate the atmosphere and may be too hard for 'cheap' satellite studies. The Moon could well become the place to study them. As previously noted, one needs to watch for possible dim lines in the gamma-ray spectrum in the multi-GeV range, and large arrays offer a definite advantage in this respect.

Returning to the high-energy part of the spectrum, the Earth's atmosphere acts as a very valuable converter for photons, and more generally as a calorimeter. One would have to replace it on the Moon by specially built devices. They would be instrumented calorimeters where the stopping material would be made from Moon rocks or dust. The main advantage would be that individual very-high-energy particles could be studied instead of the cascade outcomes which one has to be content with on Earth. One could thus study directly the primary spectrum, measure its composition, and also analyse the dominant features of scattering events at extreme energies, something that is far beyond the reach of even the next generation of colliders. To set the scale, a device 10 000 m^2 in area would be able to reach 10^8 GeV, well beyond the first kink in the spectrum of Figure 4.1. This corresponds to a hundred events per year. All this would provide very valuable information. For point sources, direct access to the incident photons would be a great advantage in view of the puzzles already mentioned.

However rates are then very low (10^{-11} cm^{-2} s^{-1}, at 1 TeV) and extensive detectors are needed to compensate for that. To set the scale, one would need a 100 m^2 area to reach beyond 1 TeV at the level of one event per day. This is good, but one should still reach three orders of magnitude down in rate in order to access the region where the effect of scattering off the cosmic background radiation is particularly marked, as shown in Figure 4.2. One could still see directly what is arriving. It is clear, however, that rate is the first problem, with the direct determination of the primary coming next. For that reason, the advantages of the Earth, with its 'built-in' converter/calorimeter provided by the atmosphere, are hard to beat when it comes to the choice of where to invest the available funding. Nevertheless, the Moon will certainly eventually be used as a base for such studies.

From a Moon base, one could also use the whole atmospheric rim of the Earth as a detector to study the Cerenkov light produced by giant air showers. This could be used to extend the detection of the cosmic-ray spectrum beyond 10^{12} GeV. Such energies correspond to only one event per 10 000 km^2 per year. The rate question is even more important when considering neutrinos. We first consider very-high-energy neutrinos. The Earth's atmosphere is a problem when considering the neutrino background associated with the weak decay of high-energy secondary particles. For the high-energy part of the spectrum, it is not a serious problem. Conditions can be assessed from the measured muon flux, which extends to 50 TeV. The calculated neutrino spectrum falls steeper with energy than the overall primary spectrum. This results from the fact that pions in particular stand a greater chance of being absorbed before they decay, as their energy increases.

For this reason also, atmospheric muons and atmospheric neutrinos have an angular distribution which has a maximum at the zenith, whereas cosmic neutrinos do not. Atmospheric absorption beats the parent–daughter relationship which, for power-law-falling spectra, gives the same drop with energy for primary and secondary particles. When considering muons associated with neutrino reactions in a detector, this atmospheric background ceases to be a problem beyond 100 GeV (Grieder 1989). One may conclude that, for neutrinos with energies of the order of 1 TeV or more, it is the size of the detector that matters. A location on Earth is fine and much progress has still to be made in providing a good case to attract available funding for that research. The present trend on Earth is to use giant water detectors, where the medium provides both stopping power and Cerenkov light. Another possibility being considered is the use of acoustic detectors spread over a quiet ice cap. This is a technique that could eventually be used on the Moon, with acoustic detectors embedded in lunar rock, benefitting from the relative quietness on the Moon. It would be tempting to use a large neutrino detector on the Moon to detect high-energy neutrinos produced by the dump of an accelerator beam on Earth. With beams in the 10 TeV range, corresponding to the coming generation of proton colliders, one will produce intense beams of high-energy collimated neutrinos. Detecting them on the Moon would also provide a great opportunity for neutrino oscillation studies. To set the scale, if one could aim such a 10 TeV beam after absorption, with a 1 km^2 detector on the Moon one could obtain 10 000 events per year (AIP 1989). Practical steering of an accelerator beam, which has to be carried out by means of a construction tangential to the Earth, is, however, limited to Moon rises and Moon sets, and the corresponding duty cycle reduces the counting rate by at least a factor of 1000.

Considering now rather low energy neutrinos, of the order of 10 MeV, the atmospheric (higher energy) and nuclear-reactor (lower energy) backgrounds appropriate to the Earth are not a serious problem as regards solar neutrinos, but would indeed be one as far as the search for neutrino relics of supernova explosions is concerned. The importance of the requisite lunar detectors pushes them into the distant future.

The question of neutrinos in an intermediate energy range, of the order of 1 GeV, for which the atmospheric background is a serious problem, remains. It arises from the combined fall in the primary spectrum with energy and the rise with energy of the reaction cross section. In this case, however, the most interesting problem seems to be the background in proton decay. This is a special issue which is discussed in the next section.

4.4 Proton decay on the Moon

Proton decay bears on a fundamental issue which is a consequence of grand unification schemes which combine quarks and leptons. The relevant energy scale is of the order of 10^{15} GeV. The expected lifetime is accordingly very long. Present results set a lower limit of 2×10^{32} yr for the most easily visible mode. While other decay modes could be dominant, one may say that nothing has been seen, all reported events being at best ambiguous. The present limit only rules out the most simple scenario based on the SU(5) Grand Unified Theory, but not others that attach great importance to values up to 10^{34} or 10^{35} yr. A gain of at least a factor of 100 in sensitivity is required, when the background due to muon and atmospheric neutrinos is already a problem (Rich et al. 1987; Pati et al. 1986). Pati, Salam and Sreekantan have long emphasised the particular advantages that a detector on the Moon would then offer (Pati et al. 1986; Rudaz 1989). To set the scale, one can consider a lifetime of 10^{33} yr. A detector of 10 kt of active material could then yield six events per year. One would have to use a relatively fine-grained calorimeter, but only a few percent of its total weight would have to be brought from the Earth. Most of the detector absorbing material could be assembled from Moon dust.

The incident flux on the surface of the Moon corresponds to a serious background (the flux of 10 GeV protons is as high as 1 $cm^{-2}s^{-1}$). The detector would then have to be completely shielded. However, an advantageous location could then be a tunnel excavated on the side of a crater, since a thickness of the order of a few hundred metres of Moon rocks should provide sufficient shielding to deal with the hadronic background. Though the neutrino background is reduced by the absence of atmosphere, it is not totally negligible. Short-lived particles with weak decay originating from primary protons interacting with the Moon rocks have some time to decay despite the density of the Moon dust, which is of the order of 2 g/cm^3. The Moon dust surface acts as the atmosphere, producing neutrinos. One can, however, estimate that this background, at the crucial GeV level, would be reduced by a factor 200 compared with that on Earth. Direct cosmic neutrinos are not expected to provide a higher background. One could then achieve the factor needed to reach 10^{34} yr and perhaps more, thereby covering a particularity interesting range.

It may well be that an increase in sophistication of the Earth-based proton decay detectors, which one may witness before even a cruder detector can be assembled on the Moon, will be such that this window in lifetime would have already lost its relevance. This could be the case with detectors becoming sensitive to a whole array of decay channels and allowing measurements not equally sensitive to the neutrino background. Yet, the possibility of installing such a detector on the Moon may still look promising two to three decades from now, if the proton decay lifetime is still out of reach.

4.5 A very large accelerator on the Moon?

We are always striving for higher energies and this is bound to continue (Jacob 1990). For a particular technology, reaching higher energies usually implies an increase in the size of the machines − a consequence of synchrotron radiation power for electron accelerators, and of magnet performances for proton machines. The present LEP(LHC) tunnel has a circumference of 27 km; the circumference of the SSC is three times as large. Accelerator operations call for a good vacuum. For such 'real-estate' and vacuum reasons, the Moon has sometimes been mentioned as offering an advantageous location for accelerators in the distant future. In this perspective of high-energy physics on the Moon, one should also consider the continuation of accelerator high-energy physics on the Moon. However, land-area and geological factors are not the present limitations, once cost constraints have been overcome. Tunnelling is currently no more expensive than cut and fill. Giant machines do not require a very large flat, hard and stable surface area. They can typically be installed a hundred or more metres underground, even beneath populated regions. The present limit on future machines is rather that of synchrotron radiation power, which increases as E^4/R, where E is the beam energy and R is the machine radius. The LEP is probably the largest circular electron machine ever to be built (100 GeV beams).

With protons, the same limitation should occur not much above the SSC energy (20 TeV beams). Indeed, synchrotron radiation at the SSC already represents a good fraction of the power that has to be removed from the cryogenic system. Great difficulties would have to be overcome in increasing the energy, and size, by a factor of two. The high-energy machines of the future will be colliders. For power-limitation reasons, they will use superconducting magnets and/or superconducting accelerating cavities, as do those of the present generation. Vacuum then comes relatively cheaply with low temperatures, and large-scale cryogenics should not give rise to major problems. One may try to overcome the synchrotron radiation limit with linear colliders. Here the problem is luminosity, which has to be obtained without the high-frequency multiple bunch crossings of a circular machine. For physics reasons, the luminosity must increase as the square of the centre-of-mass energy. A fully satisfactory technology for high-luminosity linear electron colliders in the TeV range is not yet at hand. It therefore appears that our progress towards much higher energies is technology-limited rather than size-limited. New ideas and much research and development work is needed to find ways of obtaining large energy increases per metre for intense beams, and of focussing them properly. This represents Earth-based development and construction work for the foreseeable future.

4.6 Role of Europe

Europe clearly has a role second to none in this field. The collaborations within CERN (CH) and now at the Gran Sasso Laboratory (I), have established great expertise in recent years in all relevant fields, and the prospects are therefore excellent for lunar or space experimentation.

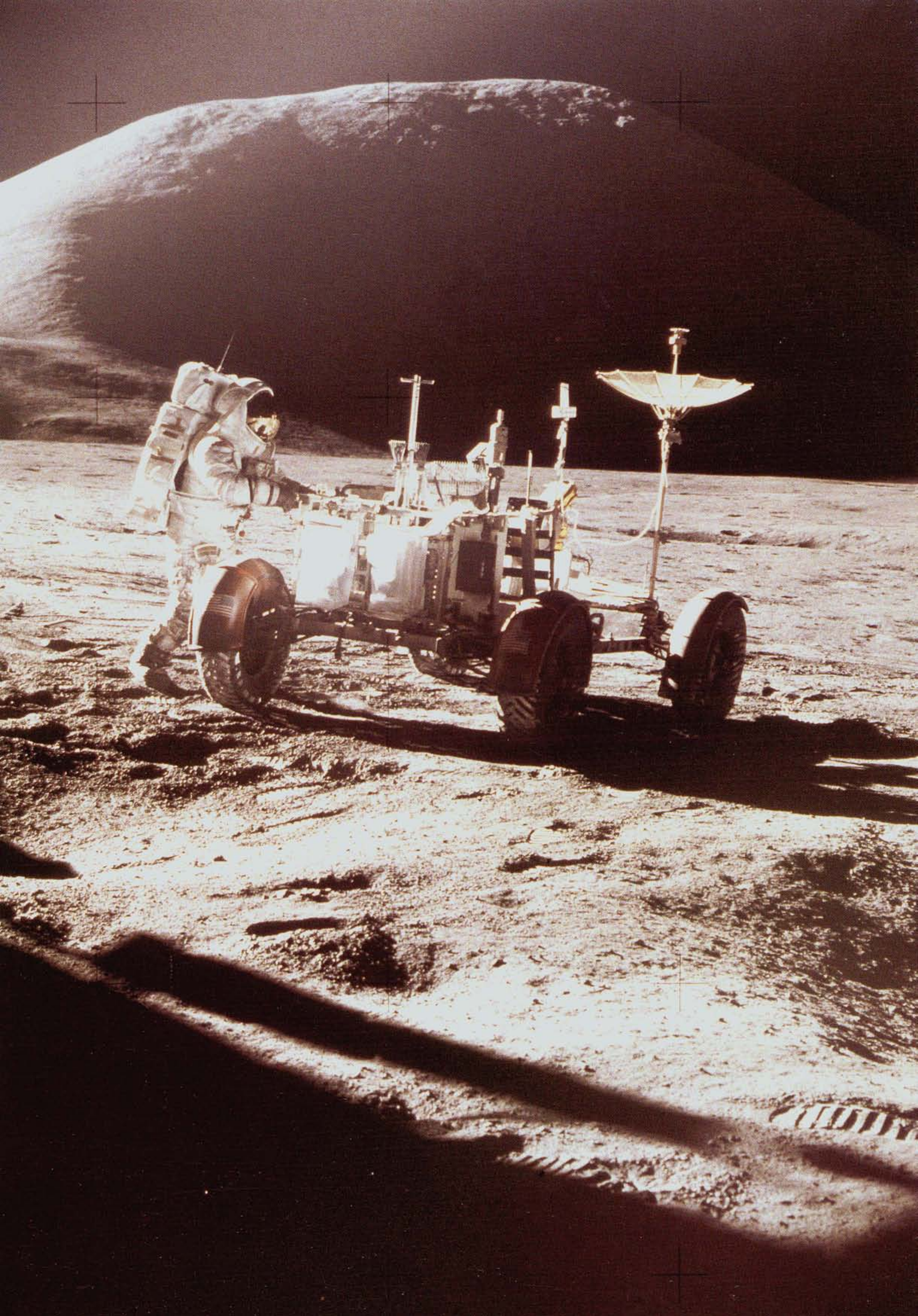
In the foreseeable future, high-energy physics research at a lunar base will be restricted to cosmic-ray-related studies. This field of research is presently changing gear, with investigators in that domain coming from traditional cosmic-ray research but also from particle-accelerator and nuclear physics. Sophisticated detection techniques, first developed with accelerator-based research, are now being applied to large-scale cosmic-ray and underground detectors. European groups are very active and Europe-based work is a very important element of that research at the world level. Many European groups have already had the successful experience of world-wide collaborations, which are likely to set the style for the scientific exploitation of the Moon.

4.7 Conclusions and recommendations

When considering ambitious, large detectors, the limitations are basically due to the detection rates. The atmosphere is often more of an asset than a problem. This research will therefore remain predominantly Earth-based for many years to come. The construction and exploitation of promising detectors is actually only just starting. Satellite studies will quickly develop with smaller detectors for composition studies, new-particle searches, and high-energy photons in the multi-GeV range, as this is the most cost-effective way to proceed. Some of these detectors could, however, also be advantageously deployed on the Moon when benefitting from the implementation of other programmes there. This is the case particularly for large gamma-ray detectors which, once installed on the Moon, could be operated by remote control and read or recorded only very occasionally. They are therefore likely to form part of the early scientific exploitation of a lunar base.

Whereas the Moon appears to be a very interesting location for the study of very energetic events, experimentation on the Earth generally appears more advantageous for the foreseeable future. The atmosphere acts as a useful converter and is thus more a welcome advantage than an embarrassing screen. Big surprises, if any, are likely to lie in the interactions of very energetic neutrinos, for which the atmosphere is transparent. The study of high-energy cosmic rays is a domain of research which is presently shifting gears, with several ambitious detectors being built or in the concept stage. Its future is Earth-based, despite the advantages that the Moon may offer. However, when large detectors are eventually installed on the Moon, they will offer very interesting possibilities. Proton decay and neutrino oscillations may still be among the fundamental questions to be answered. While the dominant part of future cosmic-ray studies will concentrate on large detectors on the Earth, many of them under water or underground, specialised modestly-sized detectors will certainly be deployed on the Moon as soon as this becomes feasible.

5. Life Sciences



Chapter 5. Life Sciences

5.1 Introduction

Despite the fact that the lunar environment lacks essential prerequisites for supporting life, lunar missions offer new and promising opportunities to the life-sciences community. The establishment of a lunar base with a permanent human presence is a challenging project for life scientists working in various fields. Among the disciplines of interest are exobiology, radiation biology, ecology and human physiology.

In exobiology, studies of, on and from the Moon will contribute to the understanding of the principles leading to the origin, evolution and distribution of life. Currently, two alternative sources of organic matter are being discussed for the emergence of life on Earth: terrestrial production of organics or extraterrestrial import. The discovery of organic molecules on the Moon would support the model scenario reflecting the delivery of extraterrestrial organic molecules to the primitive Earth. Radioastronomy, by searching for other planetary systems, especially for Earth-like planets, e.g. signals for water vapour, and the Search for Extra-Terrestrial Intelligence (SETI), may provide clues regarding whether the Earth is unique in the Universe in supporting life, or whether life is a universal planetary phenomenon. The unique environment of the Moon offers an ideal platform for a variety of exobiological studies, such as on the role of radiation in evolutionary processes (organic, chemical and early biological evolution), on the environmental limits of life, as well as on life under simulated planetary environments (e.g. Mars or early Earth).

For radiation biology, the Moon provides a unique laboratory with built-in sources for optical as well as ionising radiation to investigate the biological importance of the various components of cosmic and solar radiation. In any space mission, especially beyond the geomagnetic shielding, the radiation is of major concern not only for biological systems including humans, but also for micro-electronic equipment. Therefore, before establishing a lunar base, precursor missions will be required in order to characterise the radiation field, to determine depth dose distributions in different absorbers, and to qualify the biological efficiency of this mixed radiation environment. In addition, the installation of a solar-flare-alert station on the Moon would clearly be of benefit for radiation-protection purposes in the context of overall space activities.

One of the most challenging projects which falls into the domain of ecology will be the establishment for the first time of an artificial ecosystem on a celestial body beyond the Earth. From this venture, a better understanding of the dynamic processes regulating our terrestrial ecosystem is expected. The maintenance and propagation of life at an altered g-level and with increased radiation with limited shielding is regarded as a fundamental scientific research project. For a lunar base, a reliable Life Support System (LSS) including food supply, gas regeneration and waste management is a condition sine qua non. Considering the tremendous regenerative power of biological systems, a bioregenerative LSS concept should be gradually incorporated into the artificial environment of a lunar base. Precursor missions are required to establish and optimise artificial ecosystems using lunar resources (soil, sunlight, oxygen, and possibly water) and to obtain detailed knowledge about the adaptive stategies of organisms (micro-organisms, plants and animals) to the lunar environment (e.g. increased radiation, reduced gravity) in multi-generation exeriments.

A lunar base which will eventually have a pemanent human presence may be considered as a first step towards living on another celestial body. Building a manned lunar base that becomes gradually more autonomous will raise various problems in the fields of human physiology and health care, psychology and sociology. It will be necessary to provide a radiation monitoring system and guidelines for radiation protection, to provide health care, including the medical, psychological and sociological aspects, and finally to build up a database involving human factors with respect to a lunar base. This database should be exploited and applied for the health, safety and well-being of the astronauts in order to optimise living and working environment on the Moon. The results should also be made available to architects, engineers, agencies, politicians, policy and law people and other decision-makers and constructors.

These objectives can be achieved using a step-by-step approach compatible with the strategy described in Chapter 1 utilising lunar orbiters, to monitor the radiation history during trans-lunar cruise and in the vicinity of the Moon, and to provide information about resources on the Moon's surface. Automated surface stations including rovers would serve several disciplines: exobiology (lunar-sample analysis, cosmic-dust collection and analysis, in-situ experiments, SETI), radiation biology (long-term monitoring and shielding, dosimetric network, solar-flare-alert station), ecology (simple artificial ecosystem, e.g. micro-organisms, plants). Several of these experiments could be improved by returning samples to Earth for detailed investigation. The most sophisticated requirements have to be met for setting up a manned lunar base. Precursor automated stations should be made available in order to define feasible concepts for an efficient biological LSS and optimum radiation protection. Studies on animals should precede long-term stays on the Moon by humans. Lunar-base concepts should consider both a habitation area providing optimum conditions for human health and well-being, and an operational laboratory for extensive research in the different areas related to the Moon. To determine the boundary conditions, substantial research in human physiology will be needed.

The majority of life-sciences experiments on the Moon will be improved if human activity is directly involved. Examples of this are collection of appropriate lunar samples, setting up and analysis of in-situ experiments, refurbishment and repair of equipment. Human settlements on the Moon, per se, will be a scientifically challenging task.

European life scientists have gathered substantial experience in space-related research since their first participation in the Apollo-16 mission. In the exobiology domain, European scientists are involved in several space experiments on the impact of the extreme space environment on biological integrity, including the utilisation of ESA's Exobiology Radiation Assembly (ERA) core facility on Eureca. Since the Apollo missions, European radiobiologists have gathered data on the biological importance of the heavy particles of cosmic radiation (e.g. Biostack and Biobloc experiments, Dosimetric Mapping, CIRCE) outside and inside the geomagnetic field. The ESA study 'Protection Against Radiations in Space', has provided estimates of space-radiation hazards and their biological effects on manned space missions. Within the last decade, Europe has been increasingly involved in studies of artificial ecosystems and bioregenerative life-support systems, especially in Germany, France, and under the auspices of ESA. There is a well-developed expertise in 'man in space' research, in both fundamental research and medical operations, as demonstrated by the European manned space missions to date. This experience in the various fields will serve as an excellent basis for starting life-sciences endeavours under lunar conditions.

Within the overall scenario of lunar exploration, the interest of life sciences is twofold: on one hand, the Moon provides an ideal platform for conducting various new types of investigations, some of which cannot be done elsewhere (e.g. the establishment of an artificial ecosystem on the Moon will be a challenging project per se). On the other hand, given a lunar base, the boundary conditions for humans' safety, health, well-being and working efficiency have to be set. Scientific as well as operational aspects have to be determined. It is recommended that ESA should set up a life-sciences study group to investigate the various opportunities and goals for life-sciences research on the Moon, the scientific requirements, as well as the feasibility aspects, taking into account existing hardware concepts (e.g. ERA, Botany Facility, Aquarack).

5.2 The lunar environment as relevant to life sciences

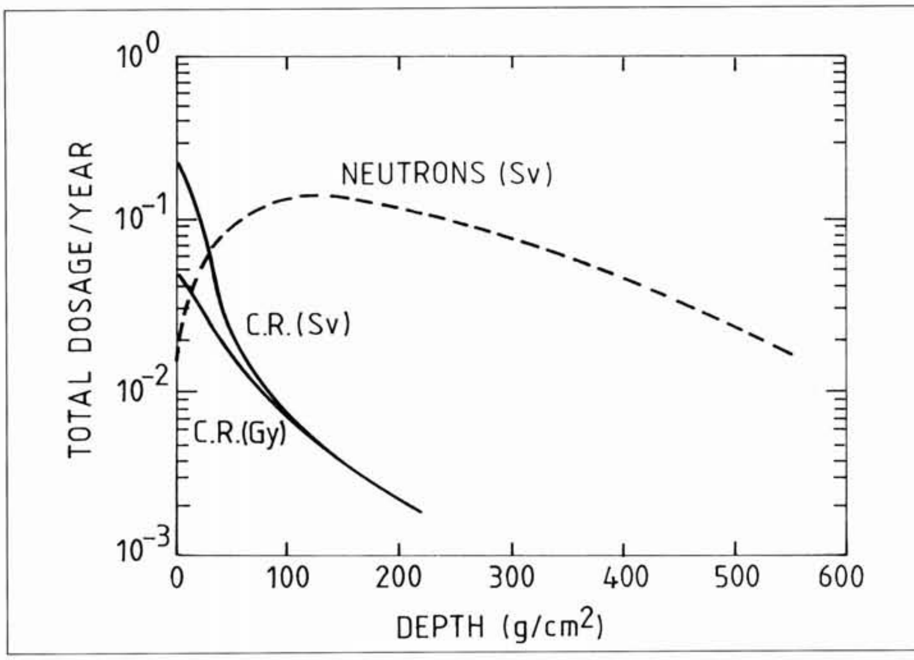
Apart from the low gravity (1/6 g), the absence of any noticeable atmosphere and magnetic field is of special importance in life-sciences terms, because it means that the lunar surface is exposed to all kinds of radiation.

On Earth at sea level there is 10^3 g/cm^2 shielding by the atmosphere. On the Moon, this corresponds to a depth of about 5 m, to obtain a similar amount of shielding by the lunar regolith. At the lunar surface, the galactic cosmic-ray particles give rise to a dose equivalent to about 300 mSv/yr (Silberberg et al. 1985). Higher doses are due to solar-flare particles, most of which stem from one or two gigantic flares each lasting only a few days. Compared with galactic cosmic rays, solar-flare particles have considerably lower energies and thus less shielding is required. The interaction of the galactic cosmic-ray particles (mainly protons with about 10% α-particles and 1% heavier nuclei) produces secondary particles, of which the neutrons are of special importance. At a depth of 1 m, equal to a shielding of 200 g/cm^2, the dose equivalent from galactic cosmic-ray particles is about 2 mSv/yr, but the dose equivalent due to secondary neutrons is about 100 mSv/yr (Fig. 5.1).

The dose equivalent due to γ-rays produced by the interaction of primary and secondary cosmic-ray particles is only about 1% of that due to the neutrons with a depth dependence closely following that of the neutrons. In the upper 100 g of the lunar regolith, about one third of the total γ-ray flux comes from the natural radio-elements K, Th and U. A large number of cosmic-ray-produced radionuclides have been measured in the samples returned from the Apollo missions. The most important of these radionuclides are ^{3}H, ^{14}C, ^{10}Be, ^{22}Na, ^{26}Al, ^{36}Cl, ^{37}Ar, ^{39}Ar, ^{53}Mn, ^{54}Mn, ^{55}Fe, ^{56}Co, ^{57}Co.

In total, the radiation due to these cosmic-ray-produced radioactive nuclides is lower than that of the natural radionuclides. This is true even when adding radionuclides with half-lives too short to be measured in returned samples such as ^{24}Na ($T_{1/2}$ = 16 h), ^{16}N ($T_{1/2}$ = 7 s) or ^{28}Al ($T_{1/2}$ = 2.2 min). The K, Th and U content in the

Figure 5.1 Depth–dose distribution due to secondary neutrons and cosmic-ray nuclei as a function of shielding by lunar soil



lunar surface rocks is similar to that of terrestrial rocks, but the K/U and K/Th ratio is about five times smaller than that of terrestrial rocks.

The Apollo Programme showed that the Moon, like the Earth, is a differentiated object with a crust, a mantle and a small core. The processes that lead to this differentiation, especially the heat source, are not yet understood. The dark lunar regions are impact basins filled with basaltic lavas from deeper regions of the Moon. Compositionally, the mare rocks are similar to terrestrial basalts, except that the FeO content is twice as high, and the opaque mineral ilmenite ($FeTiO_3$) exists in quantities of up to 25% by weight. The lighter lunar highlands are dominated by feldspar. Despite the crater density of the lunar maria being one order of magnitude lower, the samples returned from the Moon showed that mare rocks with ages around 3.6×10^9 years are only slightly younger than the highland rocks with ages around 4.0×10^9 years.

The presence of ilmenite in the mare regions and the pulverised nature (soil) of most of the lunar surface is responsible for the low average albedo of 0.072, making the Moon a rather dark object. Continuous bombardment by meteorites throughout the Moon's history is responsible for the diminution of the lunar surface rocks and their conversion into regolith of very different grain sizes, covering almost the entire surface to a depth of at least 5 m. The projectiles responsible or this diminution process, i.e. meteorites, contribute only about 2% of the regolith. Nevertheless, the metal particles of this meteoritic component, which amount to about 0.5% by weight of the regolith, could provide an important source of metallic iron for various uses.

The size/frequency distribution of the bombarding meteoritic matter leads to a turnover process with turnover times of 10^4 years for the upper 0.5 mm, 10^7 years for the upper 1 cm, and 10^9 years for the upper 100 cm.

The surface temperature of the Moon at low latitudes varies enormously during the lunar day, which equates to 29.53 terrestrial days. At the Apollo-17 landing site, the maximum temperature observed was +111°C and the minimum temperature at the end of the night was −171°C. From diurnal temperature profiles, a mean surface temperature of −57°C was calculated. However, the diurnal temperature variation of 282°C at the surface falls to only 10°C at a depth of 20 cm. At a depth of 40 cm, the variation is about 1°C at an absolute value of about −20°C. Compared with the Moon's surface temperature of −57°C, we find a temperature difference of almost

40°C. This extremely large temperature gradient generated by the heat flux from the lunar interior reflects the very low thermal conductivity of the lunar regolith. For the upper 2 cm of the latter, a very low value for the thermal conductivity of about 1×10^{-5} W/cm °C was calculated, which increases to about 1×10^{-4} W/cm °C at depths exceeding 20 cm. At high surface temperatures, about two thirds of the heat transport in the upper 2 cm is radiative and only on third conductive, indicating an extremely loosely packed soil structure.

Due to the turnover process, individual regolith grains have been exposed to the solar-wind particles. This bombardment is not only responsible for the implantation of hydrogen, ^{3}He and ^{4}He, and all other elements in solar abundance ratios, but also for the destruction of the crystal structure at the surface of the individual grains. In this way, the soil became highly reactive as the effective surface area became considerably enlarged.

The diurnal temperature variation becomes smaller towards the poles, at which temperatures definitely less than −200°C, but perhaps as low as −230°C, are expected at the floors of permanently shaded craters. It has been speculated that water ice from outgassing of the lunar interior and/or as a result of cometary impacts has been stored in such craters.

More details can be found in Chapter 1 and the references therein.

5.3 Exobiology

5.3.1 Scientific background

The main goal in exobiology is to identify the principles leading to the emergence of life from inanimate matter, its evolution and distribution on Earth, within the Solar System or beyond. To reach this goal, research has focussed on the different steps of the pathways that may be taken by the main biogenic elements − carbon, hydrogen, nitrogen, oxygen, sulphur and phosphorus − from their primary chemistry to their assembly in advanced life forms. This chain of evolutionary steps includes: (i) the cosmic chemistry of the major biogenic elements; (ii) the chemical evolution of organic molecules in water; (iii) the evolution of early forms of life; and (iv) the evolution of advanced life, including the search for life elsewhere in the Galaxy. The lunar environment, though it lacks essential pre-requisites for supporting life, may be helpful in achieving a better understanding of items (i) and (iv).

Terrestrial life can be schematically described as 'highly organised organic molecules in liquid water'. For more than sixty years, the dominant scenario referred to the formation of small bio-organic precursors in a terrestrial reducing atmosphere dominated by methane. In 1952, S. Miller exposed a mixture of methane, ammonia, hydrogen and water to electric discharges and identified four of the twenty proteinaceous amino-acids among the compounds formed (Miller 1953). Since this historical experiment, a large number of the biogenic building blocks have been successfully synthesised under putative prebiotic conditions.

Because a strongly reducing early Earth atmosphere is unlikely, geochemists now favour an atmosphere provided from volcanic outgassing dominated by carbon dioxide, nitrogen and water vapour. In a hydrogen-poor atmosphere, the production of biological building blocks such as amino acids, nucleotide bases, or lipids appears to be inhibited. If the early atmosphere was dominated by carbon dioxide and did not contain hydrogen, other sources must be sought for the supply of biological molecules.

Organic chemistry first occurred in interstellar space. So far, more than fifty different organic molecules have been identified by radioastronomy. However, there is little basis for speculating that interstellar organic molecules arrived intact to the prebiotic Earth's surface. Comets are also efficient chemical reactors. The Vega and Giotto spacecraft found that Halley's comet is richer in organic material than was generally expected (Langevin et al. 1987).

About 10 000 t of cosmic dust, derived from a number of parent bodies, reach the Earth's atmosphere every year. Comets and asteroids are believed to be the major sources of such dust, although their relative contributions are difficult to quantify. Dust particles have been collected in Earth orbit, and many of the impacting grains have a carbonaceous chondrite-type composition (Bibring 1988).

The study of meteorites, particularly the carbonaceous chondrites that contain up to 5% by weight of organic matter, has allowed close examination of extraterrestrial organic material. Seventeen amino-acids were found in the Murchison meteorite (Lawless & Peterson 1975). Some purine and pyrimidine bases, hydorcarbons, and boundary-forming fatty acids (Deamer 1986) have also been identified (Hayatsu & Anders 1981; Mullie & Reisse 1987). Micro-meteorites, also referred to as cosmic dust or interplanetary dust particles, have been extracted from deep-sea sediments (Brownlee 1985), from cryoconite (black sediment) collected from the melt zone of the Greenland ice-cap (Mauretta et al. 1987) and directly from Antarctic old blue ice (Mauretta et al. 1988). A high percentage of unmelted chondritic micro-meteorites ranging from 100 to 1000 μm in size have been observed, indicating that many particles cross the terrestrial atmosphere without drastic modification due to thermal treatment. Most of the unmelted micro-meteorites are composed of porous aggregates of tiny grains embedded in an amorphous component which is likely to be made of a carbon-rich material. Two independent estimates (Anders 1989; Chyba et al. 1990) of extraterrestrial organic infall provide a confirmed value of 10^{20} g over 100 million years (20 g/cm^2).

Chemical evolution and solar-system formation appears to occur throughout the Universe, which may therefore be rich in planets populated with life forms and perhaps intelligent beings with abilities at least comparable with our own. Using radio waves, the most efficient vehicles for information propagation across interstellar space, it should be possible to communicate over distances greater than a thousand light years.

5.3.2 Scientific objectives

■ To search for organic matter below the surface layers

Since a significant atmosphere is lacking on the Moon, the meteorites impact at their full cosmic velocities. Safe deceleration of the infalling particles is an important question. The top few centimetres of the regolith are exposed to the temperatures and proton fluxes of the lunar surface. Hence, this is the least likely place to look for lunar organic compounds. If organic molecules were delivered to the early Moon, they may still be present beneath the lunar regolith at depths of 10 m or more, as suggested by Sagan (1972).

A laboratory on the Moon would allow the analysis of a wide variety of lunar sample types and perhaps meteoritic material in pristine condition. The problem of terrestrial contamination, which in the past has been a major concern in the analysis of returned lunar samples and meteorites especially with regard to the search for organic compounds, would thereby be overcome.

■ To collect cosmic dust particles for organic-chemical or biological analysis
The non-destructive capture of cosmic dust particles is required for studying the occurrence and chemistry of biogenic elements and molecules in these particles. The Moon provides a sterile platform, uncontaminated by terrestrial material, the for collection and analysis of those individual particles. However, special means are required to slowly decelerate them.

■ To study prebiotic and early biological evolution processes by exploiting the Moon's extreme environment
The Moon's environment is dramatically different from that of the Earth. Its gravity is about 1/6 th that of the Earth, it has a high vacuum, a large difference between day and night surface temperatures, and a high influx of unfiltered solar UV-radiation and ionising radiation, such as galactic cosmic rays, and particles generated by the solar wind or solar flares. This unique environment offers ideal conditions for a variety of biological studies, including: environmental limits of life; role of radiation in evolutionary processes (organic-chemical evolution, early biological evolution); life under simulated planetary conditions, e.g. UV-climate of Mars and early Earth; life under reduced gravity.

■ To utilise astronomical instruments for studies of exobiological interest
It is still debatable whether the Earth is unique in the Universe in supporting life, or whether life is a universal planetary phenomenon, provided liquid water and organic molecules are present. This question can be partly studied by exploiting astronomical instruments which, due to the unique lunar environment and the stability of the Moon's surface, can be used in large arrays to achieve improved spectral and spatial resolutions. Of special exobiological interest is the detection and characterisation of other planetary systems, searching for Earth-like planets, e.g. signs of water vapour and artificial radio emissions from other advanced civilisations (SETI).

5.3.3 Technical requirements
For the analysis of samples collected on the Moon, the major techniques of chromatography and mass-spectrometry will be used to separate, identify and quantify the compounds of exobiological interest. This analytical equipment might be concentrated in an exobiological laboratory. For the sampling of materials such as regolith from different depths, meteorites or cosmic dust, well-developed geological and dust-collection techniques can be used, perhaps slightly modified in some cases.

The in-situ experiments on the Moon will require specific exposure facilities to allow access to all selected parameters of the lunar environment, to prepare and manipulate the test samples, and to analyse the effects after exposure.

For the astromical observations, large telescopes and radio-telescopes, interferometers and arrays, with sophisticated detectors and using advanced data handling, processing with customised hardware and software, will be required (Chapter 2).

5.4 Radiation biology and radiation protection

5.4.1 Scientific background

The Moon possesses neither an atmosphere nor a magnetic field, and so its surface is exposed to the full spectrum of galactic and solar radiation. A lunar base – either manned or unmanned – thus offers unique possibilities for studies in photo- and radiation biology. The space radiation field could be monitored and analysed in detail, in terms of both composition and temporal variation. These data would constitute invaluable input for risk assessment in further space activities.

Radiation biology on the Moon has to be looked at from two aspects, a fundamental and a practical one. In the first case, a lunar station can be considered as a laboratory with built-in radiation sources covering the whole solar ultraviolet, visible and infrared spectrum as well as solar and galactic protons, helium and heavier ions. Of special importance for radiation biology are the HZE-particles, which are unique and cannot be simulated on Earth (reviewed by Horneck 1992). They are defined as cosmic-ray primaries with charges $Z > 2$ and energies high enough to penetrate at least 1 mm of spacecraft or spacesuit shielding (Grahn 1973). On the other hand, radiation constitutes a significant hazard not only to humans, but also for micro-electronic equipment. A comprehensive investigation of radiation components and their action in physical and biological terms is hence indispensible not only for the health of people working at a lunar base, but also for the proper functioning of automated stations.

The knowledge of radiation conditions on the Moon is still very incomplete. Table 5.1 compares the lunar radiation climate with that on the Earth's surface. It may be assumed that the optical spectrum is similar to that in free space. Measurements of the ionising component made during the Apollo missions showed wide variations in daily doses, ranging between 0.22 and 1.27 mGy/day (Benton & Parnell 1988), as determined by thermo-luminescence dosimeters at the astronaut's skin. Since on the Earth the annual dose amounts to 1–2 mGy, 2–5 days on the Moon are equivalent to one year on the Earth in radiation-exposure terms. In the dosimetric data from the various Apollo missions, the fluxes of HZE-particles varied by a factor of nearly 2, largely but not exclusively due to changes in solar activity, which modulates galactic cosmic rays. The total dose may be broken down into contributions of different components (Schaefer et al. 1972), showing that 75% is produced by protons, 15% by electrons and gamma rays, 7.5% by spallation 'stars', 2.5% by HZE-particles and 0.5% by fast neutrons. According to this analysis, the mean quality factor (using present recommendations) is 2. The dose quivalent at the lunar surface has been calculated as 250 mSv/year (0.68 mSv/day), a value that should be compared with the measured results. The large variations demonstrate that there is a clear need for more systematic measurements.

Table 5.1. Radiation fields on the Moon compared to those on Earth

Radiation	Moon	Earth
Mean Total Background Dose	0.22—1.27 mGy/day	1—2 mGy/year
Cosmic Radiation	~95% of mean total dose	30% of mean total dose
Galactic Cosmic Radiation (GCR)		~100% as secondary particles of cosmic radiation field
• Flux	~3/cm^2.sec	6.4×10^{-3}—1.9×10^{-2} photons/cm^2.sec
• Energy	10^8—10^{20} eV/nucleon	Wide range: 0.1—20 GeV (muons) 10^{-8}—10^2 MeV (neutrons)
• Composition	98% nuclei consisting of: 87% protons 12% α-particles 1% HZE (Z>2) 2% electrons	 ~80% muons ~20% electrons <1% of primary cosmic-ray particles (protons, charged pions, neutrons)
Solar Cosmic Radiation (SCR)		
• Flux	~10^{-3}—10^2 particles/cm^2.sec	Minor contribution to cosmic background radiation on Earth's surface
• Energy	<10^{10} eV/nucleon	
• Composition	<98% protons, 1—4% α-particles, 0.01—0.1% HZE	
Solar-Wind Particles		
• Flux	10^8/cm^2.sec	Will be deflected by the Earth's magnetosphere
• Energy	<4×10^3 eV/nucleon	
• Composition	Electrons, protons, α-particles	
Solar-Flare Event (up to 10/yr)	0.4—0.6 Gy/hr (during flare event)	
• Flux	Up to 1000/cm^2.sec	
• Energy	10^6—10^9 eV/nucleon	
• Composition	Protons, α-particles, heavy ions	

Biological samples were flown on trans-lunar trips in a Biostack arrangement with Apollo-16 and 17 (Bücker & Horneck 1975) and effects were traced back to the traversal of single HZE-particles of cosmic radiation in a variety of biological specimens, such as bacterial spores, plant seeds and animal eggs. Most effects point to damage to the genetic material, such as mutations, tumour induction, chromosomal aberrerations, cell inactivation or development abnormalities. For inactivation of Bacillus subtilis spores, cross-sections of about 5 μm^2 were determined that largely exceed the geometrical cross-section of the spore protoplast of about 0.25 μm^2 . From these data a superimposition of two different inactivation mechanisms is inferred: a short-range component reaching up to 1 μm that may be traced back to the δ-dose, and a long-range one that extends at least to somewhere between 4 and 5 μm off the particle's trajectory.

5.4.2 Scientific objectives

Studies in the following fields need to be performed:

■ Characterisation of the radiation field on the Moon, its variation, and its modulation by different absorbers

The composition of the radiation field and its temporal variation should be measured, paying particular attention to the dependence on solar activity. Solar flares, especially anomalously large events, should be carefully recorded to facilitate means for their prediction and early warning. The production of secondary radiation should be assessed by measurements under defined shielding materials with special regard to spallation events ('stars') and nuclear fragments as well as neutrons. Detectors should be such that the energy spectra of all primary and secondary radiations can be recorded (Table 5.1)

- **Establishment of a solar-flare-alert station**

Intense particle radiation due to solar-flare events builds up to a peak within a few hours or less. Monitoring of X-ray precursors provides between 30 min and 4 h of warning. A computer-based prediction system has to be established taking into account careful registration of heavy nuclei and the temporal development of their energy distribution. As a central element of a solar-flare-alert station, such a system would be of immense value for the protection of humans engaged in all kinds of space activities.

- **Study of the effect of solar radiation on biological systems**

The intensity and spectral composition of the solar UV transmitted to the Earth's surface are controlled by the stratospheric ozone layer. A decrease in stratospheric ozone leads to an increased contribution of shorter wavelengths (UVB: 290 nm $< \lambda <$ 320 nm). Because of the nonlinear wavelength dependence of biological reactions, these changes in UVB level may have important consequences both for humans and for the species composition of ecosystems. The damage to the major biological target molecule DNA from UV radiation at 290 nm is 10 000 times greater than at 320 nm. A lunar outpost would provide an ideal laboratory for investigation of the biological effects of the full solar spectrum and of its modification by simulated ozone layers of different depths. The establishment of an artificial ecosystem on the Moon would allow the consequences of possible changes in the solar spectrum for the development of life on Earth to be assessed.

- **Biological effects of solar and galactic ionising radiation**

An attempt should be made to establish a long-term colony of biological systems (ranging from tissue cultures to suitable vertibrates) which may serve as a model system for man on the Moon. Long-term effects of the combined action of protons and HZE-particles in conjunction with reduced gravity can be studied. The main emphasis should be placed on teratogenic and genetic alterations, as well as tumour formation and reduced life expectancy.

- **Assessment of radiation hazards for man on the Moon**

The lunar regolith material is primarily composed of SiO_2($\sim$40%), TiO_2 ($<$7.5%), Al_2O_3 ($<$27%), Fe_2O_3 ($<$4%), FeO ($<$15.7%), MgO ($<$9.5%) and CaO ($<$15.7%). Its mean density is approximately: 1.4 – 1.9 g/cm^3. When the recommended radiation shielding is more than 400 g/cm^2, this would result in a regolith layer of 2 – 3 m. At the time of a rare giant flare event, shelters with $>$700 g/cm^2 of regolith (more than 4 – 5 m) are required for successful radiation protection (Neal et al. 1988). These shielding requirements result from the depth/dose distribution in lunar soil with respect to cosmic-ray particles and secondary neutrons (Fig. 5.1).

The relative abundance of radionuclides in lunar material is not only of fundamental scientific interest, but would also provide a database for the assessment of the possible hazard posed to spacecraft components by induced radioactivity.

5.4.3 Technical requirements

■ **Radiation measurement**

The dosimetric devices should consist of active dosimeters and spectrometers with sufficient, but not extreme, sensitivity and resolution. They should be installed in the orbiters, in stations on the lunar surface, e.g. networks, and at different depths in the shielding. Proportional and solid-state counters will presumably suffice. To detect spallation and fragmentation products, volume detectors are necessary. AgCl crystals appear to be a good choice in this respect, although they are not tissue-equivalent. An alternative are cloud chambers operated in continuous mode. The whole equipment can be operated by telecommand and does not require a permanent human presence ('man-tended').

■ **Solar-flare-alert station**

This can be a part of the radiation measurement system, supplemented by appropriate data handling and analysis. It can also be man-tended.

■ **Biological effects of solar radiation**

To study the action of solar UV on biological systems, appropriate life support is mandatory, such as the supply of nutrients (including water) as well as gas and temperature control. The assessment of chromosomal aberrations, mutations and neoplastic transformation at the cellular level is not possible without permanent human observation. Skin-tumour formation in hairless mice can be studied in 'automated' animal colonies that are regularly visited by scientists. UV effects on aquatic artificial ecosystems may be monitored by telescience if maintenance without outside control proves to be possible.

■ **Biological effects of cosmic radiation**

The same conditions as above apply here. Cellular studies of active systems cannot be performed without human handling. Tumour and longevity investigations may be carried out 'automatically' with regular checks and scoring by scientist visitors. Some simple biological dosimeters could be installed for long-term monitoring of the biologically effective radiation dose.

■ **Radiation protection**

Small portable dosimeters should be developed in order to monitor continuously the radiation exposure of humans on the Moon.

5.5 Artificial ecosystems and biological life-support systems

5.5.1 Scientific background

The establishment of a lunar base will lead to long-term stays by humans in an isolated artificial environment which not only has to fulfil all the basic needs for survival, but also provide suitable living and working conditions. The development of a reliable life-support system, including food supply and waste management, is therefore mandatory for human habitation. In the initial phase, this life-support system may be 'alimentary', i.e. with all water, respiratory gases and food provided from the Earth. For long-term purposes, however, this approach would result in tremendous logistical problems. The life-support system has therefore to become more and more 'regenerative', utilising physical and chemical methods to recover water and the atmosphere. A more or less autonomous food supply, however, cannot be managed with these methods. Considering the tremendous regenerative power of biological systems, a step-by-step introduction of plant food production in combination with biological atmosphere regeneration, water recovery and biological waste management will be a condition sine qua non. A lunar base would provide an excellent opportunity to develop and test sophisticated Bioregenerative Life Support Systems (BLSSs) until they are reliable enough to replace alimentation and/or possibly dangerous chemical methods.

The scenarios for lunar or martian bases developed in the USA, USSR and Japan are primarily based on physico-chemical life-support systems, but include the step-by-step integration of biological systems. NASA presently favours the concept of Controlled Ecological Life-Support Systems (CELSSs), dedicated to the production of food for astronauts mainly via the controlled growth of higher plants (MacElroy et al. 1989). There are several approaches to higher plant CELSSs which have resulted in a widespread research programme in the USA, including the breeding and selection of suitable plant species and the development of related hardware. One of the most impressive projects is the 'plant growth chamber' ('breadboard facility') at Kenedy Space Center, which is able to deliver the daily food for 'almost' one astronaut. This clearly demonstrates the difficulties of a fully biological life-support system. In the USSR it was shown with the BIOS-3 experiment by Gitelson and co-workers (1989) that it is possible to maintain humans for longer periods of time in a closed environment by not only producing higher plant food for sufficient nutrition, but also oxygen for respiration by means of air/hydroponic (substrate-bound or substrate-free plant growth in nutrient solutions) cultures. Gitelson proposes the development of artificial biospheres for biological life support in the context of a long-term international programme, and his approaches seem to be very realistic for future developments. The NASA concept for a lunar base (Cipriano & Ballard 1990) includes plant CELSS as an essential element; the scientific endeavours of Salisbury (1991) regarding lunar farming are an important step in this direction.

The Biosphere-2 project near Tucson, Arizona (USA) is the most ambitious approach to creating an isolated 'second biosphere', enclosing about 13 000m^2 of ground in a giant greenhouse containing several biomes, ranging from tropical rain forest, to an ocean and marshlands, to the desert. Eight humans are disposed to live there in total isolation from 'Biosphere-1' – i.e. the Earth – for two years without any input except energy. The life-support concept of Biosphere-2 involves not only plants, but also animals for food production, such as teleosts (aquaculture system), birds (chicken) and mammals (dwarf goats, pigs). This project might be the precursor for a lunar or planetary base in some respects, but it has to be pointed out that it is designed for normal Earth-gravity conditions (Allen 1990). A paper study by NASDA has elaborated a similar concept for a highly sophisticated biological life-support system involving micro-algal cultures, higher plants and aquatic and terrestrial animals, which is intended to be verified in three evolutionary phases (CELSS Experiment Concept of Space Station Missions, 1984). Nevertheless, the introduction of biological life-support systems can be very complicated in general, due to possible instabilities, plant and animal diseases, etc., and will only be possible using a step-by-step strategy in parallel with physico-chemical systems. This requires a long-term research and developmental programme which can only be verified through international cooperation. On the other hand, the Soviet BIOS experiments showed the biological systems to be more stable and reliable than the technical systems.

BLSS may be divided into two main categories. The first is represented by the type defined by NASA as CELSS, which has as its main purpose the production of higher plant food for the astronauts. The problems with this approach are mainly threefold. On the one hand there is the problem of plant species selection and breeding. Higher plants growing on Earth show altered physiology under reduced gravity and it was reportedly a very difficult task, taking about two decades, to grow a higher plant from seed to seed in space. This is a very important and limiting problem because in any event BLSSs have to be 'reproductive' systems which continuously produce biomass. The second problem with CELSS is the energy supply. For photosynthesis, the plants require a certain amount of light which cannot be provided economically with the power available at a space, lunar, or planetary station. Therefore possibilities for the optimal utilisation of solar energy have to be developed, based for example on the existing Himawari collector (a solar light collector using Fresnel lenses, which only cover the spectra from 400 to 760 nm, thus excluding UV and IR radiation; the solar light is distributed to plants by fiber optics). The third problem is the nutrients for the plants. In the existing CELSS models, both hydroponics and soil culture are verified. In the case of hydroponics, it will be difficult to elaborate self-regenerating systems, whereas in the case of soil-based higher plant cultures a complicated waste-processing and recycling system is necessary. There are several promising approaches in this context, such as the 'soil-bed reactor' for the elimination of organic compounds in the atmosphere by soil micro-organisms (Frye & Hodges 1990). In any case, the recycling procedure has to involve the waste from the crew also. In the BIOS-3 experiment, this was successfully achieved for more than two years with urine and the water from feces, without microbial infection problems affecting the crew.

The last point provides a link to the second category of BLSS. These may, but need not necessarily, produce food for the crew, although for economic reasons this role should be of primary interest. They serve, moreover, for the regeneration of the atmosphere, consuming the carbon dioxide produced by the crew during photosynthesis and thus delivering oxygen for respiration. In this connection, however, it should not be forgotten that plants in the dark consume oxygen by respiration. In this case, microalgal bioreactors seem to be the most promising approach, but the USSR's BIOS-3 experiment has already shown that the use of higher plants is also a practicable means of atmospheric regeneration. The introduction of photo-heterotrophous microalgal species provides additional capabilities for optimising such systems and combining algal food production with that of oxygen.

The problem of food production raises the question of whether this should be restricted to plant food. Although there is in principal no necessity to convert plant biomass into animal biomass, the 'human factor' might be a reason to make an attempt in this direction. Particularly on long-term missions lasting more than a year, there might be a basic demand for animal food if the lunar or planetary base crew does not consist specifically of vegetarians. The social event of a 'meal' together with other people is an important factor in life and this also involves the aroma and the view of a well-arranged table with different kinds of food, insofar as this is possible in reduced gravity. In this respect, fish production in aquaculture seems to be the most promising and practicable approach. It would be possible to construct intensive aquaculture systems for fish-protein production on a space station or at a lunar or planetary base.

5.5.2 Scientific objectives

Comprehensive BLSSs are complex networks of biological and technical systems because the single components need control and data-recording devices for scientific and safety purposes. Moreover, the interdependencies and interactions of the components have to be coordinated, stabilised and automated by complicated process-control systems. In addition, the systems have to be redundant and must be supplemented by physical/chemical emergency instrumentation. In the ideal case, a BLSS fulfils Gitelson's perspective of an artificial biosphere which works as an autonomous closed system. This in itself consists, however, of several subsystems with the biological ones as a rule representing artificial biomes or artificial ecosystems. Even an agricultural area is an artificial ecosystem, but those of a BLSS (or the BLSS as a whole) are different because of their characteristic of 'closedness'. This closedness may be total or partial, the latter case being the most common. As a rule, some waste products cannot be recycled and even the artificial biosphere needs energy input from the environment.

Hence the basics of BLSS research are more or less closed artificial ecosystems. Basic artificial-ecosystem research can, however, be performed in two different ways. The first step involves the so-called 'physical model', where an artificial ecosystem with a limited number of parameters is constructed by the scientist, who is able to monitor and to observe it for long periods. This descriptive approach must be supplemented by a manipulative one in which defined system parameters are changed and the system's reaction is observed. This experimental approach is very complicated in most cases because very complex monitoring and data-acquisition systems have to be implemented. The higher the number of system parameters, the more the complexity increases. In long-term multi-generation experiments, for example, this is the only way to study adaptations of organisms and evolutionary alterations. The other step is the 'mathematical model', which is often referred to simply as 'modelling'. In this case an ecological system is 'constructed' with mathematical equations in a complicated computer simulation. The difficulty of this approach lies in the fact that linear equations are very often used to describe nonlinear processes (most biological processes are nonlinear), and the real handicap is that evolutionary processes can never be predicted by mathematical elaboration. Nevertheless, the mathematical model may be a valuable tool if used in combination with the physical (experimental) model.

From another standpoint, BLSS research can be classified according to three other criteria that have already been mentioned in another context:
- biomass production and utilisation
- element recycling
- system theory, dynamics and control.

These terms summarise the main scientific goals of BLSS research in general, and a major part of the corresponding work can be done via laboratory research on Earth.

For a lunar-base laboratory, the more special scientific goals are threefold:
- basic research on artificial ecosystems under lunar conditions
- applied research to test the suitability and reliability of artificial ecosystems for BLSS utilisation
- step-by step implementation and testing of BLSS for the lunar base.

The main research work will fall in the botanical field because the effective growing of higher plants combines food production, atmosphere and water regeneration in a nearly ideal way, as pointed out above, and seems easily adaptable to lunar soil also. In combination with that, however, microbiology will play an important role both in atmosphere regeneration and in waste management (see soil-bed reactor mentioned previously). With respect to possible animal protein production, the most important field is reproductive biology, including genetics and embryology. The latter in particular will be one of the important objectives of physiological research on the Moon. The effects of reduced gravity can easily be investigated, since the residual gravity facilitates the construction and maintenace of life-support systems, thereby allowing natural mating of vertebrates, even of mammals. These topics close a loop to radiation biology by scanning possible radiation hazards, especially for vertebrates which may serve as models for humans.

5.5.3 Technical requirements

First attempts to solve some of the scientific problems summarised above can already be made using a simple automated surface device (testing radiation effects, especially). A simple automated phytotron (plant growth chamber) for a lunar outpost would be one example of a first scientific step leading to more complex devices. The scientific disposition of the research programme for a lunar station, however, can only be verified in a relatively voluminous life-sciences laboratory. For basic research purposes, a phytotron with a comprehensive monitoring system for all relevant parameters is an essential prerequisite, as is the basic laboratory equipment to perform biochemical analyses such as protein content as well as the quantitative determination of inorganic substances in soil or plant nutrients. The laboratory equipment has to include incubators for microbiological purposes, refrigerators and centrifuges.

For the applied research, phytotrons large enough for food production and atmosphere and water regeneration are needed. Even in this case, comprehensive monitoring systems are essential. Another precondition is a sufficient light supply for the plants, preferably utilising solar energy with suitable collectors. The equipment needed for food processing depends upon the degree of complexity involved. In the simplest case it has to include a biological waste processor, a dryer and a combustion device for non-degradable inedible plant parts.

For animal experiments, habitats are essential both to tetrapods and for fish. The laboratory equipment has to include blood sampling devices and suitable photometers for, for example, enzyme and hormone analyses. Histological processing of frozen or fixed tissue-sample sections is called for, as well as light-microscope examination. In all cases, means for sample storage have to be provided.

5.6 Human physiology and medicine

Man's presence on the Moon can be regarded as a resource because of his key role in the whole operational scenario of the mission. However, the environmental conditions are fundamentally different from those on the Earth insofar as life is only possible in an artificial atmosphere and will be subject to 1/6 g, which is the gravitational force of the Moon.

In this section the operational aspects will only be touched upon, the main concern here being the basic research in the field of human physiology and medicine which can most profitably be conducted at a lunar base. The results of these studies are expected to shed light upon fundamental physiological questions of importance for our understanding of our terrestrial life, but will of course also be important for our understanding of man's adaptability to weightlessness and his ability to make long-duration interplanetary flights.

5.6.1 Scientific background

Man has developed under 1 g conditions and has adapted during his evolution to his upright position, with the effects of changes in body position being counteracted in different ways depending on the nature of the physiological subsystems involved. It is therefore to be expected that the gravity field of 1/6 g on the Moon will influence man profoundly and trigger a string of adaptational processes, the study of which will be a rewarding scientific objective in itself. As far as the different physiological subsystems of the body are concerned, certain gravity-dependent processes have been described for the cardiovascular system and for the pulmonary function. This is also true for the sensory inputs from visual, tactile, and other inputs from skin, tendons and muscles, which together with sensory inputs from the inner ear make up our perception of the vertical.

The fields of interest to be studied on the Moon will thus lie primarily in the neurophysiology and cardiovascular physiology domains, but important study objectives are also to be found in the fields of oxygen metabolism during rest and exercise, calcium turn-over, blood formation, pharmacokinetics, embryology, development, and radiation effects. These same topics have been identified as the most profitable for exploration also during weightless spaceflight. In order to study the adaptational processes during 1/6 g, it is important to identify their gravity thresholds. Measurements of adaptational processes within the different subsystems need to be performed on both man and animals.

The objectives in the basic-physiology field will be important for securing the health and well-being of man on the Moon and permitting him to live and work there. It will therefore be important to study the problems of countermeasures against the ill effects of the reduced gravitational field and to define the medical care that must be provided at the Moon base.

5.6.2 Scientific objectives

■ **Neurophysiology**

Strictly speaking, man only has one gravity sensor, which is the vestibular organ forming part of the labyrinth of the inner ear, the semicircular canals of which constitute our acceleration sensor. The two types of sensors work closely together in the interpretation of sensory inputs for the perception of spatial orientation.

However, other sensory inputs are also important for our perception of spatial orientation, including tactile inputs from the skin (both superficial and deep skin receptors of different kinds), tendon- and muscle-stretch receptors, and last but not least visual inputs. During weightlessness, the space adaptational syndrome developed by most astronauts is explained as a mismatch between the visual and vestibular cues, which generally subsides after 48 h. It is suspected that similar perturbations might be seen on the Moon, depending upon how well the astronauts adapt during their travel between the Earth and the Moon. However, 1/6 g might create a kind of Moon adaptational syndrome analogous to the space adaptational syndrome.

An important neurophysiological question is how the Moon's gravity will influence the motion pattern and how man will adapt thereto, for instance, lifting weights, running and walking, and how other reflexes connected to posture will be modified. It is known from the Apollo missions that man can move in a sort of kangaroo-jump fashion, but this might only be adopted for locomotion during Extra-Vehicular Activity (EVA) on the lunar surface. Moving inside a spacious shirt-sleeve environment might be something different.

■ Cardiovascular system

The heart

When man adopted his upright position during his evolution, it created special problems for the cardiovascular system because the heart is a pressure pump with little or no suction ability. The maintenance of adequate blood pressure for the brain and the heart muscle itself therefore became problematical in that the body's position in relation to the gravity vector is critical for the optimum functioning of the cardiovascular system. During evolution, however, man has developed a regulatory feedback system which counteracts the effects of gravity on the cardiovascular system. The regulatory system is based upon an intricate interplay between systemic and local neural and hormonal processes, which maintains homeostasis within certain limits during changes in body position.

Volume regulation

It is known from previous spaceflights that gravity reduction induces cardiovascular adaptation, which is believed to be triggered by an increase in central venous pressure. However, this pressure returns to its control level within 24 h, while at the same time the circulating blood volume is believed to contract due to increased urine production. Recent progress in the field of volume regulation physiology has shown that this phenomenon is probably connected with the production of the natriuretic polypeptide by the heart and by the peptide urodilatine in the kidneys. Study of this phenomenon under reduced gravity on the Moon will provide valuable supplementary data.

The lung

The lung is highly dependent upon gravity for its functioning because of a gravity-dependent distribution of the ventilation perfusion ratio. The lowest parts of the lungs with respect to the gravity vector will be blood-filled, while the upper parts contain more air because the traction of the weight of the lungs keeps the alveoli distended. The upper parts receive less blood because of the low pressure in the lung artery. This gives a complex function to the lung which is not yet fully understood. Study of these phenomena during weightlessness is very appealing to the lung physiologist because in that situation the intrinsic qualities of the cardiopulmonary system will dominate the lung function. Preliminary results from NASA's SLS-1 mission in May 1991 unexpectedly showed that not all signs of ventilation-distribution inequalities had disappeared (Blomqvist, pers. comm. 1991). This makes it even more interesting to study this phenomenon under reduced gravity.

Peripheral circulation
Very little is known about the adaptation of the peripheral cardiovascular system to weightlessness. It can be inferred from the lack of hydrostatic pressure that the mean capillary pressure in those parts of the body below the level of the heart increases, while the reverse is the case above it. The capillary pressure is close to being equalised in the supine position, but important hydrostatic gradients remain. When standing, sitting or walking, blood is redistributed and the capillary pressures vary with the functioning of the body. The area below heart level can be characterised by a high-pressure vascular bed, while above heart level a low-pressure vascular bed develops with thinner vascular walls and less powerful constrictors in the resistance vessels than in the high-pressure region.

This important adaptation to terrestrial life is the reason why astronauts develop facial oedema, while leg tissue fluid is partially absorbed into the blood stream. Both positive and negative hydrostatic gradients relative to the heart are wiped out during weightlessness and we then have an equalised arterial blood pressure which will alleviate the high-pressure vascular bed's daily increased pressure during terrestrial life and increase the arterial pressure in the low-pressure vascular bed day and night. Evidently the cardiovascular bed does not adapt easily to this change in hydrostatic pressure because the facial oedema does not go away during even long space flights when other cardiovascular functions have adapted. It is expected that astronauts will also develop facial oedema on the Moon, and the study of capillary pressure in this context is therefore of both scientific and operational interest.

Autonomous nervous system
Important reflexes which serve the cardiovascular regulation and participate in the redistribution of blood originate in the stretch-sensitive areas of the large arteries (aorta and carotid artery) and in the aterior of the heart and lung artery called the high and low baroreceptors, respectively. The adaptations of these reflexes are important study objectives both for spaceflight and on the Moon. Different stimuli might be induced to challenge these reflexes, e.g. by exercise, lower body negative and positive pressure, neck suction, and temperature variation, whereby the regulatory processes in the cardiovascular system itself can be studied.

■ Oxygen metabolism
Both lung function and cardiovascular function change in reduced gravity and so the conditions for oxygen transportation and thereby tissue metabolism might also change. Study of these phenomena during a mission to the Moon will be important to interpret data obtained both in weightlessness and in ground-based investigations. The rate of oxygen uptake in the lungs will be measured during rest and exercise and eventually during variations in oxygen pressure and for different body temperatures.

Data from Apollo missions showed no difference in the maximum oxygen uptake rates of those astronauts who had been exposed to the Moon's gravity and those who did not land on the Moon (−21% and −17% in the maximum oxygen consumption rate, respectively), which can be taken as an indication that the Moon's gravitational field is below threshold for that particular physiological parameter.

Calcium metabolism
During weightlessness a negative calcium balance is produced which cannot be counteracted. Whether this reduction is hormonal in nature or related to the lack of stress on the weight-bearing bones is not yet completely clarified.

It remains a fact that even when rigourous exercise regimens are adhered to, it has not been possible to prevent a constant calcium loss. It is therefore thought that the Moon's gravity is above the threshold for bone decalcification.

- **Blood formation**

The formation of red blood cells is negatively affected during weightlessness, perhaps as a consequence of changes in the bones where the blood-forming tissues are located. It is hypothesised that blood formation might be unaffected or affected only to a minor extent by the Moon's gravity. Study of this field and of the field of bone formation is of course of extreme importance for long-duration missions, and for the understanding of the astronaut's recovery physiology.

- **Pharmacokinetics**

Results from weightlessness show increased blood concentrations of some pharmaceuticals, but there is no definitive information available concerning changes in pharmacokinetics. This is an important medical problem when treating a sick astronaut and for certain preventive procedures, and must be investigated for the Moon also.

- **Radiation**

There are currently insufficient guidelines for the radiation protection of man on the Moon. It is thus of paramount importance that better knowledge be achieved by long-duration exposure of animals to cosmic radiation with various shielding factors and protective techniques in order to prepare for long-duration interplanetary flights.

- **Countermeasures**

The problem of countermeasures must be studied in the light of ethical considerations. The question of for how long it is acceptable to study the spontaneous adaptation processes before countermeasures must be started cannot be answered today. A carefully planned strategy needs to be adopted taking into account both results from spaceflights and Earth-based studies during bed rest.

Even the results from year-long Russian spaceflights are very sketchy regarding the kind of countermeasures to be applied. The Moon might provide an interesting environment for addressing the problem of prescribing such countermeasures, because the risk to man, and thus the criticality of the planning, might be less pronounced.

■ **Medical care**

The extent to which medical care can be provided depends upon, and will increase with, the size of the population of the Moon base. The medical-care facility could thus range from a medical kit to a section of the lunar base equipped as a small emergency room, with at least one or perhaps two medical doctors available.

■ **Psycho-physiological support**

The psychological problems likely to affect the crew of a lunar base can be addressed by careful study of the results from spaceflights in the coming years.

5.6.3 Technical requirements

The three primary goals for human physiology and medicine are: (i) basic physiological research on man and animals; (ii) medical care; and (iii) countermeasures. To address all three problem areas, a rather complicated set of equipment has to be installed at the lunar base.

■ **Basic physiological research**

This will be performed as a further development of Anthrolab, which contains analysers for respiratory gas, an electrocardiograph, a pulse oximeter, plethysmographs, a radioactivity measuring device, turn-table, an echocardiograph, etc. For on-board blood and urine samples, equipment used by other scientists working in the biology field might be shared.

For animal experimentation, a work station must be available to be shared with other life-sciences disciplines. This also holds true for other measuring devices.

■ **Medical care**

This part of the base will contain a kind of doctor's consultation and investigation room combined. There will be conventional laboratory equipment for measuring blood and urine samples, there will be kits for heart-lung investigations, blood pressure measuring device, ECG equipment, echocardiograph, lung-function measuring device, equipment for eye and ear, nose and throat investigations and operations, a dentistry kit, X-ray system, and possibly a magnetic-resonance device. There will be an operating table, with the necessary kit for smaller operations such as an apendectomy or minor urinary or gastric/oesophageal interventions. Generally speaking, the same medical equipment cannot be used for physiological or countermeasure studies.

■ **Countermeasures**

Muscular exercise: The kind of muscular exercise to be undertaken is yet to be decided upon, but recent studies seem to indicate that a kind of rowing machine can be constructed which can be used both for cardiopulmonary function training (long-duration exercise at a low level of muscle tension) and for muscle function training per se (short-duration contractions at a high level of muscle tension).

Lower-body negative pressure: This countermeasure was used during Skylab missions, but might not be meaningful on the Moon.

Other countermeasures may become apparent based on the experiences of the first Moon missions.

5.7 Usefulness of a lunar base

The exobiological experiments will be improved by direct human involvement, e.g. collection of appropriate lunar samples, analysis, refurbishment and repair of the equipment. However, planetary protection is an important issue and the COSPAR Recommendations need to be applied. For organic/chemical analyses of lunar samples in particular, precautions have to be taken to minimise contamination with material of terrestrial origin.

A lunar base provides observatory and laboratory sites with an extremely tenuous atmosphere (no absorption, no turbulence, all wavelengths) low radiation pollution (S/N-enhanced), ultra-high stability (to 1 nm), cool environments at the poles (cryogenics, infrared instruments), high vacuum, no magnetic fields, no ionosphere. Man-made radio interference is the worst foe of the search for artificial extraterrestrial signals, and it has been steadily increasing over the years. Thus SETI is very interested in locating its radio telescopes on the far side of the Moon, which is still completely free of such interferences, provided that side is the subject of future protective policy.

Advanced studies on artificial ecosystems are not possible without establishing a lunar base. The most important general component is a sufficient energy supply, preferably from solar cells, which should not pose any serious technical problems with the present state of technology. Supply of nutrients, and above all water, is obviously more difficult and recycling plants should be designed. For most purposes, the permanent presence of humans is not necessary; a man-tended station would suffice at least for purely physical and 'automated' biological experiments.

BLSS development is primarily coupled with long-term manned spaceflight. There is therefore a predictable interaction between the construction of a lunar base and the evolution of BLSS from basic research through test phases until final implementation. The lunar base is – besides a well-equipped Space Station – an essential prerequisite for BLSS research and long-term testing in space. All three evolutionary stages of BLSS mentioned above are not possible without manned laboratories/test facilities. In a final and highly sophisticated version, a BLSS may operate in an automated fashion, but it then falls outside the pure research domain and belongs to the life-support equipment. Because BLSS research is cross-linked with various other biological disciplines, such as radiation biology, genetics, plant physiology, animal physiology, human physiology, etc., it can contribute in a broad sense to the scientific utilisation of a lunar station. On the other hand, however, the lunar base will represent an essential facility in the development of any highly reliable and economic BLSS.

5.8 Role of Europe

In the following fields of research, European scientists have special experience or are even leading authorities:

5.8.1 Exobiology

In exobiology, studies on the impact of extreme environments (free space, extraterrestrial solar radiation) on biological integrity have already been performed by European scientists on the Apollo-16 and Spacelab-1 missions, and others are planned for future missions, including Spacelab-D2 and Eureca. Eureca's Exobiology Radiation Assembly (ERA) core facility has been specifically developed for in-situ exposure experiments. European scientists have contributed to the future planning studies of both NASA, e.g. Exobiology in Earth Orbit (DeFrees et al. 1989) and ESA, e.g. ERA Follow-On Scientific Study (Horneck et al. 1989).

European bio-astronomers are in a strong position, reflecting the general astronomical situation: with the four main partners Germany, France, the United Kingdom and Italy, Europe is a close second to the USA. The European Southern Observatory (ESO) is one of the leading astronomical institutes. In the SETI domain, France is a European leader with its large Nançay radiotelescope 200 km south of Paris, the second largest in the world. Efforts are being made to organise a SETI global network.

5.8.2 Radiation biology and protection

Since the Apollo missions, European radio-biologists have gathered comprehensive data on the spectral composition of heavy ions of cosmic radiation, inside and outside the geomagnetic field, and on the biological importance of this particulate radiation in space, e.g. Biostack, Biobloc, dosimetric mapping, CIRCE. The Biostack programme, directed and coordinated by scientists at DLR in Germany, unites twelve different institutions from Europe (Paris, Toulouse Manchester, Brno, Frankfurt, Gießen, Homburg/Saar, Kiel, Marburg, Siegen), the USA (San Francisco, Berkeley), and Russia (Moscow). The focus of this programme is to understand the biological effectiveness of the cosmic radiation, especially the HZE particles. For that purpose, the Biostack concept has been developed to reveal the effects of single HZE particles on individual test organisms, a method that is now being widely used by various groups interested in heavy-ion radio-biology. This research work, which includes experiments in space as well as at accelerators, also deals with in-flight dosimetry and the impact of the microgravity environment on radiation response, as well as radiation-protection considerations for human beings undertaking space activities. Within the ESA study PARIS (Protection Against Radiations In Space 1987), European radio-biologists have already provided estimates of space-radiation hazards and their implications for manned space missions.

5.8.3 Artificial ecosystems

Compared with the comprehensive studies on artificial biosphere concepts in USA, USSR and Japan, activities in Western Europe are mainly restricted to basic artificial ecosystem research, which is as valuable for the solution of terrestrial ecological questions as for the establishment of BLSSs in space or on other celestial bodies, e.g. the Moon. Such research includes biomass production and utilisation, biological waste management and recycling, carbon-dioxide gas exchange in open systems, as well as the development of combined artificial ecosystems, e.g. Melissa and CEBAS – Aquarack. The latter combined aquatic animal-plant artificial ecosystem may serve as the precursor of an aquatic BLSS. In this framework, theoretical and experimental approaches to mixed algal reactors and to artificial self-regulating ecosystems are the focus of interest.

Although the activities in ecosystem research mentioned above are not directly targetted towards the development of a lunar-base BLSS, they may contribute substantially to the solution of several unresolved questions in CELSS or BLSS development. The basic concept of the German CEBAS – Aquarack project is already implemented in the NASA scenario of the lunar base as an aquaculture system dedicated primarily to food production, as a result of cooperative DARA/NASA activities in this project. All European efforts in the field of basic BLSS research, however, are of no value in the context of lunar-base research if there is no information transfer between the performing, coordinating and administrative bodies which – for reasons of scientific quality, efficiency and economy – have to elaborate a common catalogue of requirements and problems to be resolved. In the framework of an international artificial ecosystem research programme as proposed by Gitelson (1991), European scientists would be able to provide valuable contributions to lunar-base BLSS research.

A survey of European activities in artificial ecosystem research can be found in the Proceedings of the DARA/CNES Workshop on Artificial Ecosystems (André 1991).

5.8.4 Human physiology and medicine

There is a long European tradition of human-physiology research stretching back more than 100 years. In particular, there has been a long-standing interest in man's adaptation to different environments such as high altitudes, deep diving, cold and hot climates, exercising under different conditions, being trained or untrained, man or woman. It was therefore not surprising that, when the possibility of investigating man's adaptation to weightlessness arose, experienced and high-level researchers in various laboratories around Europe took up this new challenge. Furthermore, the models used to simulate weightlessness and to vary the central venous pressure have resulted in renewed interest in the regulation of circulation and fluid volume, which has led to a succession of publications over the last decade or two. This interest should also be seen against the background of the new and more exact methods of analysing hormones that have appeared. The European space physiology and medicine community is anchored in the scientific tradition of the universities and other well-established and recognised laboratories and institutes, and is well prepared for the challenges of manned spaceflight.

5.9 Conclusions and recommendations

Within the overall scenario of lunar exploration, the interest on the part of the life scientists is twofold: on the one hand, the Moon provides an ideal platform for conducting a variety of new investigations, some of which cannot be carried out elsewhere (e.g. the establishment of an artificial ecosystem on the Moon will be a challenging project per se). On the other hand, the boundary conditions for human safety, health, well-being and working efficiency at such a lunar base have to be established. The scientific as well as the operational aspects need to be examined. It is therefore recommended that ESA should set up a life-sciences study group to investigate the various opportunities and goals for life-sciences research on the Moon, addressing both the scientific requirements and the feasibility of the hardware concepts under consideration, e.g. ERA, Botany Facility or Aquarack.

6. Exploitation of the Moon


USA

Chapter 6. Exploitation of the Moon

6.1 Power generation by thermonuclear fusion

6.1.1 Introduction

Thermonuclear fusion is one of the very few options (fusion, fission and solar) for providing useful and plentiful energy to mankind. It exploits the fact of the very high binding energy per nucleon at the low end of the atomic mass scale, thus fusing together light nuclei such as protons, deuterium, tritium or helium-3 to produce helium-4 provides much more energy per elementary process than the fission of uranium. Whereas fission typically yields 2×10^7 kW/kg fissile material, typical fusion values are 2×10^8 kW/kg fusion cell. However, because the particles to be fused carry electric charges of equal sign, and thus repel each other, the effective fusion cross-section is small, and the relative kinetic energy of the particle required to achieve their fusion is correspondingly high. The only reasonable way to produce a net energy output is to heat a plasma consisting of the fuel constituents to temperatures above 10 keV and to confine it long enough for the fusion processes to occur at a sufficient rate.

This is a difficult task, and all the present effort is concentrated upon using the fuel offering the largest fusion cross-section together with a high reaction energy, which is a DT mixture as described by the following process:

$$D + T \rightarrow {}^4\text{He} + \text{n} + 17.6 \text{ MeV}$$

Of these two fuel constituents, D and T, only deuterium is abundantly available on Earth, whereas tritium has to be bred. This is done by exploiting the neutron of the above reaction which, according to the following reactions, breeds tritium in a lithiumin-containing blanket surrounding the reaction chamber:

$$\begin{aligned} {}^7\text{Li} + \text{n} &\rightarrow {}^4\text{He} + \text{T} + \text{n}' - 2.5 \text{ MeV} \\ {}^6\text{Li} + \text{n} &\rightarrow {}^4\text{He} + \text{T} \qquad + 4.8 \text{ MeV} \end{aligned}$$

If desired, isotope enrichment of lithium is possible at rather low cost.

6.1.2 State-of-the-art fusion development

The state of the art of fusion development can best be assessed by noting that, at present, the four World fusion programmes (Euratom, Japan, USA and USSR) are jointly designing the ITER facility, an engineering fusion test reactor for confining a burning DT plasma, and for using it for the development of fusion technology. This demonstrates the confidence existing in fusion circles that the physics database needed for the creation of a burning plasma is at hand. Remaining questions are of a quantitative rather than a qualitative nature. Some further supporting experiments are under way which are expected to clarify some specific ITER-related questions before starting the construction of the faclility. If everything runs as expected, ITER could begin producing results in the next decade. A demonstration fusion reactor, the natural step to follow ITER, and designed to demonstrate that fusion power can be produced on a competitive basis, would then be expected to begin operating around 2025. This indicates not only that fusion power is well on the way to becoming a reality, but also that the inherent time scale for its development, being determined by a sequence of construction and operation of big devices, is rather long.

As previously stated, all of the above activities are based on DT fuel. Unfortunately, nature has combined the highest fusion cross-section of DT with a number of properties of this fuel which are unwanted and not all inherent in the fusion process in general. These are:

- One of the reaction products is a neutron which carries 80% of the reaction energy, leading to the need for arranging a blanket around the plasma for energy conversion and for a shield, and causing activation of the structure surrounding the plasma.
- Tritium, one of the fuel constituents, has to be bred in a lithium-containing mantle situated around the plasma, but inside the magnetic-field coils, thereby using expensive space filled with magnetic field.
- Tritium is an unstable isotope which requires specialised measures for its safe containment.

6.1.3 Alternative fuels

Although these points are considered to be manageable without excessive effort, and although DT-fusion is already considered to be very desirable for its environmental properties, it is nevertheless appropriate to check whether there are other fuel combinations available that are more desirable from an environmental point of view. In fact, there are two other fuels worth considering, DD and D^3He. No other fuels offer sufficient yield.

DD works according to the following reactions:

$$\begin{aligned} D + D &\rightarrow {}^3He + n + 3.3\ \text{MeV} \\ &\rightarrow T + p + 4.0\ \text{MeV} \end{aligned}$$

with the two branches occurring at about equal rates. Two of the reaction products, T and ^3He, will further react with D, the initial fuel, according to the reactions above and below, to yield in total:

$$3D \rightarrow {}^4He + p + n + 21.6\ \text{MeV}$$

The improvement provided by using this reaction is only partial. The fuel side is positive, because there is only one constituent which is stable, abundantly available, and can be transported to the reactor site without any radioactivity precautions. Breeding of fuel is not required. However, on the side of the reaction products there are still neutrons and tritons (the latter only transiently). Considering the disadvantage of the much lower cross-section of DD in comparison with DT (and thus the much greater difficulties in confining a plasma of this type), there is no incentive to address DD as a potential fuel at this time.

6.1.4 The case of D^3He:

The case may be different for D^3He:

$$D + {}^3He \rightarrow {}^4He + p + 18.4\ \text{MeV}$$

This process on both the fuel side and the side of the reaction products involves only stable elements. The reaction energy is as high as for the other fuels considered here.

This is altogether very positive. However, the cross-section of the reaction is much smaller than for DT, requiring, for the same fusion power density, an approximately five-fold increase in burn temperature, and the confinement of a plasma pressure about one order of magnitude higher than for DT, which is close to the limit of what is possible based on the laws of physics. This is alleviated somewhat by the fact that all the reaction products are charged particles and thus available for maintaining the reaction temperature of the plasma (for DT this is only one-fifth), and that a shield of reduced thickness reduces the volume filled with magnetic field. The possibility of excluding radioactively unstable elements altogether would certainly provide an enormous challenge in trying to solve the additional physics and technology problems connected with the use of this process.

Unfortunately, there are two further obstacles connected with the use of D^3He. There are also possible DD reactions among the particles of one of the fuel constituents, and there is practically no ^3He available on Earth. The first point would re-introduce all the radioactivity discussed before, though on a reduced level. This level could be further reduced by changing the fuel mixture from 1:1 to, for example, 1:9, which would reduce the DD reaction rate by a factor of 25, compared with the D^3He reaction rate being reduced by less than a factor of 3. A further possibility would be offered by proper spin polarisation of the D nuclei so as to reduce the DD reaction rate even further. However, this process has not yet been demonstrated to be viable, and also the reduction factor is still in dispute.

Under these circumstances, the conclusion is that at least the first generation of fusion reactors will have to work with DT. DT fusion has to be demonstrated before the additional physics problems connected with advanced fuels can be addressed meaningfully, because this work has to build on the experience gathered by using DT. D^3He remains a challenging alternative though, because of the promised large reduction in radioactivity. As far as the availability of fuel is concerned, however, there is no immediate need for large quantities of ^3He. The existing and available resources of ^3He on Earth seem to be of the order of several 100 kg, originating mainly from the decay of the tritium of atomic weapons. This amount is considered sufficient for any D^3He development programme, and even for the first fuelling of a reactor. Further supplies would only be needed if and when the first reactor based on D^3He starts operating. This will certainly not be earlier than several decades from now.

There are two conceivable sources for large amounts of ^3He. The first one was considered many years ago and consists of fusion reactors designed to produce excess amounts of tritium to be stored and allowed to decay to ^3He. The idea was to site tritium-producing reactors at remote locations, whereas the ^3He produced by the decaying tritium could be used in reactors located close to regions of high population. Nowadays, this is no longer considered a practical option.

6.1.5 Lunar helium-3

The second possibility is to exploit the large ^3He resources existing on the surface of the Moon, which were generated by continuous bombardment of the lunar surface by the solar wind. Some important figures on these resources are collected in Table 6.1.

Table 6.1. Useful data on lunar Helium-3

Parameter	Value
Concentration of $^3He/^4He$	480 ppm
Concentration of helium in lunar soil	36 g/ton (with some spread)
Concentration of ^3He in lunar soil	13 mg/ton (with some spread)
Depth down to which ^3He is distributed	5 m
D^3He energy equivalent of these resources	20×10^6 GWa

It is also worth noting that 30 tons of ^3He is approximately equivalent to the annual electricity consumption of the United States. It could also be the payload of a Space Transportation System similar to the Space Shuttle, so that one such load per year would be enough to provide a power output equivalent to the total US electricity consumption.

Table 6.1 clearly demonstrates that the ^3He resources available from the Moon's surface are large enough to produce fusion energy on Earth for several centuries. The tasks to be addressed for their exploitation are:
- mining of lunar soil
- degassing
- mass separation of the gas
- transportation to Earth.

First estimates indicate that about 1% of the latent energy of D^3He would be sufficient to carry out all of these processes, from mining to transport to Earth. Energy-wise, this amount would be fully acceptable. Also as far as the economics are concerned, there are indications that exploitation of the lunar resources of ^3He would not significantly increase the cost of fusion power, provided that the helium concentrations available for exploitation are at least as high as quoted above. More exhaustive information can be found in the References.

6.1.6 Conclusions and recommendations

Thermonuclear fusion is one of the few options (fusion, fission or solar) for providing a plentiful supply of energy to mankind in an environmentally acceptable manner.

The physics and technology development effort needed for the exploitation of fusion energy is difficult and requires a step-by-step approach, which inherently leads to long development time scales.

The first fusion reactors will operate with DT fuel, because it offers the highest power output for the given parameters.

Compared with DT, use of D^3He would offer a large and significant reduction in the radioactivity associated with the fusion process and would therefore be very desirable. However, the physics and technology involved are much more demanding than for DT.

With several hundred kilograms, there is enough ^3He on Earth for any D^3He development programme, up to the first fuelling of a reactor.

Lunar ^3He could be the source for this kind of fuel if and when D^3He fusion reactors start power production. This would not happen for a number of decades. The lunar resources of ^3He would then be sufficient to cover the needs for several centuries. Exploitation of the Moon's ^3He resources would be a long-term project. An assessment of the need for a quantitative verification of the lunar resources of ^3He and their spatial distribution should be made some time in the future.

In addition, the need for conceptual studies of the mining of the lunar soil, its degassing, extraction of ^3He and its transportation to Earth, should be addressed. Some of the answers could possibly be deduced from other lunar exploitation studies.

6.2 Exospheric contamination due to human activities on the Moon

6.2.1 Introduction

It had already been deduced from optical and radio observations before the Apollo era that the Moon could not possess an atmosphere with a density exceeding 10^{-12}g/cm^3 at its surface (corresponding to an altitude $\sim$250 km above the Earth). From Apollo-mission measurements, an even lower atmospheric density limit was inferred. Morgan & Shemansky (1991) report more recent ground-based spectroscopic observations made in 1988 just above the sunlit limb of the Moon which indicate that the sodium density ranges between 40 and 80 atom/cm^3 at the subsolar base of the lunar atmosphere. They also report that these most recent observations show that the density scale height of the sodium atoms $H_{Na} = 75 \pm 27$ km at the same location. Similar results are reported for the potassium atoms: $[K] = 15 \pm 3$ atom/cm^3 and $H_K = 90 \pm 20$ km.

For such small densities, the mean free path of particles is larger than the atmospheric density scale height. Consequently, the surface of the Moon coincides with the exobase surface (for the Earth, it lies at an altitude of 500 km); the lunar atmosphere must therefore be considered as a planetary exosphere: all neutral atoms move along ballistic or escaping trajectories in the gravitational field with no collisions between them. This implies that any localised injection of new particles from a landing spacecraft or from a man-tended lunar station will quickly spread all around the Moon, in a time of the order of the ballistic free flight time of the molecules; i.e. in less than 3 h for oxygen atoms which have a mean thermal speed of 0.6 km/s.

6.2.2 Gravitational escape

All lunar light gases escaped completely from the Moon's gravitational field: for example, atomic hydrogen disappears in less than 120 min from the sunlit lunar hemisphere where the temperature is 400 K and where the thermal speed of hydrogen atoms is 2.5 km/s, which is comparable to the escape velocity at the surface of the Moon, namely 2.38 km/s. A helium atmosphere would similarly dissipate in some 3.6 h of daylight, or 1.4 yr at night, when the regolith temperature is only 100 K. Atomic oxygen and water vapour are expected to take years to escape by this gravitational mechanism in daytime, and 10^6 yr at night (Kopal 1974).

The heavier atoms and molecules could also react with surface rocks to form solid components. This does not, however, apply to the 'inert' gases like argon, krypton, or xenon, which do not form compounds and which are sufficiently heavy for their escape rate to be moderately low: the mean thermal velocity of argon at the surface temperature of 400 K is 0.4 km/s, i.e. six times smaller than the escape velocity (Kopal 1974).

6.2.3 Additional removal mechanisms

In the case of inert gases, additional (and faster) removal mechanisms have been identified:

(i) collisions of neutral atoms with the solar energetic particles (Herring & Licht 1959, 1960), and more importantly

(ii) photo-ionisation of the neutral atoms combined with electric-field acceleration of newly formed ions.

Indeed, electric fields are induced around the Moon: (i) by photo-electric emission from the sunlit lunar surface (Hinton & Taeusch 1964); and (ii) by the dynamo action of solar-wind convection through the lunar exosphere (Manka & Michel 1970).

The latter dynamo action due to the streaming solar wind produces an electric field equal to $E = -V \times B$ in a frame of reference fixed with respect to the Moon (V and B are the solar-wind bulk velocity and magnetic-field intensity, respectively, in this frame of reference) (Manka & Michel 1970). Therefore the average lifetime against loss by electric field acceleration depends mainly on the time required for solar radiation to ionise such a heavy neutral particle orbiting in the lunar exosphere.

The time required for solar radiation to ionise a particle is about 10^7 s (0.3 yr) for most gases, but approximately 10^6 s for argon. This is much shorter than the lifetime against gravitational escape (190 yr for Ne, 1.3×10^{10} yr for Ar). Whether an accelerated exospheric argon ion escapes or is accreted in the lunar soil depends primarily on the direction of the electric field E at the lunar surface (i.e. in the direction of V and B in the interplanetary medium).

The electrostatic force exceeds the gravitational force by a factor $eE/mg = 5000$ for $^{40}Ar^+$ (Manka & Michel 1970). This indicates that the solar-wind electric and magnetic fields control the time of residence of any photo-ionised atom in the lunar atmosphere-ionosphere. This implies also that the time of residence of heavy atoms and molecules released in the lunar atmosphere from a man-made station or spacecraft will be determined by their characteristic ionisation time scale (0.3−0.03 yr), more than by their rate of gravitational escape as neutral particles.

6.2.4 Non-Maxwellian velocity distribution

The gravitational escape flux of neutral particles (also known as the 'Jeans escape flux'), is predominant for the lightest atoms only (H, H_2, He); the value of this dispersion rate is calculated with the assumption that the velocity distribution of particles at the exobase is Maxwellian; it was also generally assumed that the rough nature of the lunar regolith exobase should produce directional uniformity of the velocity distribution near the surface of the Moon (Hodges 1973).

These concepts were questioned in 1977 by Shemansky et al. (1977). They noted that the absence of a high energy tail in the distribution of down-going exospheric atoms should be reflected in the up-going distribution as a deficit of escaping atoms. Furthermore, they emphasised that the nature of the interaction of the atmospheric particles with the regolith surface is of critical importance for the determination of: (i) the dominant loss mechanisms; (ii) the global distribution of the gas in the lunar exosphere; and (iii) the influence of the source geometry and energy distributions.

The early lunar exosphere models have been generated on the basis of a complete thermal accommodation of all particles to the surface of impact. The introduction of more realistic thermal-accommodation coefficients introduces varying degrees of complexity in the modelling process, depending on the magnitude of these coefficients. As a consequence, the surface temperature tends to lose its control over the local atmospheric energy distribution for a given exospheric particle species, and the exobase temperature tends to be more uniform than in the case of the earlier thermally accommodated models. Moreover, energy accommodation at the gas/surface interface also alters the details of the exospheric altitude distribution. The exobase temperatures of 'old' (orbiting exospheric) and 'new' (released from the regolith) atoms are not necessarily equal. These two populations of the same particle species distribute in the lunar exosphere with two different scale heights H_1 and H_2.

6.2.5 Monte-Carlo model calculations

Atmospheric modelling becomes more and more complicated when such details of surface interaction are taken into account. The number of adjustable parameters such as: (i) variability in the heats of absorption; (ii) nature of the surface roughness; (iii) magnitude of the accommodation probability; and (iv) day/night temperature asymmetry, make a unique and analytical solution difficult to obtain. Therefore, Hodges (1973), Smith & Shemansky (1978) and Hodges (1980) suggested that the only viable numerical method for solving this complicated problem must be based on Monte-Carlo (MC) simulations. Hodges (1980) presented original exospheric-model calculations based on the MC method, and checked its validity in a series of cases when analytical solutions are known (e.g. Hodges et al. 1972). This new numerical tool (MC method) also permits arbitrary specification of time-dependent physical processes affecting atom creation and annihilation, atom-regolith collisions, adsorption and desorption, and non-planetocentric acceleration.

Figure 6.1 illustrates the results of one of the MC simulations published by Hodges (1980). It shows the concentration versus lateral distance λ from the point source where new atoms fuse omni-directionally with a Maxwell-Boltzmann velocity distribution. The concentrations are scaled by $10^{-z/H}$ to separate adjacent altitude plots, where z is the altitude and H the atmospheric density scale height of the neutral atoms considered. The dots which show the results of the MC simulations fit rather well with the solid line corresponding to the analytical solution determined for this simple case by Hodges et al. (1972).

Figure 6.2 shows the vertical flux of atoms as a function of lateral distance from the temperature discontinuity at $x > 0$, the location of the lunar terminator. The x-axis is calibrated in units of local scale height: H_1 for $x < 0$ and H_2 for $x = 0$. The solid line is an analytical solution for this 'simple case', while the dots correspond to the results obtained with the MC method. This good correspondence between the analytical and MC methods lends confidence to the MC simulations in the case of

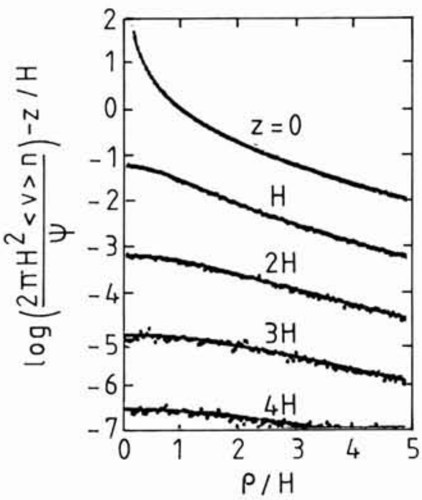


Figure 6.1 Normalised concentration versus lateral distance from a Maxwell-Boltzmann point source

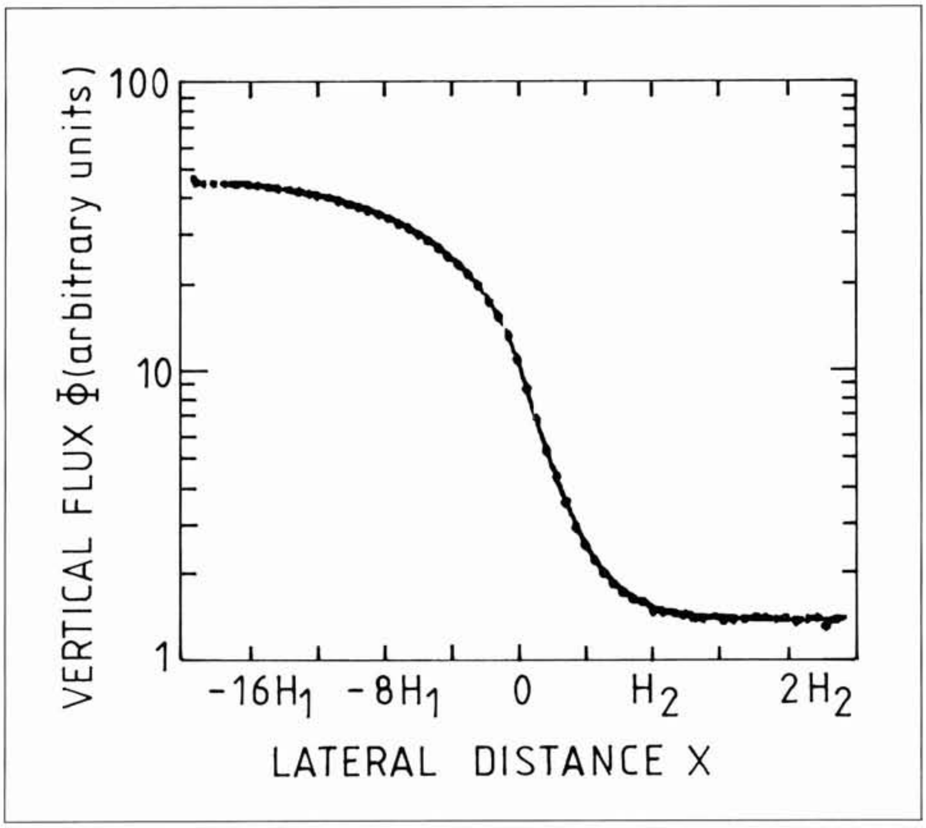


Figure 6.2 Vertical flux as a function of lateral distance from temperature discontinuity at x=0

more complex velocity distributions and non-symmetrical boundary conditions, when analytical solutions are not obtainable.

6.2.6 Complexity of the lunar exosphere models

For each atomic species or molecule, the interaction with the regolith material is different; it has to be determined by appropriate laboratory experiments with lunar-like regolith. Furthermore, the photo-ionisation rate is different for each type of atmospheric particle, this also makes the modelling exercise more complicated and specific to each particle species to be imported by future human activity on the surface of the Moon. This discussion shows that the road to more comprehensive and realistic lunar exospheric models not only depends on the availability of more sophisticated numerical simulation tools (such as the MC method), but will also require a lot of laboratory work by solid-surface geophysicists, geochemists, and physico-chemists to analyse the interaction of different types of lunar soil materials with all relevant atoms present in the Moon's exosphere, or molecules that man will import there.

From the preceding resume and a recent review entitled the 'Limits to the Lunar Atmosphere' by Morgan & Shemansky (1991), it can be concluded that a full understanding of all the physical processes of importance in determining the distribution of different atoms and molecules in the lunar exosphere is certainly not as easy as was thought twenty or thirty years ago. A comprehensive study and modelling of all of these processes is not yet available. For instance, from the recent review by Morgan & Shemansky (1991), it can be concluded that the production of sodium and potassium observed spectroscopically in the lunar exosphere is controlled by continuous impact vaporisation of regolith by meteorites falling on the Moon, which is an additional process not taken into account in most previous models of the lunar exosphere.

Modelling the impact of a permanent or semi-permanent lunar station (man-tended or not) is even more complicated because the source of injected atoms and molecules will change with time, contrary to the stationary model simulations of Hodges (1980). Moreover, future models will have to take into account the fact that the complex molecules that will be injected into the lunar exosphere will undergo photo-dissociation and adsorption in different hemispheres at the lunar surface. Consequently, a comprehensive and realistic assessment of the actual atmospheric distribution and the long-term effects of lunar stations on the evolution of the Moon's exosphere is currently beyond reach and would require a more detailed study by a working group composed of physicists, geochemists and modellers.

6.2.7 'Back of an envelope' calculations

It is not necessary to await the conclusions of a long and detailed study to infer that new spacecraft landings and the presence of manned stations will have a significant effect on the current atmosphere of the Moon as well as on its surface, at least over time scales of the order of the dissipation rate of the exogenic molecules that will be injected (and permanently in the case of permanent settlement).

Indeed, the total mass of particles orbiting in the lunar exosphere is estimated to be less than 1200 t, assuming a scale height of 50 km (for argon) and a surface density smaller than 10^{-15} g/cm^{-3}. This latter density can be inferred from the cathode pressure gauges mounted on the lunar surface during the Apollo-12 and 14 missions, indicating that the maximum atmospheric density appeared to correspond to some 2×10^5 particles per cm^3 at night, and about 100 times more during daytime (Johnson 1971). The derived daytime densities were $5-10 \times 10^7$ atoms/cm^3 (Johnson et al. 1972), but what fraction of this was native to the sunlit atmosphere of the Moon is unknown, as is the degree to which the initial large flux of polluting species and internal outgassing of the spacecraft may have compromised the subsequent performance of the Cold Cathode Gauge Experiment of Apollo's ALSEP package. Kopal (1974) estimates that the total amount of atmospheric hydrogen does not exceed 1 kg of free mass, and that of xenon about 1 t. From the recent spectroscopic measurements by Potter & Morgan (1988), it can be inferred that the total amount of sodium in the lunar exosphere is less than 2.6 kg, less than 1.5 kg according to Tyler et al. (1988). Similar estimates indicate a total potassium mass of less than 0.8 kg, and a comparable amount of oxygen atoms.

Such amounts are small in comparison with gas already released into the lunar environment by human activities. Between 1966 and 1967, the soft lunar landers and orbiters imported enough gas to have already altered the composition of the lunar exosphere. Furthermore, the quantities of gases released around the Moon by the successive Apollo missions could have profoundly altered its atmospheric composition. Kopal (1974) estimated that each of the seven Apollo missions left about 10 t of exhaust gases on the Moon. Including the early American and Soviet hard and soft landers, the total contribution may be little short of 100 t. If this amount were spread uniformly around the Moon, it would produce a gaseous envelope with a density close to 10^{-10} g/cm^{-3} 100 m above the lunar surface. The molecular weight of the constituent particles (some 36% H_2O, 32% N_2, 13% H_2, 9.6% CO, 3.7% CO_2, 1.9% H and 1.6% OH, according to Aronovitz et al. 1968) is sufficiently high to make the rate of gravitational dispersal of most of these molecules very slow compared with the time interval between successive Apollo missions. Fortunately, we have seen above that the photo-ionisation time of most molecules is less than 0.3 yr, and that due to the solar-wind electric field acceleration the time of residence of most heavy molecules in the lunar exosphere should only be of the order of 0.3 yr.

It is also of interest to compare the amount of gas released by human activities with that from natural sources. The solar wind is continuously transporting to the Moon gases originating from the solar atmosphere. The rate of gas acquisition by the Moon due to the solar wind amounts to 1200 t/yr of hydrogen ions (3×10^8 protons cm^{-2} s^{-1}), 250 t/yr of He, 0.95 t/yr of nitrogen, 6.3 t/yr of oxygen, 1.5 t/yr silicon, etc. It can be calculated that comparable influxes of silicon, carbon, and other heavy elements result from meteoroid impacts on the surface of the Moon. Any permanent man-tended lunar station (including the fuel needed for landing and lift-off) which would release oxygen, silicon, iron, aluminium and hydrogen into the exosphere at yearly rates higher than these figures would therefore surpass natural contributions to the lunar environment.

6.3 A Moon-based archive of the terrestrial biosphere

Records of the evolution of life on Earth and of the development of human existence and culture have generally been compiled in a very haphazard manner. Many records have been destroyed forever due to unfavourable environmental circumstances or deliberate human actions and wars. Humanity should be able to improve considerably on this situation, and leave a record of the evolution of our biosphere, with its rich scientific and cultural contents and variety, for our descendants on Earth or for possible visitors to the Solar System. The lifetime of these records needs to be substantial, bearing in mind that the Earth–Moon system will survive in its present state for a period comparable with its present age (approximately 4 billion years). The existence of a lunar base provides an opportunity to create such a long-duration archive.

The Moon offers a number of advantages for such an archive. In addition to the absence of the deteriorating effects associated with atmospheric conditions and oceans, biological and chemical destruction can be better controlled and hopefully eliminated. Arbitrary destruction by those bent on modifying cultural and historical records becomes much more difficult, and a systematic biasing of the record can be minimised if not eliminated, in the face of the international cooperation that will be needed anyway if we are to exploit the Moon. By locating the archive underground and near the lunar poles, cyclic thermal variations would be minimised. The remaining negative environmental effects would be from cosmic rays and meteorite impacts.

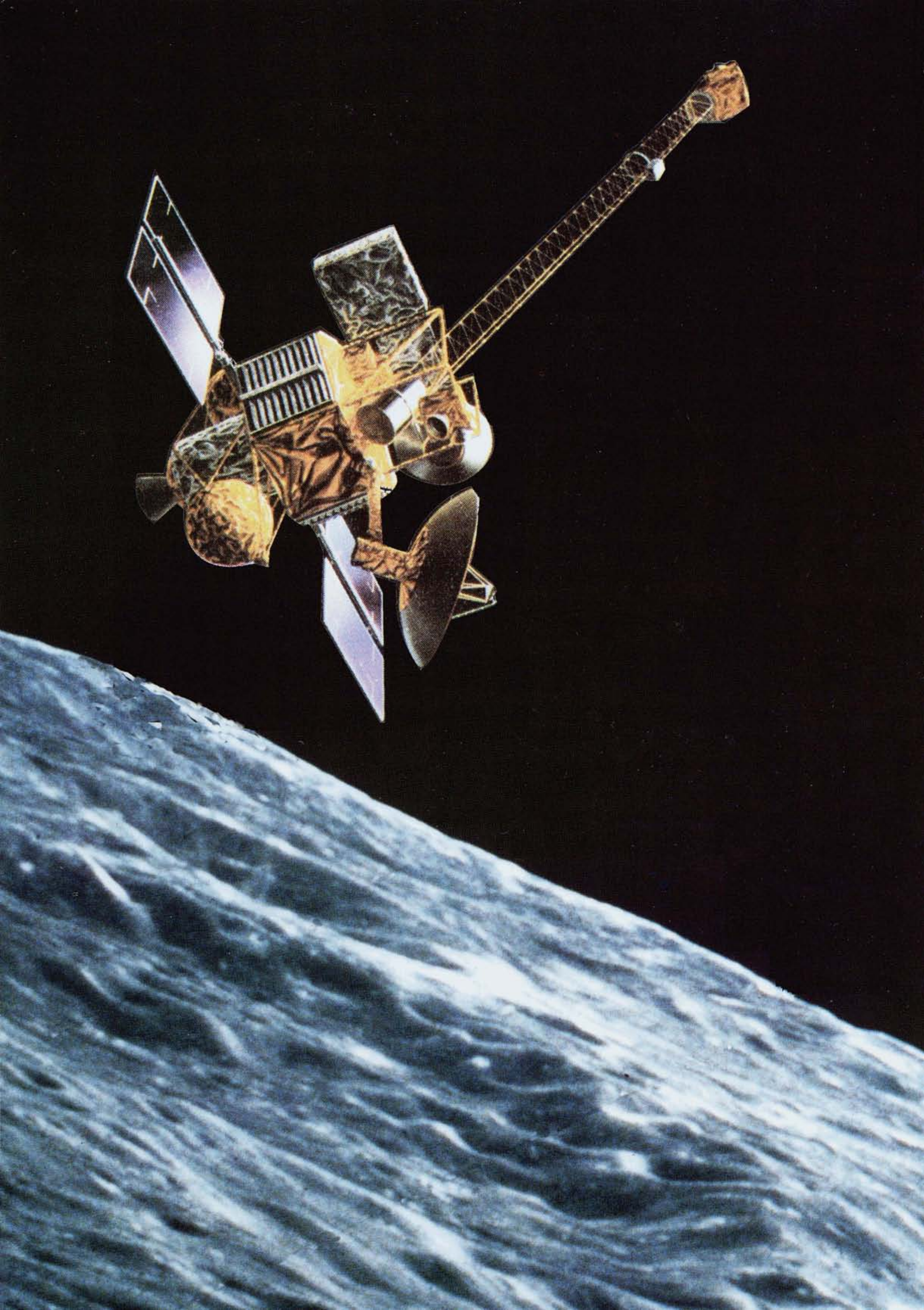
The contents of the archive could include biological, cultural, linguistic, religious, and scientific records. Biological records should, for example, include actual specimens recording the wide range of life forms on Earth and the genetic coding of living beings. The cultural record should represent the enormous variety of human expression and relationships. Linguistic records should be aimed not only at preserving the many forms of the spoken and written word, but also at providing the key to the interpretation of the other records in the archive (e.g. Rosetta Stones). Religious records should show us to be spiritual beings with a frequent need for the supernatural. The scientific records should record the development of scientific thought and knowledge in the many natural and social sciences.

The above is just a first tentative list of possible contents for a Lunar Archive. Developing this content is in itself a major research effort with many intellectual rewards. The creation of such a Lunar Archive also raises a number of practical issues: How to limit its size? How to select its contents? How to make it durable? How to make it 'readable'? How to protect it? How to make it robust?

Two parallel tracks need to be followed in the archive's development: the first is the technological/scientific track related to creating an archive that can survive for long periods of time and which is accessible; this involves the scientific and engineering community. The second track needs to focus on the contents of such an archive, and all interested communities should be involved. A strategy should be developed for implementing such an archive, initially on a small, exploratory basis.

The fortunate availability of a companion 'planet' with the properties of the Moon provides us with an excellent opportunity to create a permanent record of our biosphere. The process of establishing such a record is in itself a very challenging and intellectually stimulating exercise which will bring together representatives from a variety of fields that hitherto were not part of our space research programmes. Beyond this near-term benefit, the Lunar Archive will record our existence for the benefit of future life forms on Earth, and possibly elsewhere also.

7. Conclusions and Recommendations



Chapter 7. Conclusions and Recommendations

Whilst the Moon, after the Apollo and Luna programmes, is now the best-known planetary body other than the Earth itself, its scientific exploration remains far from complete in that major questions are still unanswered. Its potential as an ideal location for carrying out both astronomical observations and scientific experiments is yet to be exploited. For these reasons, the current study concentrated essentially on scientific issues, and in particular the identification of unique scientific goals to be achieved on the Moon. Three broad avenues requiring further in-depth studies were considered, in which Europe has played, and can continue to play, an important role:
- Lunar Science (*Science of the Moon*)
- Astronomy and Astrophysics, in the broadest sense (*Science from the Moon*)
- Life Sciences (*Science on the Moon*).

7.1 Lunar Science

A renewed exploration programme, with the prospect of the establishment of a lunar base, represents an outstanding opportunity for lunar science. The prominent problems that still remain to be addressed are the following.

In the field of geophysics, much has yet to be learned about the internal structure of the Moon and the question of paleo-magnetism remains to be resolved.

The main goal in geology is to extend our knowledge to the whole lunar surface by means of remote sensing, automatic rovers and field geology in the context of a lunar base.

Similarly, in chemical- and mineralogical-composition terms, a global mapping is required, together with representative sets of in-situ analyses combined with sample returns.

The interaction of the Moon with its environment (meteoroids, solar wind, cosmic rays) has dominated its surface evolution over the last three billion years. On the one hand, the formation of a regolith is representative for many Solar System bodies, but the lunar regolith also represents a unique record of the interplanetary medium's evolution over several billion years. Moreover, the regolith can be considered as a material resource containing significant amounts of all but a few key elements.

Based on such considerations, a balanced lunar-exploration programme should include the following four steps:

- A lunar polar orbiter, providing complete geophysical, geochemical and geological mapping, and full high-resolution coverage of the lunar surface with up-to-date remote-sensing techniques.
- In-situ studies at selected representative sites: surface stations, including seismic observatories, and rovers performing traverses across geological boundaries. These elements would determine the local chemical and mineralogical composition, as well as provide site documentation. Studies conducted by ESA for the Marsnet Phase-A study are likely to yield relevant technical inputs regarding means for lunar in-situ investigation.

- Sample returns, from selected representative sites: the highlands, the far side in general, and the high-latitude regions are not (or only poorly) represented in the previous set of nine sampling sites. Deep-drilling cores extending to the base of the regolith and beyond would also be of high scientific value.
- A lunar base would allow investigations by field geologists, which would be extremely valuable for selecting and documenting new lunar samples. A lunar base would also help in placing and maintaining surface stations and rovers.

Recommendation 1

Regarding lunar exploration (Science of the Moon), major questions remain to be answered, for which a four-step approach is recommended involving:
(i) a lunar polar orbiter; (ii) a network of surface stations and/or rovers; (iii) a sample-return strategy for the most relevant sites on the Moon; (iv) a lunar outpost/base, which would make field investigations possible and also provide support for sampling and surveying activities.

7.2 Astronomy and Astrophysics

In the area of astronomy and astrophysics, the Moon offers some unique advantages over ground- or space-based observations. These range from mechanical and thermal stability, to the possibility of deploying large observational facilities. These advantages call for the development of very sensitive high-resolution capabilities for astrometry and imaging in the range from the ultraviolet to submillimetre wavelengths, for the deployment of long-baseline interferometric facilities.

With baselines of up to 1 km, providing angular resolutions of $50-100\times10^{-6}$ arcsec in the visible at limiting magnitudes of better than $m_v=20$, significant advances would become possible in galactic and extragalactic research as well as in planetary science and solar physics.

High-precision astrometry of single objects would yield significant improvements in extragalactic distance scale, as well as thorough mapping of the dynamics within our Galaxy, well-determined luminosities and masses for all types of objects found within our Galaxy and, to a lesser precision, for extragalactic objects. Imaging at these resolutions from the ultraviolet to the infrared would include cometary nuclei and their activity, the surface and atmosphere of planets and stars (including the Sun), accretion phenomena, interacting binaries, active galactic nuclei and quasars. In the submillimetre range, new observations of Solar-System objects, circumstellar envelopes, interstellar matter and the formation of stars and planetary systems, and of extragalactic objects, would be possible.

Indeed, we have no reason to doubt that the Moon is the ultimate location for interferometry. The logical strategy is to build upon current Earth-based long-baseline-interferometry technology to achieve an intermediate-baseline (<100 m) space-based optical interferometer (which has its own scientific justifications), and ultimately to construct a very-long-baseline (1 km) lunar instrument.

The electromagnetically clean environment of the Moon's far side lends itself to the deployment of a Very-Low-Frequency (VLF) observatory, based on large arrays to achieve high angular resolution. This would pave the way for important new observations.

Of particular interest is the observation of free absorption in H II regions, which can lead to derivation of the galactic emissivity and its variation along the galactic plane.

The highly magnetised planets in the Solar System produce a powerful VLF emission which is in marked contrast to the intrinsic emission of the Sun, enhanced by spectral-signature and polarisation characteristics. This suggests that Jupiter- or Earth-like planets in nearby stellar systems could be detected with a large Moon-based VLF telescope.

In the context of a lunar base, an important function of lunar VLF studies would be monitoring of the evolution of solar-flare events through imaging at low frequencies for prediction and protection purposes. However, the installation of a large VLF facility on the far side of the Moon would require the infrastructure of a well-established lunar base. This would not be available in the initial phases of lunar exploration. An early option could be to deploy simple wire antennas remotely at surface stations, thereby providing early access to VLF observations.

Recommendation 2

In the field of astronomy and astrophysics, imaging by interferometry has been identified as the most promising and unique approach. We recommend deepening of the interferometric studies already initiated, following the step-by-step approach proposed in Section 2.2.6:

(i) Plan for a precursor 100 m-class, space-based, imaging optical interferometer that provides an order-of-magnitude increase in performance over ground-based systems.

(ii) Initiate a conceptual study for very-long-baseline interferometers in readiness for their location on the Moon to cover the ultraviolet/visible/infrared and the submillimetre ranges.

In addition, we recommend carrying out a conceptual study for an early programme in VLF astronomy from the far side of the Moon, using simple antennas remotely deployed at surface stations.

7.3 Life Sciences

Within the overall scenario of lunar exploration, the interest for life sciences is twofold: on the one hand, the Moon provides an ideal platform for conducting various new types of investigations, some of which cannot be carried out elsewhere (e.g. the establishment of an artificial ecosystem on the Moon will be a challenging project per se). On the other hand, given the prospect of a lunar base, boundary conditions for human safety, health, well-being and working-efficiency have to be set. Scientific as well as operational aspects have to be looked into.

In the exobiology domain, studies *of, on,* and *from* the Moon would contribute to the understanding of the principles leading to the origin, evolution and distribution of life. The discovery of organic molecules on the Moon would support the model scenario reflecting the delivery of extraterrestrial organic molecules to the primitive Earth. Radioastronomy, by searching for other planetary systems – especially Earth-like planets – may provide clues for the debate as to whether the Earth is unique in the Universe in supporting life, or whether life is a universal planetary phenomenon. The Moon also offers an ideal platform for a variety of exobiological studies, such as on the role of radiation in evolutionary processes, on the environmental limits of life, as well as on life in simulated planetary environments (e.g. Mars or early Earth).

For radiation biology, the Moon provides a unique laboratory for investigating the biological importance of the various components of cosmic and solar radiation. On any space mission, especially those beyond the geomagnetic shielding, the radiation environment is of major concern not only for biological systems, including the human body, but also for sensitive electronic equipment. The installation of a solar-flare-alert station on the Moon would clearly be of benefit for radiation-protection purposes in the context of overall space activities.

One of the most challenging projects falling into the ecology domain would be to establish an artificial ecosystem for the first time on a celestial body beyond the Earth. This venture could be expected to provide a better understanding of the dynamic processes regulating our terrestrial ecosystem. A reliable Life-Support System (LSS) is a *sine qua non* pre-condition for a lunar base. A biogenerative LSS concept should then be gradually implemented within the artificial environment of the base. Precursor missions would be required to establish and optimise artificial ecosystems using lunar resources, and to gather detailed information on the adaptive strategies of organisms to the lunar environment via multi-generation experiments.

In the framework of a stepped approach to lunar exploration, life-sciences objectives could be achieved by implementing the following strategy:

- A lunar orbiter, accommodating active dosimetric packages, would monitor the radiation history during trans-lunar cruise and in the vicinity on the Moon.
- Automated surface stations, including rovers, would serve the needs of several disciplines. Several of these experiments could be improved by returning samples to Earth for detailed investigation.
- The most sophisticated requirements have to be met for the setting up of a manned lunar base.

Recommendation 3

In the life-sciences domain (Science on the Moon), two major interrelated goals have been identified: (i) to establish an artificial ecosystem beyond the Earth; (ii) to provide information pre-requisite to the establishment of a manned lunar base, including monitoring solar activity (predicting and alerting). It is recommended that these goals be further pursued through a study approach.

Recommendation 4

To implement the above studies, we recommend the setting up of specific study teams involving appropriate expertise to cover the various scientific thrusts that have been identified. Moreover, the Lunar Study Steering Group (LSSG) should continue its activity of supervising and coordinating these studies.

Several fields considered in the present study, such as second-generation VLF astronomy (involving large structures), high-energy physics (cosmic rays), and the exploration of the Moon for Helium-3 (^3He) as a fuel for fusion, require the availability of a well-established lunar base and are therefore not suitable for early implementation. These fields should be reconsidered at a later stage in the study.

Important technical areas such as technology development, transportation systems, lunar infrastructure and the role of man have not been addressed in the study. In the next study phase, however, technical requirements will emerge from the mission concepts considered.

Recommendation 5

We recommend enlarging the scope of the overall study activities to include these areas, in particular transportation requirements, technology development and human exploration, and evaluating ESA's possible role in each of them by exploiting the technical expertise available in the various Directorates of the Agency.

Implementation of the scientific-exploration activities will require the development of appropriate infrastructures and will eventually lead to the presence of man on the Moon's surface. As briefly outlined in the study, this could ultimately disturb the lunar environment to a degree detrimental to science. To preserve the Moon's unique character, therefore, it will be essential to minimise any such disturbances.

Recommendation 6

Future exploration activities on the Moon must involve extensive scientific investigations prior to any significant disturbance of the lunar environment. Considering the future plans of the various space agencies, a 'lunar preservation plan' should be agreed upon at international level as early as possible.

The present study has identified fields of scientific investigation in which Europe has the potential to play a significant role in any future lunar exploration programme. Such a programme will require extensive scientific knowledge and technical expertise and considerable financial means, implying the need for international collaboration in the broadest sense, particularly with those space agencies already involved in the preparation of lunar programmes (NASA and ISAS).

Recommendation 7

Lunar investigations are preferably to be performed within the framework of a broad international collaboration. We recommend that such international coordination be initiated early in the study phase.

Annex 1

Annex 1

Astronomy Using VLBI with an Earth—Moon Baseline

J. Kovalevsky

1. Introduction

The system of positions of extragalactic radio objects will soon become the basis of the definition of a fixed reference celestial frame to which all motions of stars or objects in the Solar System are to be referred. This is now being worked out by the International Astronomical Union. Therefore, at the time of the establishment of a lunar base, quasars and other point-like extragalactic sources will be the primary fiducial points of all astrometry.

The current proposal deals only with the astrometric aspects of Moon–Earth VLBI, independently of other applications such as the study of the structure of the sources.

2. Achievable accuracy

The Moon–Earth baseline is of the order of 4×10^{11} mm. This will be (at the epoch) measurable to 2–3 mm accuracy using lunar laser telemetry, with two to three colours to eliminate the effect of refraction (such an instrument is already in an advanced planning stage). Clock synchronisation is possible by means of direct continuous observation. This means that it can achieve the precision of present hydrogen masers (10^{-13} or 10^{-14}). This will allow the determination a priori of the position of the central fringe to within a few wave periods, resulting in a reasonably small quantity of data to be collected. Technical efforts should then be devoted to obtaining a significant sample of the radio wave. With the advent of picosecond electronics, this should allow the work to be carried out at $l > 1$ cm with a sample of 10–20 points per wavelength. This would then result in an angular precision of better than 1 μarcsec.

3. Scientific objectives

Some expected astrometric returns are:
- reference frame to better than 1 μarcsec (and 1 μarcsec/yr)
- access to proper motions of some galaxies and quasars
- monitoring the Earth–Moon distance to a fraction of a millimetre. Such a sensitivity on lunar orbit and libration would open up large domains for new interpretations of the dynamics of the Earth–Moon system
- parallaxes of distant radio stars might also be obtained to an accuracy of 1 μarcsec, providing a 1% accuracy to the measurement of the distance to the centre of the Galaxy.

4. Conclusion

This VLBI technique applied to astrometry would be the most powerful method of leading astrometry to a microarcsecond level of accuracy. It seems unlikely that astrometric instruments based only on the Moon could achieve such an objective.

Annex 2

Annex 2
A Possible Lunar Instrument for Relative Astrometry

E. Høg

Summary

The concept of establishing a telescope on the Moon for accurate observation of relative parallaxes and proper motions has been investigated. Up to a million stars in 500 selected fields of about 0.5×0.5 square degrees can be observed to faint magnitudes. An accuracy of $\sigma = 2$ marcsec is achieved by Hipparcos for stars at 11 mag, whereas this accuracy would be obtained at $V_{2mas} = 19.9$ mag with a lunar-based telescope with an aperture of 60 cm. At $V_{0.1mas} = 15.5$ mag, a standard deviation of 0.1 marcsec is expected for parallaxes and annual proper motions.

With a telescope of $D = 100$ cm and a low background, these limits are 2.0 mag fainter.

The error of 0.1 marcsec on relative proper motions is better than the best measurements currently available in only two or three fields, and it would be obtained for a hundred times as many fields in the sky. This could probably not be achieved by any other means for such faint stars for many decades to come.

These possibilities have significant implications for the cosmic distance scale, dynamics of clusters and the Galaxy, masses of stars and clusters, theory of stellar interiors, discovery of planetary systems, rotation of the Moon, etc.

We shall see that the very slow rotation of the Moon is a fundamental consideration for utilising a CCD without being drowned in read-out noise. The slow rotation of the Moon makes it quite unsuitable for all-sky coverage, for which a free-flying satellite such as Hipparcos is the obvious choice. It therefore seems of interest to limit astrometry from the Moon to cover only a small fraction of the sky.

1. Introduction

An astrometric instrument on the Moon must be justified by the important scientific results it could obtain. The chance of it being realised and subsequently functioning properly is greater if its design is kept simple. The following proposal is intended to meet these criteria.

The three most important properties of the Moon as a base for astrometry are:
(i) the vacuum, allowing the optics to perform at the diffraction limit
(ii) the Moon itself, as a stable platform
(iii) the very slow rotation of the Moon, compared with a free-flying satellite like Hipparcos, and with the Earth.

The simplicity requirement leads to the consideration of only relative astrometric observations within the telescope's field of view. Thus, relative parallaxes and proper motions with an accuracy about 0.1 marcsec will be obtained in fields of less than a square degree. Large angles are more difficult to measure with such accuracy and are not considered here. However, the possibilities should be carefully studied as previously suggested: global astrometry by reference to the direction of the local vertical, or with the use of angular reference optics, or by means of interferometry.

A telescope for accurate observation of parallaxes and proper motions is discussed, choosing 60 cm aperture as a baseline. Relative astrometric parameters for all stars to a limit of $V=15.5$ mag in 500 selected fields of about 0.5×0.5 square degrees can be obtained. An accuracy of 0.15 marcsec for parallaxes and 0.10 marcsec/yr for proper motions is expected if the observations are pursued for five years (Table A2.1). This parallax error is ten times smaller than that obtained by Hipparcos or by the best current ground techniques. The Hipparcos accuracy of 2 marcsec can be obtained down to magnitude $V_{2mas}=18.9$.

The current best relative proper motions have an accuracy of $0.1-0.15$ marcsec/yr and are obtained only in the Pleiades and two other clusters. Since this accuracy depends primarily on old epoch photographic plates with a limiting magnitude of $B=14$, or about $V=13.5$, it cannot be significantly improved upon in the forthcoming decades.

The average sky contains 330 stars brighter than $V=15.5$ per field of 0.25 square degrees, and many more in clusters. The 500 fields would cover more than 150 000 stars down to this magnitude.

Perturbations in the motions of stars due to faint companions or planets could be detected with much more success than by photographic astrometry.

The relative parallaxes obtained directly in a field must be converted to absolute parallaxes by a study of all stars in the field, including the use of the best possible photometric parallaxes of the most distant stars. It is estimated (F. van Leeuwen, priv. comm.) that this conversion would normally reduce the accuracy by a factor of four (at $V=15$

Table A2.1 Accuracy (std. dev.) of observations. For $D=60$ cm telescope: 100 s observation on vertical slits; the longitude from one field crossing of a star at the ecliptic; 30 field crossings in five years on the Moon; the std. dev. of the parallax is 1.4 times larger than of annual p.m. Factors G and H should be applied to the previous columns to obtain the accuracy for an aperture $D=100$ cm, and for $D=100$ cm and a background lowered to 25%, respectively

V (mag)	100 s vertical slits (mas)	1 crossing longitude (mas)	5 years parallax (mas)	5 years pr. motion (mas/yr)	G $D=100$ cm	H $D=100$ cm low I_b
12	0.232	0.050	0.013	0.009	0.354	0.333
13	0.414	0.089	0.023	0.016	0.350	0.306
14	0.808	0.173	0.045	0.031	0.343	0.267
15	1.73	0.370	0.095	0.066	0.338	0.228
16	3.99	0.855	0.220	0.153	0.336	0.198
17	9.67	2.075	0.535	0.371	0.331	0.180
18	24.0	5.14	1.327	0.921	0.329	0.171
19	59.9	12.82	3.31	2.294	0.330	0.167
20	150.1	32.12	8.29	5.750	0.329	0.166
21	376.8	80.64	20.8	14.43	0.329	0.165

from 0.1 to 0.4 marcsec), but this remains a very good result, compared with, for example, the 2 marcsec from Hipparcos down to $V=11$ mag. The conversion of relative to absolute proper motions is not quite as difficult. The conversion from relative to absolute astrometry will be greatly enhanced by the faint limiting magnitude, and in some cases extragalactic stars or quasars may be used.

2. Telescope and location

The instrument should be located at the lunar equator so that most of the sky can be 'seen'. It should be located some 60° or more from the Sun–Earth point to avoid being disturbed by the reflection from the Earth. Location at the Moon's limb or on the far side is to be preferred.

An astrometric instrument should utilise the stability of the Moon as a platform. The instrument itself and the vicinity should therefore be protected against direct sunlight. This can be achieved by means of a narrow structure that produces a shaded area, oriented east-west since the Sun always passes within 1.5° of the zenith of an observatory at the equator. The instrument itself should consequently be mounted horizontally, east–west in the shaded area (Fig. A2.1). Super-insulation of the telescope and the immediate surroundings is required.

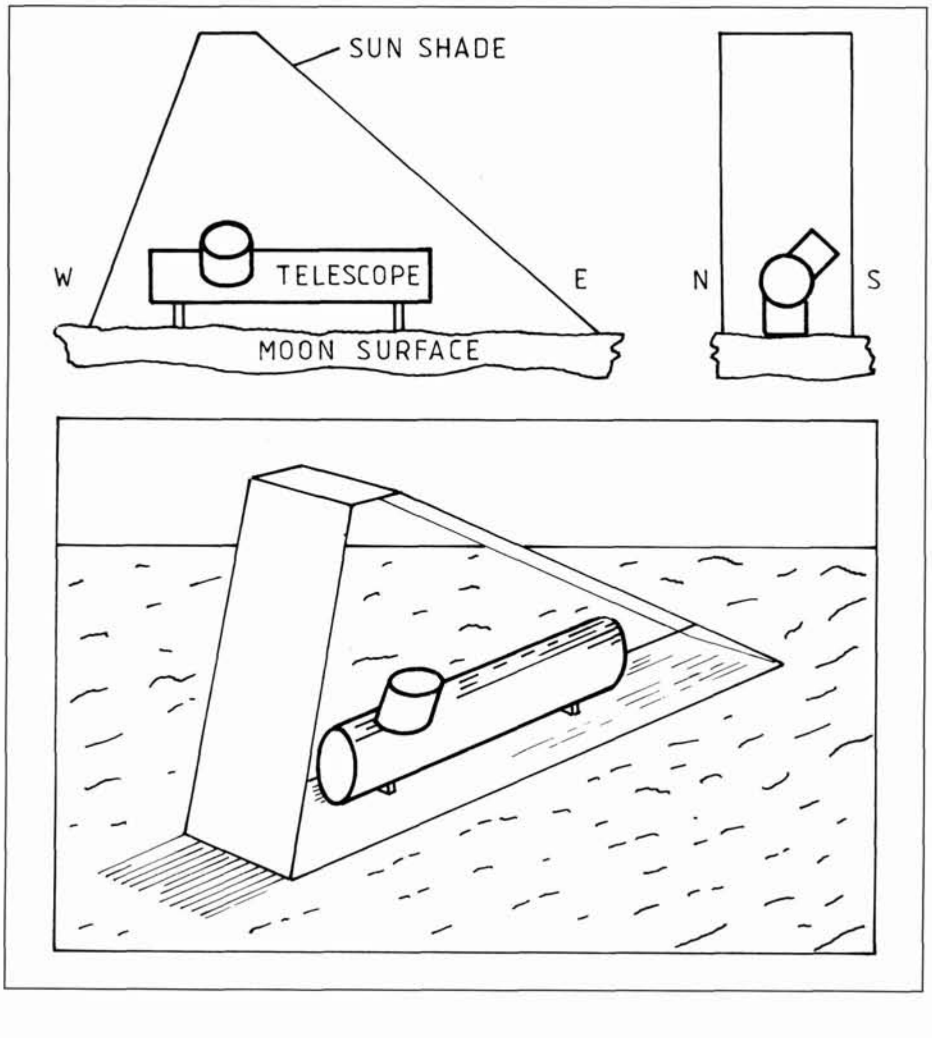


Figure A2.1 Meredian telescope on the Moon

Observations should be carried out at the meridian because a single rotation about an east–west axis would be sufficient to direct the instrument to any required field of the sky when it crosses the meridian. A star of lunar latitude B relative to the Moon's equator will move with a speed of $v = 0.55 \cos B$ ″/s. If the field of the telescope is 0.5°, a star will stay in the field for at least 0.9 h, which will allow very accurate observation with a moderate aperture even of faint stars.

A simple meridian instrument (Fig. A2.2) requires a flat mirror at 45° to the east–west line to direct the light into the telescope. The flat mirror with a micrometer, or the whole telescope, is rotated before observation to point at a new field arriving at the meridian.

If a 1 h observing time is allocated to each field, up to 650 fields could be observed once per month, and twelve times per year if observation could be continued during the lunar day. In the following, it is assumed that observation is efficient only during the lunar night, and that 30% of this time is required to rotate the mirror for calibration, etc. This leaves 1 h six times per year for observation of 500 fields, corresponding to a total useful observation period of only 35% of real time.

The fields may be located anywhere on the sky, but not too close to either the ecliptic or its poles. Nearer to the poles than approximately 10°, the motion is too slow. Nearer than 10° to the ecliptic, any observation is obstructed by the shade structure.

Optical system
The optical system (Fig. A2.2) is a 'glass meridian circle', originally designed for measuring large angles, such as the traditional meridian circle (see Ref. 4). For use on the Moon, it offers the advantages of being a horizontal instrument and a simple reflective telescope.

The main mirrors S1 and S2 are flat and spherical, but with slight deformations. A field of at least 0.25 square degrees without aberrations can be achieved if the aperture ratio is less than 1:10. This implies a focal length of at least 6 m. The whole instrument would then be 7 m long.

The auxiliary mirrors S2 and S4 are rigidly connected to S1 and S3, respectively, though not necessarily by solid glass as shown. The centre of curvature of S4 is at

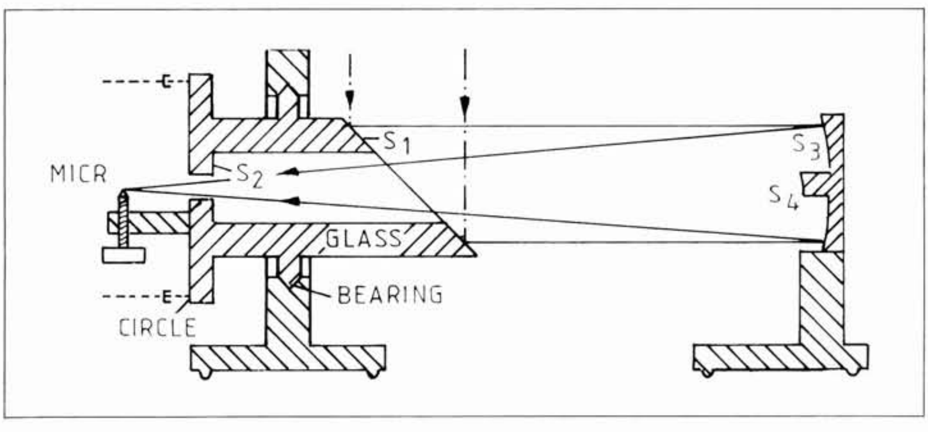


Figure A2.2 Optical principal of Moon meridian telescope

the focus and serves to determine the tilt of the S3 – S4 double mirror. The plane S2 mirror determines the tilt of the S1 – S2 double mirror by auto-collimation.

3. Detector system

The proposed modulation and detection system consists of a focal grid (Fig. A2.3) and a CCD detector. An optical relay system forms an image of the grid on the CCD which detects the modulated light from the moving stars, similar to the Hipparcos IDT. The light from a programme star is integrated during a sampling interval by a number of pixels centred on each star. The counts of electrons in these pixels are added, giving one sample per star per interval. A numerical analysis of these counts will provide the grid coordinates for each star as a function of time, and hence the positions of the programme stars in the field of the telescope.

(a) Star mapper and auto-collimation
The focal grid contains a star mapper of a 'vertical' and several inclined slits, all imaged by the relay system on the same CCD as the main grid. The star mapper is used to calibrate and occasionally verify the setting of the telescope. An accuracy of 0.1 arcsec can easily be obtained from a single crossing of a star of known position.

The small field *F* contains slits suitable for measuring the images of illuminated pinholes in the field *E*. These images are formed by auto collimation on the auxiliary mirrors of the telescope.

(b) Main grid
The main grid with an area about 30′ ×30′ contains slits of width w,s = 0.28 and period 0.74 arcsec. These values are derived from the Hipparcos values w,s = 0.46, 1.208 arcsec by scaling inversely with the aperture diameters D and proportionally with the wavelengths λ_{max} (Table A2.2), i.e. by a factor 0.610. In this way, the same modulation is obtained as that of Hipparcos, chosen to be the optimum for single and double stars.

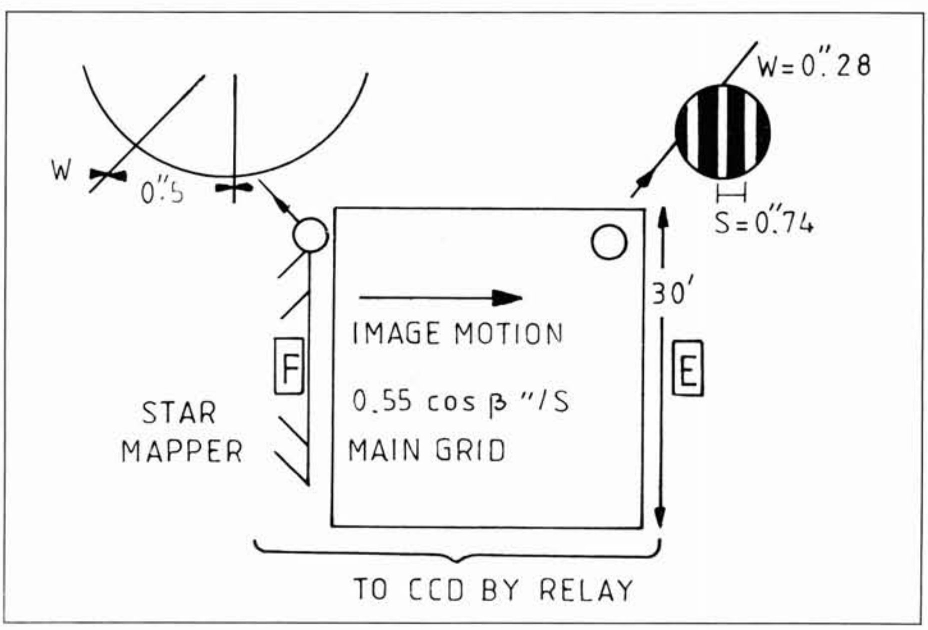


Figure A2.3 Focal grid of Moon meridian telescope. The main grid contains slits with width and period as shown for the 'vertical' direction, perpendicular to the image motion. Other slits at 45° inclination are not shown. The fields *E* and *F* contain light sources and slits for auto collimation. All slits, including the star mapper, are imaged on a single CCD

Table A2.2 Optical transmittance T at λ_{max} of detection chain, taking into account all payload elements. I_0 is predicted for a star of $V=10$, $B-V=0.6$ (a blank means that a number is equal to the preceding column).
Performances: V_σ is the visual magnitude at which the accuracy σ is reached with a five-year Moon mission. Number of stars per field, N, with $V<V_\sigma$ on the average sky

	Hipparcos	$D=60$ cm	$D=100$ cm	$D=100$ cm low I_b
Spectral range (nm)	375—750	375—800		
λ_{max} (nm)	475	600		
FWHM (nm)	200	(>) 200		
Detector type	IDT, S20	CCD (RCA)		
T_{optics}	0.73	0.79		
$T_{detector}$	0.08	0.73		
T_{total}	0.060	0.57		
Source	Reference, 1 p. 63	Reference 1+2, p. 143		
$\log f$ (erg cm^{-2} s^{-1} nm^{-1})	-11.38	-11.50		
Source Reference 3, p. 207				
D, diameter (cm)	29	60	100	100
A, area of pupil (cm^2)	660/2	2827	7853	
I_{ph}, photons at pupil (Hz)	65800	542000	1505000	
I_0, predicted for $V=10$ (Hz)	1503	117600	327000	
I_0, observed for $V=10$ (Hz)	707	(55000)	(153000)	(153000)
Source ESTEC.STATUS.11-7				
w, s (arcsec)	0.46, 1.21	0.28, 0.74	0.168, 0.444	
M_1, M_2	0.74, 0.26	0.74, 0.26		
I_0 for $V=14$ (Hz)	—	1380	3833	
I_0 for $V=14$ (counts/sample)	—	180	287	
Sampling interval (s)	—	0.125	0.075	
Read-out noise (rms)	—	10	10	?
I_{read} (Hz)	—	800	1333	?
Sky background (mag)	—	15.2	15.2	?
I_{sky} (Hz)	—	1200	3333	?
I_b (Hz)	—	2000	4667	1167
Performances:				
V_{2mas} (mag)	11	18.9	20.1	20.8
$V_{0.1mas}$ (mag)	—	15.5	16.8	17.5
$N(V<V_{0.1})$ (stars per field)	—	330	930	1570

A major difference from Hipparcos is required regarding the direction of the slits. These were all 'vertical' (Fig. A2.3), perpendicular to the stellar motion, whereas relative measurement in two dimensions requires some slits to be inclined. It is proposed to divide the main grid into 30 vertical bands of 60 arcsec containing in sequence 81 vertical slits, 57 slits inclined at 45°, 57 inclined at 45° with respect to the other side, etc., repeated ten times.

The two directions of inclined slits would be sufficient for two-dimensional measurement, but since more information about parallaxes is contained in the ecliptic longitude, it is probably worthwhile including the vertical slits, which enable direct measurement of the longitude.

(c) CCD detector
The focal grid is imaged onto one end of a rectangular CCD containing 300×800 square pixels. Each pixel corresponds to 6×6 arcsec in the focal plane. The relay

optics should be designed to smear the star image on a circle of about 12 arcsec diameter in order to minimise the modulation due to individual pixels. An area of 5×5 pixels (= 30×30 arcsec) centred on each programme star will integrate the light during a sampling interval T1=0.125 s. After each sampling interval, the charges of the 300×400 pixels from the main grid and the star mapper are shifted by 'frame transfer' to the other end of the CCD in about 2 ms, after which the integration can restart. The read-out by binning will take place during the integration at the other end of the CCD.

The read-out noise under such circumstances is 15−20 rms electrons with present techniques, a value expected to improve to less than ten electrons within ten years. This noise figure applies to the read-out of a single pixel and also to the read-out of an area of, for example, 5×5 pixels with binning. It also applies if, for instance, 400 areas are read out.

Read-out noise is inversely proportional to the square root of the time interval used, in this case 100 ms. This long interval is due to the slow rotation of the Moon. A scanning satellite such as Hipparcos must rotate 100 times faster so that the sampling interval must be 100 times smaller for a given telescope resolution. Each sample will therefore contain 1/100 of the photo electrons, together making the signal-to-noise ratio per sample due to readout 1000 times smaller, whereas that due to photon noise only decreases to a tenth. This is the reason a CCD is not directly usable for a scanning astrometric satellite.

(d) A CCD at the focus?
For accuracy, it seems important that a modulating grid be used instead of direct imaging on a CCD because the irregularities in a CCD are believed to be larger than those in a specially manufactured grid. If direct imaging is considered, the individual pixels would have to be read out, resulting in a higher read-out noise than with binning. Ideally, if none of these problems existed, a CCD with smaller pixels corresponding to the telescope resolution would of course be more efficient, and this possibility should be carefully analysed with realistic figures for CCD irregularities.

The optimum pixel size should be approximately equal to the above slit width, w = 0.28 arcsec. The proposed field would therefore require a chip with 6000x6000 pixels or a mosaic of smaller chips.

The maximum possible gain in relation to the photon noise with a grid is probably equal to a weight factor s/w = 2.6 in both coordinates, no matter how small the pixels are. This gain should be compared with the problems of the large number of pixels, their irregularities and the read-out noise.

4. Performance

The phase of the modulated signal generated by the main grid contains all the astrometric information. The statistical error from photon noise is evaluated below, including sky background and read-out noise from the CCD. Other error sources from the grid, the optics and the detector should be negligible. Grid effects would be four times smaller than those for Hipparcos due to the 6 m focal length, and they would be constant for a given star if the star always crosses almost the same parts of the grid.

Detector and optical effects would be quite constant for the same reason. These effects, however, can be calibrated from the star observations utilising the completely smooth rotation of the Moon, and scans with some overlap in the latitude direction.

(a) Optical transmittance
The spectral transmittance of the Hipparcos main detector chain is given in Reference 1. The values from 'spectral range' to T_{total} given in the Hipparcos column of Table A2.2 are derived from here and need no further explanation.

Corresponding values in the $D = 60$ cm column are derived from the Hipparcos transmittance in Reference 1 and the CCD quantum efficiency in Reference 2. This is on the conservative side in several respects. The Hipparcos optics do not use the red sensitivity of a CCD. The relay optics, especially, could be improved. Nevertheless, the FWHM is about 300 nm (from 425 to 725 nm), much more than the 200 nm adopted in Table A2.2.

Thanks to the high value of $T_{detector}$ of the CCD compared with the S20 cathode, the value of T_{total} is ten times higher for $D = 60$ cm.

(b) Count rates
The flux f at $\lambda_{\max}$ in Table A2.2 for a star of $V = 10$ and $B-V = 0.6$ is obtained from Reference 3. The telescope diameter D and area A of the entrance pupil are given. The number of photons passing through the pupil is

$$I_{photons} = A\, f\, FWHM/(hc/\lambda) \tag{1}$$

and the mean intensity of the modulated signal is

$$I_0 = I_{photons}\, T_{total}\, w/s \tag{2}$$

This predicted I_0 is 1503 Hz for Hipparcos, while only 707 Hz is observed. This discrepancy needs an explanation, but for the present a corresponding lower value of only 55 000 Hz for $D = 60$ cm is adopted, only half of that predicted. Thus

$$I_0 = 55000\ 10^{-0.4(V-10)} \tag{3}$$

This results in $I_0 = 1380$ Hz for $V = 14$ (Table A2.2), or 180 counts per sample with a sampling interval TI $= 0.125$ s.

A read-out noise of y electrons rms should be added quadratically to other noise. It is therefore equivalent to a background intensity of y^2 counts/sample with Poisson distribution, or

$$I_{read} = y^2/TI \text{ Hz} \tag{4}$$

giving 800 Hz in Table A2.2.

The background of the mean sky is equivalent to 215 stars of 10th mag deg^{-2} in the visual. This is taken from Reference 3, for faint stars plus zodiacal light plus diffuse galactic light. It is equivalent to one star of 15.2 mag in V per sky area of 490 arcsec^2. This is the area of the slits inside 36×36 arcsec, which is larger than deemed necessary in Section 3 above.

The resulting count rate is I_{sky} in Table A2.2, and the total background of 2000 Hz is

$$I_b = I_{read} + I_{sky} \tag{5}$$

(c) Photon noise
The photon noise of the best possible estimate of the astrometric information is discussed in Reference 1. The Cramer-Rao statistical limit is

$$\sigma^2_{CR} = (s/2\pi)^2 N^{-1}/g(M_1, M_2) \tag{6}$$

where s is the split period (mas), T is the observing time (s), $N = I_0 T$ is the total number of photons, M_1 and M_2 are the modulation coefficients, and

$$g(M_1, M_2) = \text{average } [(M_1 \sin x + 2M_2 \sin 2x)^2/(1+M_1 \cos x + M_2 \cos 2x)] \tag{7}$$

taken over $x = 0$ to 2π. The average over twelve points between $x = 0$ and π was used. Equations (6) and (7) are due to Lindegren (priv. comm.), while Equation (3.5) in Reference 1 is incorrect. The background effects are taken into account by modifying I_0, M_1, M_2 as in Reference 1.

The resulting accuracy is σ_1 for $T = 1$ s observation on 'vertical' slits. Table A2.1 gives the accuracy for 100 s observation. The variances at vertical and inclined slits in the direction of motion are

$$\sigma^2_{vert} = \sigma^2_1/T \text{ and } \sigma^2_{incl} = 2\sigma^2_1/T \tag{8}$$

The observation of a star during a complete field crossing at the ecliptic becomes less accurate than at other latitudes because of the shorter crossing time, but this is adopted in Table A2.1 for simplicity. Crossing the ten triple bands takes 30 T where $T = 109$ s. This observing time is assumed for simplicity in the following, although stars starting within the field will be observed for a shorter time. The resulting variances of longitude (Table A2.1) and latitude are obtained from

$$\sigma^2_L = (20T)^{-0.5}\sigma_1 \tag{9}$$

and

$$\sigma^2_B = (10T)^{-0.5}\sigma_1 \tag{10}$$

(d) Astrometric parameters
We consider J observations spaced at equal distances in time Δt and of accuracy σ_0. The variance of the mean position is

$$\sigma^2_{pos} = \sigma^2_0/J \tag{11}$$

The parallax has the largest effect in longitude and is the same for all stars if observations are well distributed on the parallactic ellipse, as they will be from lunar night oberving. Neglecting latitude observations in order to be conservative and simple, the variance is

$$\sigma_{par}^2 = 2\sigma_0^2/J \tag{12}$$

where we take $\sigma_0 = \sigma_L$. A mission of five years gives $J = 30$ crossings and the accuracy in Table A2.1. The accuracy for positions becomes identical to that for parallax if we take $\sigma_0 = \sigma_B$ to be conservative.

The variance of the proper motion for not too small J is

$$\sigma_{p.m.}^2 = 12\, J^{-3} \Delta t^{-2} \sigma_0^2 \tag{13}$$

Again, the conservative value $\sigma_0 = \sigma_B$ is adopted for the accuracy of proper motion in Table A2.1.

The parallax, proper motion and perturbations of the motion of a star will be less affected by the effects of grid, optics and detector than the relative positions, if the stars always cross nearly the same parts of the grid. It is therefore reasonable to believe that the accuracies in Table A2.1 can in fact be achieved.

(e) Stability

The scale value in the direction of longitude will be determined by the Moon's rotation. This is a great advantage over photographic astrometry where the scale value is uncertain unless accurate reference stars are available in the field. This has sometimes led to spurious determination of expansion of a star configuration. For the lunar instrument, the scale value in the latitude direction cannot be directly determined, but if the scale changes with time this change would be equal in longitude and latitude, allowing safe detection of a very small expansion or contraction.

If observation of a field is continued for several hours, the stars in a band of several degrees length and 0.5° width will be connected, provided the instrument is very stable. The drift should be less than 0.1 marcsec/h corresponding to the accuracy of one crossing, and preferably much smaller.

(f) Aperture and background

The presentation has, for the sake of clarity, focussed so far on a single baseline proposal with a telescope diameter of D=60 cm. A diameter of 100 cm has been considered, and column G in Table A2.1 gives the factor with which the previous columns should be multiplied to give the accuracy for D = 100 cm.

The assumption on background was conservative. The sensitive area on the CCD of 30×30 arcsec^2 could probably be reduced to 10×10, so that I_{sky} = 370 Hz. The number of pixels on the chip should probably be increased by the same factor. The read-out noise would not change, but it would then dominate the sky contribution. It is however reasonable to assume that future technology will permit a total I_b of only 25%. Column H in Table A2.1 gives the factor.

The last lines in Table A2.2 on performance have been derived by means of the last three columns in Table A2.1. V_σ is defined as the magnitude at which the accuracy σ is reached for annual proper motion with a five-year mission.

It appears from Table A2.2 that σ = 2marcsec can be obtained for D = 60 cm at eight magnitudes fainter than that with Hipparcos. This 2 marcsec magnitude becomes V = 20.8 mag for D = 100 cm and the low background.

The super-accuracy of 0.10 marcsec is obtained at V=17.5 for D = 100 cm and low background.

5. Concluding remarks

- The flux discrepancy should be resolved.
- Colour indices of all stars could be obtained by providing a few vertical bands in the focal plane with colour filters.
- The optical aberrations of the telescope should be calculated.
- To what extent is Equation (6) valid at very low modulations and faint stars?
- The possibilities of using a CCD directly in the focus without a grid should be analysed.
- The proposal by Gatewood for an astrometric telescope in connection with the Space Station should be studied.
- The problems of conversion from relative to absolute parallaxes and proper motions should be studied for the given circumstances.
- The optimum scientific utilisation of the above possibilities should be studied.

References

1. Perryman M A C et al. 1989, The Hipparcos Mission, ESA SP-1111, Vol. I.
2. Linde P 1987, Preprocessing of CCD Images, *Proceedings of 'Nordisk Forskerkursus'* (R Florentin Nielson, Ed.).
3. Allen C W 1973, Astrophysical Quantities III.
4. Høg E 1973, Design Study of a Glass Meridian Circle, *Mitt. der Astron. Ges.* No. 32, p. 120.

Annex 3

Annex 3

An Advanced Lunar-Based Solar Observatory

O. von der Lühe

Summary

Concepts for an advanced lunar-based solar observatory have been developed according to the scientific objectives discussed in the 'Solar Physics from a Lunar Base' report. The need for high angular resolution implies long baselines in the collecting aperture of the observatory's telescope. One of the main obstacles for the implementation of this goal is the disposition of the large solar flux that an instrument with the required angular resolution is likely to collect. Surprisingly, the solar-surface intensity is too low to provide enough photons to fulfil simultaneously all scientific goals concerning angular, spectral and temporal resolution, and the sensitivity for a reasonable telescope in the visible spectral domain. Trade-offs between these goals will have to be made on a case-to-case basis.

A number of concepts for the optical layout of an observatory are presented. It is concluded that a Michelson interferometer array with a diameter of 10 m and about twelve array elements, consisting of 1.6 m Gregorian telescopes mounted on a common structure, is the best option.

1. Introduction

The report by the Solar Physics Section of the Lunar Science Study Team, entitled 'Solar Physics from a Lunar Base', presents a set of scientific goals for an advanced, lunar-based solar observatory. The primary scientific goal is an increase in angular resolution by a factor of 100 over that currently and routinely achieved with ground-based telescopes, i.e. approximately 1 arcsec. Current developments indicate that by the turn of the century advanced technology and data-analysis methods (adaptive optics and speckle methods), as well as metre-class space observatories, will improve angular resolution to about 0.1 arcsec. The additional increase in resolution by another factor of 10 cannot be realistically achieved without a major effort, such as the establishment of a large-scale lunar observatory.

Many of the other specifications, such as spectral and polarimetric resolution and, to some extent, time resolution, do not represent an advance beyond current capabilities, and these specifications are as significant as the increase in angular resolution. Many concepts for spectroscopic and polarimetric equipment in existence today will most likely play a major role in the setting up of an advanced lunar observatory. The achievement of higher angular resolution will have the greatest impact on the concept, and is discussed at length in this Annex. The dimensions of the observatory depend on the observed wavelength. Achieving the resolution goal for visible wavelengths of about 500 nm results in 10 m structures pointing to the Sun, adopted here as representing the upper size limit of a structure to be used on the lunar surface. Hence, a 0.01 arcsec resolution does not appear to be feasible for much longer wavelengths, i.e. in the infrared range beyond 1 μm. This annex concentrates on an observatory concept fulfilling the angular-resolution requirement at visible wavelengths. Altering the basic concept to accommodate shorter wavelengths (near- and vacuum-ultraviolet) is relatively straightforward.

The second section revises the goals and derives some fundamental specifications for an observatory. This allows the assessment of the main characteristics, including the dimensions of the observatory, as well as possible fundamental problems. Most of the goals are also discussed in the context of current technological capabilities, in order to identify needs for technology development. The third section presents possible strategies for the implementation of the goals. Major elements of the observatory are identified and described in some detail.

2. Implementing science goals

2.1 Angular resolution

The scientific goal is to resolve 0.01 arcsec, corresponding to 7.4 km on the solar surface, within as broad a wavelength range as possible. The angular resolution of a telescope is fundamentally limited by diffraction at its entrance aperture, resulting in a diffraction profile in the telescope focus as the image of an idealised, point-like source – Point Spread Function (PSF). The diameter of the PSF is proportional to the diameter of (or the largest dimension within) the entrance aperture, and inversely proportional to the detected wavelength of the radiation. Therefore, a necessary prerequisite for achieving high angular resolution is a sufficiently large baseline B in the entrance aperture of the telescope, i.e. the largest span within the entrance opening of the instrument. There are various options for achieving a long baseline, including (i) a connected collecting area of sufficient diameter (single-aperture telescope), and (ii) several disconnected collecting areas which combine the light in a coherent fashion, and cover an area of sufficient size (multiple-aperture interferometer).

The intensity distribution of any astronomical object can be described as being composed of an infinite number of two-dimensional sine and cosine patterns in the sky, resembling straight fringes with varying spacing and orientation. The amplitude and phase of a term with a given spacing and orientation can be obtained by evaluating the two-dimensional Fourier transform of the object's intensity distribution. The frequencies of the Fourier components can be expressed in 'line-pairs per arcsecond', where a line-pair comprises a full sine period (one bright and one dark line). A sine intensity pattern with an angular frequency of one line-pair per arcsecond therefore has one full period covering one second of arc on the sky. Traditionally, the (two-dimensional) frequency space of Fourier components is referred to in interferometry as the 'UV domain' or 'UV plane'.

Any aberration-free imaging instrument can be viewed as a linear filter that transmits the object's Fourier components at various frequencies with differing efficiency. The Fourier-component transmission efficiency as a function of angular frequency is called the 'Modulation Transfer Function' (MTF) of the telescope. Because of diffraction at the entrance aperture, no information is transmitted by the instrument for Fourier components at angular frequencies above a certain limit, called the diffraction limit of the telescope. The diffraction limit is at an angular frequency of $2\text{x}10^5$ B/λ line-pairs per arcsec, where B is the maximum baseline, typically the diameter of the telescope. A resolution of 0.01 arcsec corresponds to a diffraction limit of $R_\alpha=100$ line-pairs per arcsec. The baseline for a given observation wavelength λ is therefore:

$$B = 2\times10^5 R_\alpha \lambda, \text{ or in this case, } B = 2\times10^7 \lambda \quad (1)$$

Table A3.1 summarises the required baseline as a function of wavelength.

Table A3.1 Baseline as a function of wavelength

Spectral range	λ (nm)	B (m)
EUV	10	0.2
	100	2.0
UV	200	4.0
Visble	500	10.0
	1000	20.0
Opacity minimum	1600	32.0
Near-IR	2000	40.0

The requisite baselines may not be feasible for wavelengths longer than 500 nm, and most certainly not with a filled-aperture instrument. The baseline to some extent determines the collecting area of the instrument, and the large amount of solar energy incident on the instrument becomes a concern. A 10 m single filled-dish telescope has a collecting area of some 70 m^2, and the resulting collected flux amounts to some 100 kW, most of which must be disposed of somehow. Other options are diluted-aperture telescopes, such as interferometers, or telescopes whose collecting area consists only of a thin ring (annular apertures), thereby reducing the collecting area while preserving the required baselines.

It is important to design a diluted aperture in such a way that complete information from the source to the resolution limit is instantly available in order to meet the time-resolution requirements. This means that all Fourier components in the solar image are measured simultaneously. Linear array configurations, which do not provide instantaneous two-dimensional coverage of all Fourier components, should therefore not be considered.

2.2 Temporal resolution

The scientific goal of achieving a temporal resolution of 0.1–1.0 s presents a technical problem for current detector technology. Large-scale panoramic detectors, e.g. 1024x1024 pixel CCDs, have readout times of a few seconds, and it can be expected that, due to technology evolution, the problem will be solved by the time faster detectors are needed.

The goal of 0.01 s temporal resolution presents no problem for today's small-format detector arrays (e.g. 32x32 Reticon arrays), but may be difficult for large panoramic detectors to handle. In the latter case, dedicated development of fast-readout detectors must be undertaken.

The exposure time limits the number of detectable photons, a factor that is significant for signal-to-noise considerations and will be discussed later.

2.3 Spectral resolution

A scientific goal of $R_\lambda = \lambda/\Delta\lambda = 5\text{x}10^5$ does not seem to present a problem. Several options for highly efficient spectrometers exist today. Again, this requirement will limit the number of detectable photons and have a significant impact on the signal-to-noise ratio.

2.4 High polarimetric accuracy

A requirement to measure polarised light with a precision of 10^{-4} (Stokes I) places stringent limits on instrumental (de-)polarisation effects. An optical design exhibiting high rotational symmetry up to and including the polarisation modulator is mandatory. In the case of a diluted-aperture instrument, this necessity leads to a symmetric configuration which, to a certain extent, will cause redundancy.

2.5 Sensitivity

The polarimetric accuracy goal implies a signal-to-noise ratio of 10^4 for a single measurement (one angular resolution element, one spectral channel, one polarimetric channel, one exposure) in the continuum. Hence, at least 10^8 photons must be detected in order to overcome the noise from just the photons.

With current technology, there is a significant problem in the detection of such a large number of photons in a single measurement. CCD pixels typically have capacities which fall short by 2 to 3 orders of magnitude of the initial 10^8 photo-electrons. The scale of this problem indicates that either the goal needs substantial descoping, or that substantial technology development, possibly exploring entirely new approaches to high-flux detection, is necessary.

A fundamental problem arises, however, due to the limited luminosity of the Sun. The number N_o of photons per resolution element, per time unit, and per unit wavelength available at the entrance aperture of the telescope is independent of the diameter of a filled collecting aperture. This is because the diameter of the resolution element (and hence the area on the solar surface that contributes to the solar constant) varies inversely with the diameter of the instrument. The increase in collected flux brought about by the increased instrument diameter is exactly cancelled by the decrease in 'collecting area' on the solar surface. The only way to increase the collected flux within a solid angle of 0.01 arcsec is to increase the diameter of the collecting area beyond the baseline required for a 0.01 arcsec resolution.

Table A3.2 Solar photon flux per instrument resolution element at visible wavelengths

λ (nm)	N_o ($Å^{-1}$ s^{-1})
400	0.52×10^9
500	1.40×10^9
600	2.19×10^9
700	2.87×10^9

The number of photons N_o is wavelength-dependent and has a broad maximum in the red/infrared regime. Table A3.2 lists N_o for visible wavelengths. A few simple calculations quickly reveal that the solar luminosity is too small to permit the achievement of all scientific goals in terms of angular resolution, spectral resolution, and temporal resolution with a sensitivity of 10^{-4} in intensity, even if the instrument had an unrealistic efficiency of 100%. This is also indicated in Figure A3.1, in which the achievable spectral resolution of approximately λ=500 nm is shown as a function of exposure time, for a measurement within a single resolution element of 0.01 arcsec. The limit of the curves shown is the SNR of that measurement, assuming an instrumental across-the-board efficiency of 10%. The hatched area indicates the desired regime of spectral and temporal resolution. If simultaneity is an absolute must, the diameter of a filled aperture must be increased to a minimum of 32 m in order to collect enough photons for a 1 s exposure with the desired SNR within a 10 mÅ spectral band. In this case, the extra 1.2 MW solar power outside the spectral channels of interest pose a particularly daunting problem.

It is quite clear that a trade-off between the scientific goals will have to be made. A realistic estimate of instrument efficiency would be 10%, divided equally between the efficiency of a diluted pupil and optical throughput. A simultaneous achievement of scientific goals is then missed by a factor of 10^2-10^3. Relaxing the requirements of the angular resolution and SNR results in the optimum solution; descoping either of them by a factor of 10 increases spectral and temporal resolution by a factor of 100. For example, integrated intensity measurements over a 0.3x0.3 arcsec area can be made within spectral bands of 10 mÅ , 0.1 s, and an SNR of 10^4. Similarly, a 0.1 x 0.1 arcsec area within 1 s, again with a spectral resolution of 10 mA and an SNR of 10^4 could be observed. Other combinations would be measurements within 1 s, an SNR of 10^3, and highest spectral and angular resolutions. All of the other

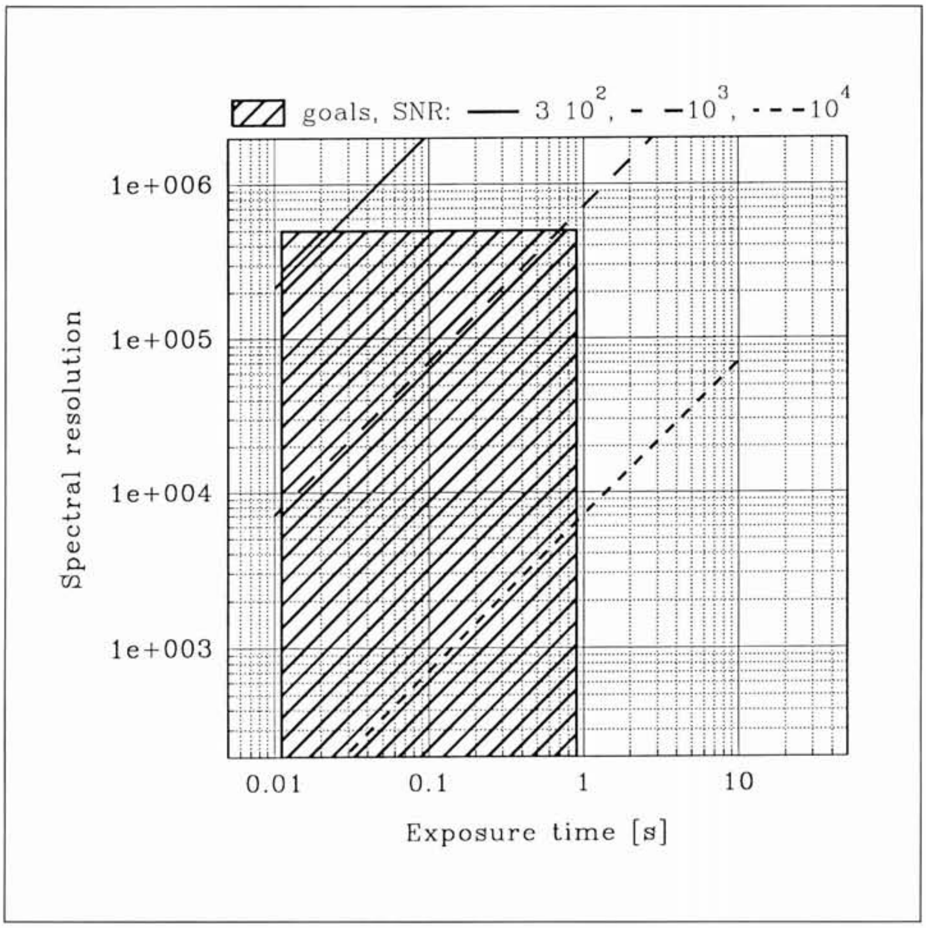


Figure A3.1 Spectral and temporal resolution relationship as a function of signal-to-noise, at λ = 500 nm

scientific goals, apart from SNR, can be achieved with an SNR of 300. It is important to be aware that at an angular resolution of 0.01 arcsec, the very high signal-to-noise goals may not be necessary if magnetic fields are very concentrated, and therefore very strong. It should be noted that sensitivity is also a function of angular frequency, which is particularly important when the pupil of the telescope deviates substantially from a filled circle, and the MTF becomes relatively small at medium to high frequencies. Such cases are presented in the next section. The primary noise source in the data will be that of the photons, and the noise spectrum will be 'white', i.e. independent of angular frequency, and dependent only on the number of detected photons as long as the observed structure has low contrast. The signal-to-noise ratio in the Fourier domain is then proportional to the MTF. Low MTF values at higher frequencies imply that measurements of small-scale structures have greater noise; this implication needs to be accounted for when trade-offs are made.

3. Elements of an advanced lunar solar observatory

3.1 General considerations

A lunar solar observatory will have many aspects in common with existing solar instruments. The basic components include:

- A structure which supports the main optical elements, the focal-plane assembly, and the post-focal-plane assembly in a stable fashion, ensures their alignment, and provides pointing towards the region of interest on the Sun.
- The main optical elements which collect sunlight and direct it towards the instrument focus where a solar image is formed. These elements may consist of a single primary telescope mirror only, or an array of smaller collecting elements.
- Possible focal-plane assembly that assures that the light received from the main optical elements forms a usable solar image.
- The post-focus instrument assembly which contains spectrometers, filters, polarimetric equipment, detectors, etc. This unit achieves most of the scientific goals.

- A possible control system which helps the structure align the optical components to the required precision, thereby relaxing the requirements on the stability of the structure itself. This will probably include an active closed-loop system similar to present active optics.

The entire observatory could be one compact system pointed exclusively at the Sun and could include all post-focus instrumentation. Another option would be to split the system into two parts, one of which combines the light-collecting mechanism pointed at the Sun, the other the focal-plane and post-focal-plane assemblies at a fixed location above or below ground, to which the sunlight is directed in a Coudé optical arrangement. Placing the post-focal equipment below ground may provide radiation protection both for sensitive equipment and humans. This option is shown schematically in Figure A3.2 which represents a 12-element interferometric-array telescope on a tower structure 15 m high. The light is collected by the 12 smaller telescopes on the perimeter and directed towards the secondary-optics unit in the centre of the main structure. Part of the sunlight is directed through the central vertical pipe towards an underground facility. The support building and the astronauts on the ground provide an impression of the scale.

3.2 Structure

The support structure provides the following:

- Support to the major elements of the telescope to ensure their alignment. Alignment requirements will be of the order of the observed radiation wavelength and can be relaxed if an active control system is used. The structure itself may play a part in

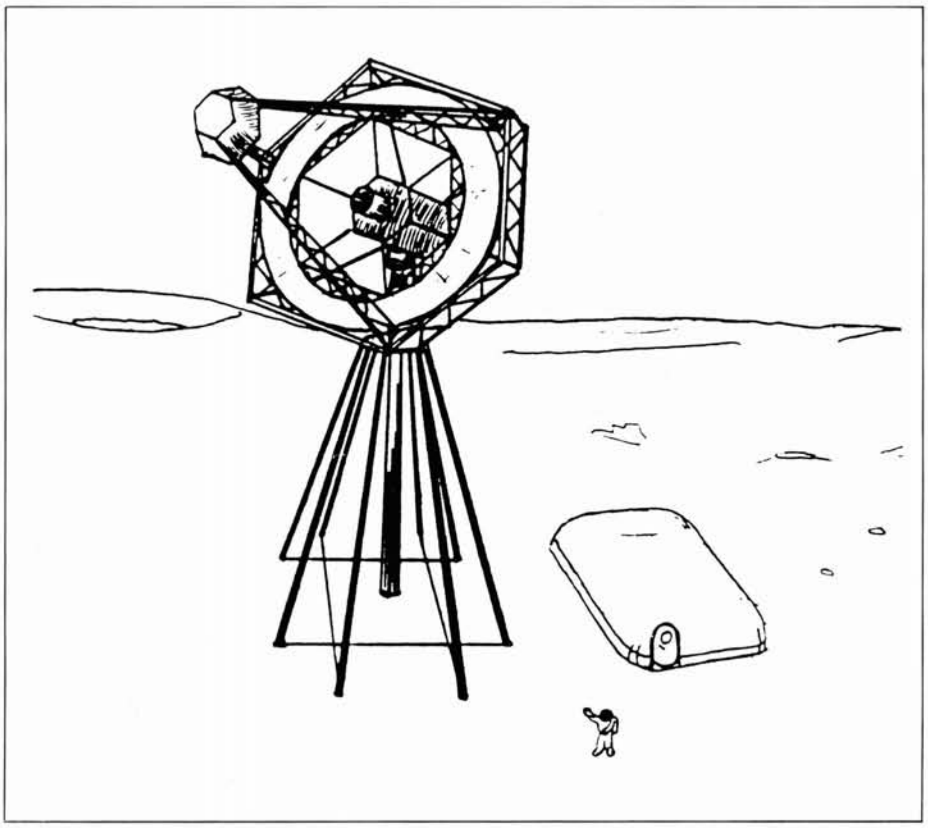


Figure A3.2. Artist's impression of a possible lunar solar observatory

alignment control by implementing active members. The structure will also accommodate light-feeds to a possible underground laboratory.
- Pointing to and tracking of the Sun.
- Shielding from environmental hazards (lunar surface material contamination, micrometeorites, etc.) and support of a charged-particle deflection system to protect the main optics from direct exposure to solar wind.
- Elevation of the telescope to a height that minimises contamination with lunar surface material.

The design of the structure is closely linked to the design of the main optics. It will probably consist of prefabricated lightweight elements and pre-assembled substructures, which will undergo final assembly on site.

3.3 Main optics

The main optical elements constitute the 'telescope' of the observatory. They provide the baseline which achieves angular resolution in line with the scientific requirements, and they collect a sufficient amount of photons. At the same time, they help in limiting the solar flux to manageable levels. The latter two are conflicting functional requirements. As pointed out earlier, all scientific goals cannot be met simultaneously with a reasonable effort. The need to limit solar flux at the secondary optics unit greatly complicates matters.

One option is to select a few spectral bands of interest and to treat the surfaces of the main optics such that only these are directed towards the instrument focus. The main optics can then consist of a filled aperture of approximately 10 m diameter, whereby the photon flux within the windows is maximised. The great disadvantage is that the scientific versatility of the resulting observatory would then be severely limited. The unused light should be transmitted by the primary optics unit, rather than being absorbed.

Another option is to reduce the collecting area of the main optics. Full spectral coverage can now be retained at the expense of an across-the-board loss in photons. Such a configuration is referred to as the 'diluted pupil' approach and, because it is scientifically more attractive it will be investigated here in some detail. The following general requirements will be imposed on the main optics:
- The maximum baseline should result in an angular resolution of 100 line-pairs per arcsec at a wavelength of 500 nm and is therefore taken to be 10 m.
- The pupil configuration should be such that instantaneous coverage of angular frequencies in the Fourier transform of image intensity is provided within the resolution limit.
- The configuration should exhibit rotational symmetry in order to minimise instrument polarisation.

The choices considered here are annular-aperture telescopes and interferometric telescope arrays.

■ Annular-aperture telescopes

An annular-aperture telescope can be regarded as an ordinary telescope with an exceptionally large central circular obstruction. The collecting area consists of a ring with an outer diameter D of 10 m. The ratio ϵ of inner and outer diameters determines the reduction of the collecting area and hence the photon flux. The collecting area A is given by:

$$A = A_o (1 - \epsilon^2) = \frac{\pi}{4} D^2 (1 - \epsilon^2) \quad (2)$$

where $A_o = 78.5\ m^2$ is the collecting area of a filled 10 m aperture.

The imaging performance of an annular-aperture telescope can be judged by its Modulation Transfer Function (MTF). Figure A3.3 shows the MTF for various values of ϵ. There is a rapid decrease at low angular frequencies, followed by a more or less constant plateau that extends to the maximum resolution achievable, at 100 *lp*/arcsec. The sensitivity with which Fourier components in the solar image can be measured at these angular frequencies is evidently reduced in comparison with that of a filled aperture. However, the MTF of the annular telescope up to an angular frequency of (1-ϵ) times the diffraction limit of 100 *lp*/arcsec is comparable with that of a filled aperture with a diameter of (1-ϵ) times 10 m, with a much larger sensitivity because the collecting area of the annular telescope is larger by a factor of $(1+\epsilon)/(1-\epsilon)$.

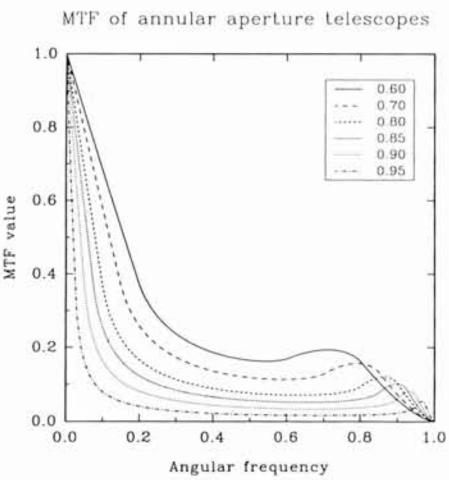


Figure A3.3 Annular aperture MTF, with ϵ as a curve parameter

The point spread function produced by this type of telescope resembles the familiar Airy discs with increased intensities in the higher order rings as ϵ increases. Therefore, images taken with such a telescope should be deconvolved (corrected for the MTF) in order to arrive at meaningful photometric results. The MTF can be measured by frequent observation of bright stars (blue giants).

Table A3.3 summarises some properties of annular aperture telescopes. It is evident that a significant reduction in collecting area (by more than 30%) is achieved only for $\epsilon > 0.8$. The main optics can consist of segments of the annulus assembled into the structure on site. Each segment will have the shape of an off-axis parabola or hyperboloid. Since all segments are identical, they can probably be fabricated by replication.

Table A3.3 Annular aperture telescope properties

ϵ	A (m^2)	A/A_o	MTF (at 50 lp/arcsec)
0	78.5	1.0	0.43 (filled aperture)
0.6	50.3	0.64	0.17
0.7	40.1	0.51	0.12
0.8	28.3	0.36	0.08
0.85	21.8	0.28	0.06
0.9	14.9	0.19	0.04
0.95	7.7	0.098	0.02

■ Interferometric telescope arrays

The main optics of an interferometer consist of a number of disconnected collectors directing sunlight to the final focus. The optical path lengths to all collectors, from the source to the final focus, must be equal for the interferometer to operate properly. A significant advantage of an interferometric array with many elements is its modularity; just a few (3 or 4) array elements could be incorporated during an initial phase with reduced performance in resolution and sensitivity while the system is tested and initial scientific programmes are performed. Further elements can be added later, reducing development risks and flattening the funding profile.

The most efficient configuration for the collectors is on the perimeter of a circle; this leads to nearly uniform coverage of the UV plane within the desired resolution limit. Also, the configuration must be extremely symmetric to minimise instrument polarisation, which leaves only the regular multigons if all elements have the same diameter. From the front, such a configuration looks like a segmented, annular-aperture telescope with circular and disconnected segments. The minimum diameter

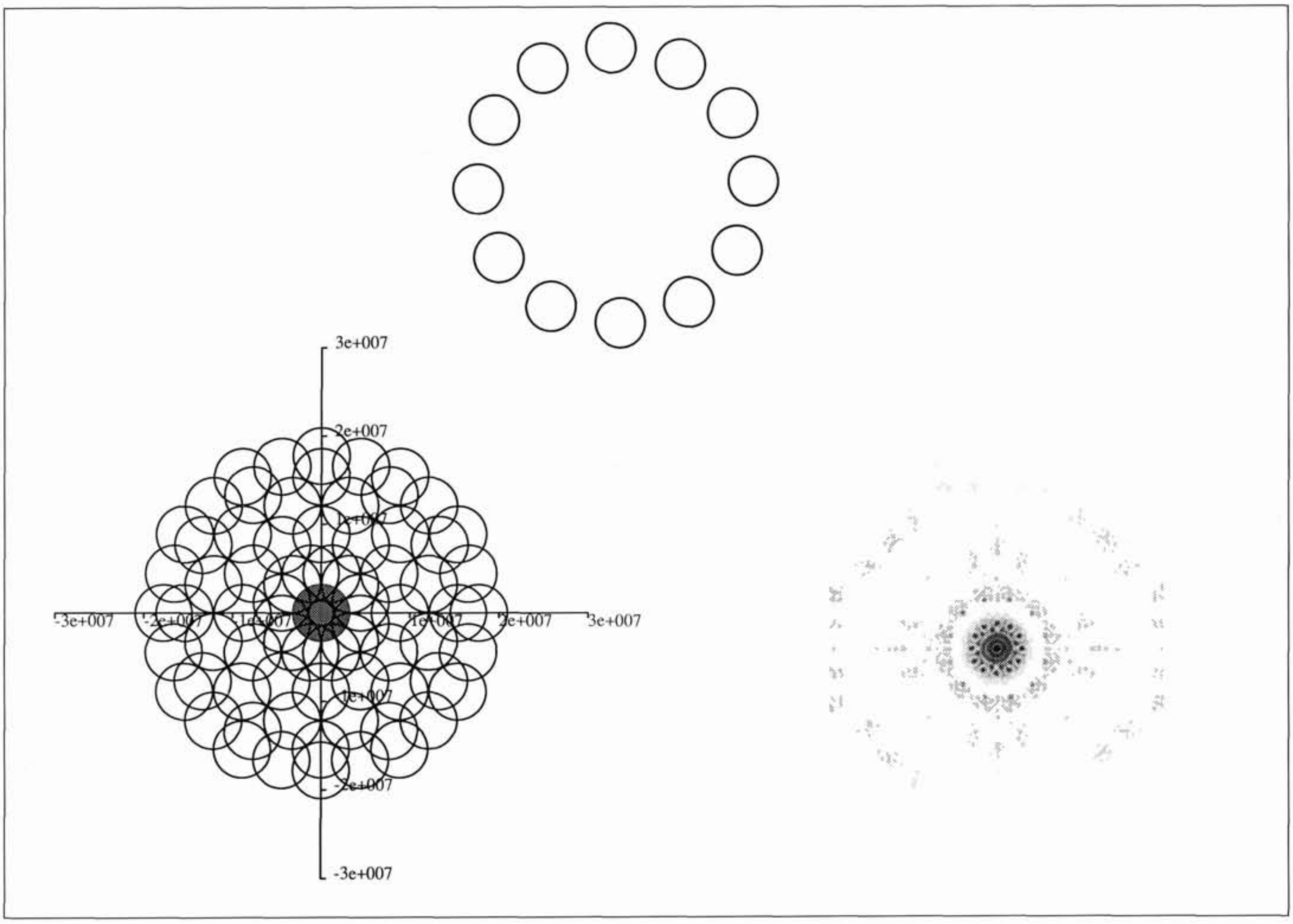


Figure A3.4 Two-dimensional renderings of the configuration (top-left), UV plane coverage (top-right) and PSF (bottom) of the 12-element array telescope shown in Figure A3.2. The configuration consists of twelve 1.6 m diameter telescopes arranged regularly over a 10 m diameter circle. The UV plane coverage is shown in units of 0.5 μm wavelengths. The PSF is shown in a logarithmic intensity scale in order to emphasise the background halo

D of a single element of the interferometer and the spacing *d* of its nearest neighbours then follows directly from the number of elements, the 10 m maximum baseline, and complete UV coverage requirements. Table A3.4 summarises parameters for some arrays for up to 12 elements. Figure A3.4 shows the configuration, the coverage of the UV plane and the point spread function for the array with the 12 elements presented in Table A3.4.

The imaging properties of an interferometer can again be judged by inspection of its MTF. For an array with *N* elements, the MTF shows a rapid decrease at low angular frequencies, followed by a complex pattern at intermediate frequencies with a mean level of 1/N. The MTF has a mean value of $1/(2N)$, close to maximum angular resolution. The two-dimensional coverage shows a strong contribution near the frequency origin which extends to an angular frequency corresponding to the resolution limit of a single interferometer element, D/λ, and a pattern of partially overlapping circular patches of the same size as the central region, but at a much lower level. Figure A3.5 shows representative cross sections of the MTF and the PSF of the 12-element array shown in Figure A3.4. Table A3.5 summarises MTF values at approximately half the maximum resolution.

The point spread functions of the interferometer arrangements consist of the Airy disc of a single interferometer element (with diameter *D*) with an imposed complex fringe pattern. The high-resolution information resides in these fringes. The PSF includes a bright sharp core which represents the equivalent diffraction-limited point spread function of a 10 m filled-aperture telescope if the interferometer is aligned with an accuracy better than a small fraction of the mean observing wavelength. The more elements the array contains, the more its point spread function will resemble that of an annular aperture with an annulus width *D*. Image data taken with an array need to

Table A3.4 Solar interferometric array configuration parameters

N	d (m)	D (m)	A (m^2)	A/A_o
3	6.4	3.6	30.3	0.39
4	4.7	3.33	34.9	0.44
6	3.9	2.18	22.5	0.29
8	3.0	2.13	28.5	0.36
12	2.3	1.60	24.1	0.31

Figure A3.5 Representative cross-sections of the MTF (left) and the PSF (right) of the 12-element array

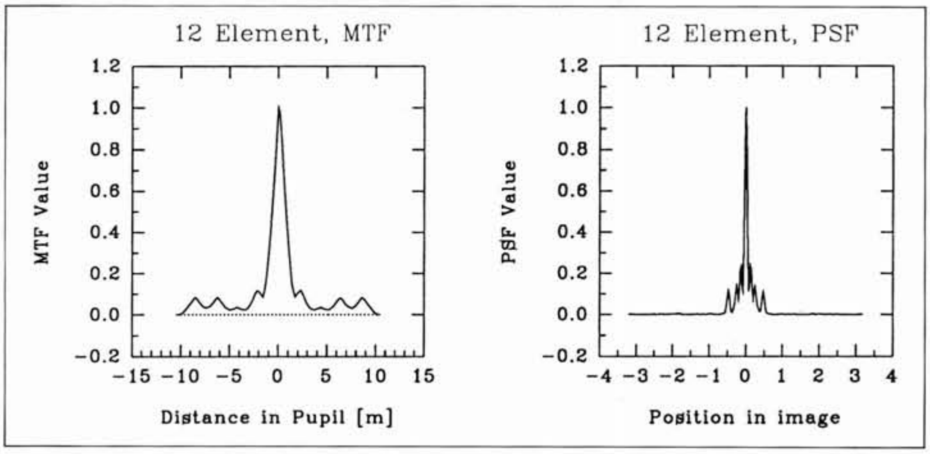


Table A3.5 Solar interferometric array MTF levels

N	Mean MTF at 50 lp/arcsec
3	0.167
4	0.250
6	0.167
8	0.125
12	0.083

be calibrated for the MTF. As with the annular-aperture telescope, the MTF can be measured by frequent observation of bright stars.

Two principal concepts are involved in constructing an imaging interferometer. The Fizeau concept looks very much like an incomplete annular telescope. All interferometer elements are identical, and form a common parabolic or hyperbolic surface. They need to be adjusted relative to each other with a precision of a fraction of a wavelength. Light from the elements is directed towards a secondary optics assembly which forms a solar image at the final focal plane. A solar observatory based on a Fizeau interferometer concept would look very much like the annular-aperture telescope, except that the annulus is replaced by an array of circular elements, as shown in Figure A3.4. The advantages of this concept are simple optics, efficiency, and an inherently large field of view. Among the disadvantages are the need for the secondary optics structure to protrude from the primary optics structure in the direction of the Sun, and very tight adjustment tolerances, which should be controlled with an active optics system.

The other concept is the Michelson design. Each collector consists of a small, afocal telescope that directs a compressed beam of sunlight towards a central beam-combiner station. The light is transformed at this point into superimposing solar images, one from each interferometer element. If the optical paths of all elements are equal, detectable interference fringes are produced in the superimposed image. The beam-combiner station can have integrated path-length control systems ('delay lines') which allow substantial relaxation of the stability requirements of the whole structure. Precise mapping of the output pupil to the input pupil ('homothetic mapping') is required in order to maintain a useful field of view.

The cross-section through one interferometer arm is shown Figure A3.6. It comprises a small modified Gregorian telescope, using three imaging mirrors which constitute one of the array elements. An advantage of the Gregorian telescope is that there is a real primary focus where a field stop can be placed. This stop rejects all light except that from the field of interest, which could be as small as a few arcminutes (a field of 2x2 arcmin requires detectors with 24 000x24 000, or 576 000 000, pixels for critical sampling), thereby removing a very large fraction of the solar flux. The solar-flux problem is therefore basically overcome. The same primary focus image of the Sun may be used for a limb guider system that assures correct pointing of the individual elements. An image of the full solar disc for scientific purposes is best generated at lower angular resolution by using a separate auxiliary telescope.

An afocal beam of sunlight is generated by the Gregorian tertiary mirror and sent via a flat mirror towards the central beam-combiner. The three-mirror Gregorian design ensures accurate orientation of the afocal beams at the beam combiner in conjunction with the other transfer optics. The beam-combiner itself consists of a concave mirror

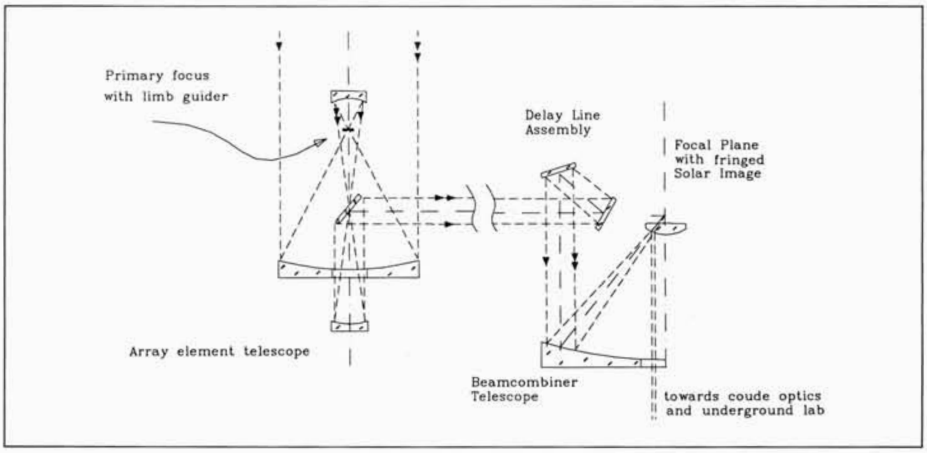


Figure A3.6 Solar Michelson interferometer optical design

that collects the beams coming from all elements, and which then generates a focal plane with a usable solar image. It is essential to arrange all beams entering the beam-combiner precisely such that the diameters of the parallel beams and their position form a scaled image of the entrance pupil of the entire interferometer, and to ensure that the orientation of the pupils be maintained. This guarantees interference fringes within the entire observed field, and eliminates polarisation reduction of fringe contrast.

The structure that directs the converging light beams towards the beam-combiner telescope consists of a rigid assembly of two flat mirrors arranged as a pentaprism. The assembly can be moved along its plane of symmetry by a few centimetres, thereby changing the length of the light path from an element to the beam-combiner without changing the position and direction of the exit beam optical axis. There is one such assembly for each array element. The structure will be used as a delay line, allowing equalisation of all optical paths.

Among the numerous advantages of the Michelson concept are:
- compact main structure, because no protruding secondary optics are needed
- relaxed stiffness requirements for the main structure, because of the presence of delay lines with a stroke of a few centimetres
- simple solution for the solar-flux problem without impaction sensitivity
- modularity (array element designs could also be used for other purposes).

One disadvantage is far more complex optics, partially offset by the fact that many of the control elements, which would probably be necessary to keep a Fizeau interferometer aligned, are already part of the Michelson concept.

3.4 Focal-plane assembly

Much of a focal-plane assembly is actually part of the main optics and therefore highly concept-dependent. This part of the observatory is of significant size only when the Michelson interferometer is used and would consist of the beam-combiner. Otherwise, the focal-plane assembly would include all functions essential for the proper operation of the telescope, but not directly contributing to the scientific output, e.g. pointing and guiding systems, pick-up optics for control and alignment systems, optics for the calibration of the telescope on bright stars, etc.

3.5 Post-focus instrumentation assembly

This assembly will contain the instruments that produce scientific data, and will implement the primary scientific goals of spectral resolution, temporal resolution, and polarimetric resolution. It will probably consist of filter graphs, spectrographs, polarimetric equipment, detectors and data compression and reduction facilities. This area is prone to rapid evolution, in particular in solar physics, and it does not seem worthwhile at this stage to make predictions as to what will be available thirty years from now.

One point worth mentioning is the desire for long, uninterrupted observation of the Sun. Owing to the immense amount of data present even in single snapshots, the observatory must clearly provide the means for sustained analysis and/or transfer of large quantities of data.

4. Conclusions

Major features of the lunar-base solar observatory are determined by the objective of high angular resolution, which sets the dimensions of the telescope aperture at approximately 10 m diameter. The discussion in the second section has revealed that fundamental limits prevent the simultaneous achievement of all scientific goals, even with concerted efforts. On the one hand, a large number of photons are needed within very narrow spectral bands for very short time intervals, and this requires a large collecting power (more than that needed for resolution) and highly efficient optics. On the other hand, the problems of dealing with the collected flux throughout the solar spectrum seems insurmountable if a filled aperture is used. One therefore has to make trade-offs between scientific goals based on the specific research, and use of a telescope with an efficiency that allows proper management of the solar flux.

Of the concepts for an advanced solar observatory on the Moon discussed in the third section, the most promising appears to be a Michelson interferometer array with a relatively large number of elements. The 12-element option shown has element diameters of manageable size (1.6 m), which can be designed as a Gregorian telescope. This immediatetely solves the solar-flux problem by means of a field stop in the primary focal plane. The entire structure will be compact, and charged-particle deflection systems can be mounted in front of each element relatively easily. The array could initially consist of only a few elements and be upgraded later.

The same structure that supports the telescope for the visible spectral regime could also carry a smaller telescope for shorter wavelengths. A scaled-down interferometer or annular-aperture telescope optimised for the near- and vacuum-UV spectral range down to $L\alpha$ (1000 Å) could be implemented within the central 2 m of the structure. Stiffness requirements will be more stringent because of the shorter wavelength, but the problem will be offset by that fact that the UV instrument would be a fifth of the size of the visible telescope.

It is worthwhile noting that in some areas addressed herein there is quite considerable expertise among European solar physicists. One project worth mentioning is SIMURIS, which incorporates a four-element linear interferometer with a 2 m baseline, to fly on the Space Station. This interferometer is designed for use in the UV and visible range. A lot of valuable experience can be gained in the pursuit of this project, which can have direct application to a lunar-based solar observatory.

A specific area of technology development relevant to solar observations (apart from the general technological issues of deployment and support of a major scientific facility on the Moon) is the development of suitable panoramic detectors. The readout speed, as well as the capacity for 'information units' (photon counts, photo-electrons, etc.), needs substantial improvement. Of course, the technological issues of building a 10 m interferometer for visible wavelengths, or 2 m for the UV, are quite difficult, but this will also be the case for similar stellar observatories.

Bibliography

Bibliography

■ Science of the Moon

Basaltic Volcanism Study Project, *Basaltic Volcanism on the Terrestrial Planets*, Pergamon Press, New York, 1981.

Bratt S R, Solomon S C, Head J W & Thurber C H 1985, The Deep Structure of Lunar Basins: Implications for Basin Formation and Modification, *J. Geophys. Res.*, No. 90, pp. 3049—3064.

Bills B G & Ferrari A J 1977, A Lunar Density Model Consistent with Topographic, Gravitational, Librational and Seismic Data, *J. Geophys. Res.*, No. 82, pp. 1306—1314.

Boss A P & Peale S J 1986, Dynamical Constraints on the Origin of the Moon, *Origin of the Moon*, Hartmann W K, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 59—101.

Cameron A G W 1986, The Impact Theory for the Origin of the Moon, *Origin of the Moon*, Hartmann W K, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 609—616.

Chevrel S & Pinet P 1990, Near-Infrared Multispectral Images for Gassendi Crater: Mineralogical and Geological Inferences, *Proc. 20th Lunar Planet Sci. Conf.* , Lunar and Planetary Institute, Houston, pp. 187—194.

Clayton R N & Thiemens M H 1980, Lunar Nitrogen: Evidence for Secular Change in the Solar Wind, *Proc. Conf. on the Ancient Sun, Geochim. Cosmochim. Acta*, Suppl. 13, Pergamon Press, New York, pp. 463—473.

Fisk L A 1980, Solar Modulation of Galactic Cosmic Rays, *Proc. Conf. on the Ancient Sun, Geochim. Cosmochim. Acta*, Suppl. 13, Pergamon Press, New York, pp. 103—118.

Geiss J & Bochsler P 1991, Long-Term Variations in Solar-Wind Properties: Possible Causes versus Observations. In: The Sun in Time, Univ. of Arizona Press, Tuscon, pp. 98 – 117.

Goins N R 1978, The Internal Structure of the Moon, PhD Thesis, Massachussetts Institute of Technology, 666 p.

Haines E L & Metzger A E 1979, The Variation of Iron Concentration in the Lunar Highlands and Resultant Implications for Crustal Models, *Lunar Planet Sci. X*, Lunar and Planetary Institute, Houston, pp. 689—718.

Hartmann W K 1986, The Impact-Trigger Hypothesis, *Origin of the Moon*, Phillips J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 579—608.

Hartmann W K, Phillips R J & Taylor G J (Eds.) 1986, Origin of the Moon, Lunar and Planetary Institute, Houston, 781 p.

Head J W 1975, Processes of Lunar Crater Degradation: Changes in Style with Geologic Time, *The Moon*, No. 12, pp. 299—329.

Hood L L 1986, Geophysical Constraints on the Lunar Interior, *Origin of the Moon*, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 361—410.

Hood L L & Jones J 1985, Lunar Density Models Consistent with Mantle Seismic Velocities and Other Geophysical Constraints, *Lunar Planet Sci. XVI*, Lunar and Planetary Institute, Houston, pp. 360—361.

Jones J H & Delano J W 1989, A Three-Component Model for the Bulk Composition of the Moon, *Geochim. Cosmochim. Acta*, No. 53, pp. 513—527.

Kaula W M 1977, On the Origin of the Moon, with Emphasis on Bulk Composition, *Proc. 8th Lunar Sci. Conf.* , Pergamon Press, New York, pp. 321—331.

Kaula W M, Schubert G, Lingenfelter R E, Sjogren W L & Wollenhaupt W R 1972, Analysis and Interpretation of Lunar Laser Altimetry, *Proc. 3rd Lunar Sci. Conf.*, Pergamon Press, New York, pp. 2189—2204.

Kaula W M, Schubert G, Lingenfelter R E, Sjogren W L & Wollenhaupt W R 1974, Apollo Laser Altimetry and Inferences as to Lunar Structure, *Proc. 5th Lunar Sci. Conf.*, Pergamon Press, New York, pp. 3049—3058.

Langevin Y & Arnold J R 1977, The Evolution of the Lunar Regolith, *Annual Review of Earth and Planet. Sci.*, No. 5, p. 449.

Lucey P G & Hawke B R 1987, A Remote Mineralogic Perspective on Gabbroic Units in the Lunar Highlands, *Proc. 18th Lunar Planet Sci. Conf.* , Cambridge Univ. Press, pp. 355—363.

Mendell W 1985, *Lunar Bases and Other Space Activities of the 21 st Century*, Lunar and Planetary Institute, Houston, 865 p.

Nakamura Y 1983, Seismic Velocity Structure of the Lunar Mantle, *J. Geophys. Res.*, No. 88, pp. 677—686.

Nakamura Y, Latham D, Lammlein D, Ewing M, Duennebier F & Dorman J 1974, Deep Lunar Interior Inferred from Recent Seismic Data, *Geophys. Res. Lett.*, No. 1, pp. 137—140.

Newsom H E, The Lunar Core and the Origin of the Moon, EOS Trans., Amer. Geophys Union, No. 65, pp. 369—370.

Pepin R O, Eddy J A & Merrill R B (Eds.) 1980, *Proc. Conf. Ancient Sun: Fossil Record in the Earth, Moon and Meteorites, Geochim. Cosmochim. Acta*, Suppl. 13, Pergamon Press, New York, 581 p.

Pieters C M 1986, Composition of the Lunar Highland Crust from Near-Infrared Spectroscopy, *J. Geophys. Res.*, Vol. 24, No. 3, pp. 578.

Ringwood A E 1979, *Origin of the Earth and Moon*, Springer-Verlag, New York, 295 p.

Ringwood A E 1986, Composition and Origin of the Moon, *Origin of the Moon*, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 673—698.

Ringwood A E, Seifert S & Wänke H 1986—87, A Komatiite Component in Apollo-16 Highland Breccias: Implications for the Nickel-Cobalt Systematics and Bulk Composition of the Moon, *Earth Planet Sci. Lett.*, No. 81, pp. 105—117.

Ryder G & Wood J A 1977, Serenitatis and Imbrium Impact Melts: Implications for Large-Scale Layering in the Lunar Crust, *Proc. 8th Lunar Sci. Conf.* , Pergamon Press, New York, pp. 655—668.

Schubert G, Solomon S C, Turcotte D L, Drake M & Sleep N 1991, Origin and Thermal Evolution of Mars, *Mars*, (Kieffer H, Jakowsky B, Snyder C & Matthews M (Eds.), Univ.of Arizona Press, in press.

Singer S F 1986, Origin of the Moon by Capture, *Origin of the Moon*, Phillips J & Taylor J G (Eds.), Lunar and Planetary Institute, Houston, pp. 471—486.

Soderblom L A & Lebofsky L 1972, Technique for Rapid Determination of Relative Ages of Lunar Areas from Orbital Photography, *J. Geophys. Res.*, No. 77, pp. 279—296.

Solomon S C & Head J W 1979, Vertical Movement in Mare Basins: Relation to Mare Emplacement Basin Tectonics and Lunar Thermal History, *J. Geophys. Res.*, No. 84, pp. 1667—1682.

Solomon S C & Head J W 1980, Lunar Mascon Basins: Lava Filling Tectonics and Evolution of the Lithosphere, *Rev. Geophys. Space Phys.*, No. 18, pp. 7—14.

Spohn T 1991, Mantle Differentiation and Thermal Evolution of Mars, Mercury, and Venus, *Icarus*, No. 90, pp. 222—236.

Spudis P D, Hawke B R, Hood L L, Schultz P H, Taylor G J & Wilhelms D E 1986, Status and Future of Lunar Geoscience, NASA SP-484, 54 p.

Stöffler D, Knoll H D, Marvin U B, Simonds C H & Warren P H 1980, Recommended Classification and Nomenclature of Lunar Highland Rocks: A Committee Report, *Proc. Conf. on Lunar Highland Crust*, Papike & Merrill (Eds.), pp. 51—70.

Symposium on Lunar Bases and Space Activities of the 21st Century, April 5—7, 1988 Houston, Texas, *Abstracts*, Lunar and Planetary Institute, Houston, pp. 260.

Taylor S R 1982, *Planetary Science: A Lunar Perspective*, Lunar and Planetary Institute, Houston, 481 p.

Tera F, Papanastassiou D A & Wasserburg G J 1974, Isotopic Evidence for a Terminal Lunar Cataclysm, *Earth Planet Sci. Lett.*, No. 22, pp. 1—21.

Walker R M 1980, Nature of the Fossil Evidence: Moon and Meteorites, *Proc. Conf. on the Ancient Sun, Geochim. Cosmochim. Acta*, Suppl. 13, Pergamon Press, New York, pp. 11—28.

Wänke H & Dreibus H 1982, Chemical and Isotopic Evidence for the Early History of the Earth–Moon System, *Tidal Friction and the Earth's Rotation II*, Brosche P & Sündermann J (Eds.), Springer Verlag, Berlin, pp. 322—344.

Wänke H & Dreibus G 1986, Geochemical Evidence for the Formation of the Moon by Impact-Induced Fission of the Proto-Earth, *Origin of the Moon*, Hartmann W K, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 649—672.

Wänke H, Dreibus G & Jagoutz E 1984, Mantle Chemistry and Accretion History of the Earth, *Arch. Geochemistry*, Kröner A et al. (Eds.), Springer Verlag, Berlin, pp. 1—24.

Weidenschilling S J et al. 1986 Origin of the Moon from a Circumterrestrial Disk, *Origin of the Moon*, Hartmann W K, Phillips R & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 731—762.

Wilhelms D E 1987, The Geologic History of the Moon, US Geological Survey Professional Paper No. 1348, US Government Printing Office, Washington DC, 302 p.

Wood J A 1986, Moon Over Mauna Loa: a Review of Hypothesis of Formation of the Earth's Moon, *Origin of the Moon*, Hartmann W K, Phillips R J & Taylor G J (Eds.), Lunar and Planetary Institute, Houston, pp. 17—55.

Workshop on Mare Volcanism and Basalt Petrogenesis: Astounding Fundamental Concepts Developed Over the Last Fifteen Years, *Abstracts*, 1990, Lunar and Planetary Institute, Houston, 58 p

Yoder C F 1981, The Free Librations of a Dissipative Moon, *Phil. Trans. Roy. Soc. London*, A-303, pp. 327—338.

Young R A 1975, Mare Crater Size-Frequency Distributions: Implications for Relative Surface Ages and Regolith Development, *Proc. 6th Lunar Sci. Conf.*, Pergamon Press, New York, pp. 2645—2662.

Young R A 1977, The Lunar Impact Flux, Radiometric Age Correlation, and Dating of Specific Lunar Features, *Proc. 8th Lunar Sci. Conf.*, Pergamon Press, New York, pp. 3457—3473.

■ Astronomy and Astrophysics

Baldwin J E 1990, *Proc. Int. Workshop on Low-Frequency Astrophysics from Space (LFAfS)*, Crystal City (Va), USA, 8–9 January 1990, p. 195.

Barrow C H 1985, The Influence of the Sun on Jupiter's Radio Emission. *Planetary Radio Emissions I*, Rucker H O & Bauer S J (Eds.), Austria.

Benson R F & Akasofu S I 1984, Auroral Kilometric Radiation / Aurora Correlation, *Radio Sci.*, No. 19, p. 527.

Boischot A et al. 1980, A New High Gain, Broadband, Steerable Array to Study the Jovian Decametric Emission, *Icarus*, No. 43, p. 399.

Chevalier R A 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 130.

Connerney J E P, Davis L Chenette D L 1984, Magnetic Field Models, *Saturn*, Gehrels T & Matthews M S (Eds.), Univ. of Arizona, Tucson.

Cordes J M 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 163.

Dulk G A 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 85.

Duric N, A Lunar Far Side Very Low Frequency Array, NASDA Conference Publ. No. 3039, p. 29

Galopeau P, Zarka P & LeQuéau D 1989, Theoretical Determination of Saturn's Kilometric Radiation Spectrum, *J. Geophys. Res.*, No. 94, p. 8739.

Galopeau P, Ortega-Molina A & Zarka P 1991, Evidence of Saturn's Magnetic Field Anomaly from Saturnian Kilometric Radiation High Frequency Limit, *J. Geophys. Res.*, in press.

Hajivassiliou C A, in preparation, 1991.

Hoyle F & Ellis G R A 1963, *S. Australian J. Phys.*, No. 16, p. 1.

Kaiser M L & Desch M D 1984, Radio Emissions from the Planets Earth, Jupiter and Saturn, *Rev. Geophys. Space Phys.*, No. 22, p. 373.

Kaiser M L et al. 1984, Saturn as a Radio Source, *Saturn*, Gehrels T & Matthews M S (Eds.), Univ. of Arizona, Tucson.

Kassim N E 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 144.

Lecacheux A 1990, *Proc. 3rd Int. Bioastronomy Symp.*, Val-Cenis, 18–23 June 1990.

LeQuéau D 1988, Planetary Radio Emissions from High Magnetic Latitudes: the 'Cyclotron Maser' Theory, *Planetary Radio Emissions II*, Rucker H O, Bauer S J & Pedersen B M (Eds.), Austria.

Okoye S E & Hewish A 1967, *Mon. Not. Roy. Astron. Soc.*, pp. 137, 287.

Razin R G 1960, *Radiophysica*, No. 3, p. 584.

Readhead A C S, Kemp M C & Hewish A 1978, *Mon. Not. Roy. Astron. Soc.*, No. 185, p. 207.

Reber G 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990.

Rees N 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 204.

Rees N 1991, *Mon. Not. Roy. Astron. Soc.*, pp. 233—244.

Reynolds R J 1990, *Proc. LFAfS*, Crystal City (Va), USA, 8–9 January 1990, p. 121.

Spangler S R & Armstrong J W 1990, Low Frequency Astrophysics from Space, Springer Verlag, Berlin.

Steinberg J L et al 1984, *Astron. Astrophys.*, No. 140, p. 39.

Stone E C 1977, The Voyager Missions to the Outer Solar System, *Space Sci. Rev.*, No. 21, p. 75.

Tsytovich V N 1951, Vestn. Mosk. Univ., No. 11, p. 27.

Winglee R M, Dulk G A & Bastian T S 1986, A Search for Cyclotron Maser Radiation from Substellar and Planet-Like Companions of Nearby Stars, *Astrophys. J. (Lett.),* No. 308, L59—L62.

Wu C S & Lee L C 1979, A Theory of the Terrestrial Kilometric Radiation, *Astrophys. J.*, No. 230, p. 621.

Zarka P & Pedersen B M 1986, Radio Detection of Uranian Lightning by Voyager-2, *Nature*, No. 323, p. 605.

Zarka P 1991, The Auroral Radio Emissions from Planetary Magnetospheres, in preparation.

■ Physics of the Sun

Space Science – Horizon 2000, ESA SP-1070, 1984

A Proposed Medium-Term Strategy for Optical Interferometry in Space, ESA SP-1135, 1990.

Bochsler P, Geiss J & Maeder A 1990, The Abundance of ^{3}He in the Solar Wind: A Constraint for Models of Solar Evolution, *Solar Physics*, No. 128, pp. 203—215.

■ High Energy Physics

Burnett T et al. 1987, *Proc. 20th ICRC*, Moscow.

Domokos G et al. 1989, *Phys. Rev. Lett.*, No. 63, p. 844.

Domokos G et al. 1990, CERN-TH 5872.

Fichtel C & Linsley J 1985, NASA 85-003.

Grieder P 1989, High Energy Neutrino Astronomy and Astrophysics, CERN Summer School.

Horgan R & Jacob M 1981, Physics and Collider Energy, CERN Summer School 1980, CERN 81—04.

Jacob M 1981, Physics and Collider Energies, Cargese Summer School.

Jacob M 1987, P-P Collider Physics: Present and Prospects, Cargese Summer School.

Jacob M 1990, Physics with Accelerators, EPAC-2, Ed. Frontieres.

Lattes C, Fujimoto Y & Hasagawa S 1980, *Physics Reports*, No. 65, p. 152.

McCusker C 1975, *Physics Reports*, No. 20, p. 229.

Pati J, Salam A & Sreekantan B 1986, *Int. J. Mod. Phys.*, A1, p. 147.

Potter A E & Wilson T L 1989, Conference Proceedings, 202.

Rich J, Lloyd-Owen D & Spiro M 1987, Experimental Particle Physics without Accelerators, *Physics Reports*, No. 151, p. 239.

Ringwald A et al. 1990, DESY, pp. 90—127.

Samorski M & Stamm W 1983, *Ap. J. Lett.*, L17, p. 268.

Schramm D & Truran J 1990, New Physics from Supernova 1987A, *Physics Reports*, No. 189, p. 89.

■ Life Sciences

Allen J P 1990, Historical Overview of the Biosphere 2 Project, *Biological Life Support Systems*, Nelson M & Soffen G (Eds.), Synergetic Press, Oracle, pp. 12—32.

Hayatsu R & Anders E 1981, Organic Compounds in Meteorites and Their Origins, *Topics in Current Chemistry*, No. 99, p. 1.

Horneck G et al. 1989, ERA Follow-On Scientific Study, ESA Contract No. 8116/88/F/ BZ.

Horneck G 1991, Radiobiological Experiments in Space: A Review, Nucl. Tracks Radiat Meas, *Int. J. Radiat. Appl. Instr.*, Part D, in press.

Langevin Y et al. 1987, First Statistical Analysis of 5000 Mass Spectra of Cometary Grains Obtained by PUMA I Ionization Mass Spectrometers in the Compressed Mode, *Astron. Astrophys*, No. 187, p. 761.

Lawless J G & Peterson E 1975, Amino Acids in Carbonaceous Chondrites, *Origins of Life*, No. 6, p. 3.

MacElroy R D, Tremor J, Bubenheim D & Gale J 1989, The CELSS Research Programme: A Brief Review of Recent Activities, *Lunar Base Agriculture: Soils for Plant Growth*, Henninger D & Mings D (Eds.), pp. 165—172.

Maurette M et al. 1987, Characteristics and Mass Distribution of Extraterrestrial Dust from the Greenland Ice Cap, *Nature*, No. 328, p. 699.

Maurette M et al. 1988, L'expedition glace bleue I de collecte de micrometeorites antarctiques de Janvier-Février 1988: Résultats préliminaires et perspectives, *ATP Planétologie,* Besançon, p. 295.

Miller S L 1953, The Production of Amino Acids Under Possible Primitive Earth Conditions, *Science*, No. 117, p. 528.

Mullie F & Reisse F 1987, Organic Matter in Carbonaceous Chondrites, *Topics in Current Chemistry*, No. 139, p. 83.

Neal V et al 1988, Advanced Extravehicular Activity Systems Requirements Definition Study, NAS-17779.

Avions Marcel Dassault 1988, Study of Biological Effects and Radiation Protection for Future European Manned Space Flights, ESA Contractor Report CR(P) 25536988/86/NL/PP(SC).

Sagan C 1972, The Search for Indigenous Lunar Organic Matter, *Space Life Sciences*, No. 3, pp. 484.

Salisbury F B 1991, Bioregenerative Life Support System: Farming on the Moon, *Acta Astronautica*, No. 23, pp. 263—270.

Savage C J, Binot R A & Novara M 1991, A European Facility for CELSS and Habitability Research, *Proc. DARA/CNES Workshop on Artificial Ecosystems*, Marseille, 24—26 October 1990, Maquette Pixeline, Marseille.

Schaefer H J et al. 1972, Nuclear Track Recordings of the Astronaut's Radiation Exposure on the First Lunar Landing Mission Apollo XI, *Rad. Res.* No. 49, pp. 245—271.

Silberberg R et al. 1985, Radiation Transport of Cosmic-Ray Nuclei in Lunar Material and Radiation Doses, *Lunar Bases and Space Activities of the 21st Century,* Mendell W W (Ed.), Lunar and Planetary Institute, Houston.

NAS 9-17779: Advanced Extravehicular Activity Systems Requirements Definition Study *Extravehicular Activity at a Lunar Base,* Sept. 1988.

NCRP Report No. 94, 1987

■ Helium-3 Fusion

NASA 1988, Lunar Helium-3 and Fusion Power, *Proc. Workshop at NASA Lewis Research Center*, Cleveland, Ohio, April 1988, NASA Conf. Publ. No. 10018.

Wittenberg L J, Santarius J F & Kulcincky G L 1986, Lunar Source of ^{3}He for Commercial Fusion Power, *Fusion Technology*, No. 10.

■ Exospheric Contamination

Aronovitz L et al 1968, *J. Geophys. Res.*, No. 73, p. 3231.

Herring J R & Licht A L 1959, *Science*, No. 130, p. 266.

Herring J R & Licht A L 1960, *Space Research,* Kallmann Bijl H (Ed.), No. 1, North-Holland Publ. Co., Amsterdam, p. 1132.

Hinton F L & Taeusch D R 1964, *J. Geophys. Res.*, No. 69, p. 1341.

Hodges R R Jr 1973, *J. Geophys. Res.*, No. 78, p. 8055.

Hodges R R Jr 1980, *J. Geophys. Res.*, No. 85, p. 164.

Hodges R R Jr et al. 1972 *J. Geophys. Res.*, No. 77, p. 4079.

Johnson F S 1971, *Rev. Geophys. Space Phys.*, No. 9, p. 813.

Johnson F S, Caroll J M & Evans D E 1972, *Proc. 3rd Lunar Sci. Conf.*, Pergamon Press, New York, p. 2231.

Kopal Z 1974, *The Moon in the Post-Apollo Era,* Reidel Publ. Co., Dordrecht, Holland, 223 p.

Manka R H & Michel F C 1970, *Science*, No. 169, p. 278.

Morgan T H & Shemansky D E 1991, *J. Geophys. Res.*, No. 96, p. 1351.

Potter A E & Morgan T H 1988, *Science*, No. 241, p. 675.

Shemansky D E, Broadfoot A L & Wallace L 1977, *Rev. Geophys. Space Phys.*, No. 15, p. 491.

Smith G R & Shemansky D E 1978, *J. Geophys. Res.*, No. 83, p. 3783.

Tyler A L, Kozlowski R W H & Hunten D M 1988, *Geophys. Res. Lett.*, No. 15, p. 1141.

Acronyms

Acronyms

■ Science of the Moon

ALSEP	Apollo Lunar Surface Experiments Package
BVSP	Basaltic Volcanism Study Project
CRAF	Comet Rendezvous Asteroid Flyby mission (NASA)
ERS-1	European Remote-Sensing Satellite 1 (ESA)
ISTP	International Solar-Terrestrial Physics Programme
KREEP	Lunar surface material rich in K (Potassium), REE (Rare Earth Elements) and P (Phosphates)
LO	Lunar Observer (NASA)
MARSNET	MARS NETwork Mission (ESA)
MO	Mars Observer (NASA)
OPT	Orbiting Planetary Telescope Mission (ESA)
POLO	Polar Orbiting Lunar Observatory (ESA)
REE	Rare Earth Elements
SNC	Shergotty-Nakhla-Chassigny-type meteorites

■ Astronomy and Astrophysics

AKR	Auroral Kilometric Radiation
AU	Astronomical Unit
AGN	Active Galactic Nuclei
CCD	Charge-Coupled Device
CME	Coronal Mass Ejections
ESO	European Southern Observatory
FIRST	Far-InfraRed Submillimetre Telescope
HH	Herbig Haro (object)
H-R Diagram	Hertzprung-Russel Diagram
HST	Hubble Space Telescope
IAU	International Astronomical Union
IRAM	Institute RadioAstronomie Millimétrique
JCMT	James Clerk Maxwell Telescope (Hawaii)
MHD	Magneto-Hydrodynamic (waves)
NTC	Non-Thermal Continuum
SAO	Smithsonian Astrophysical Observatory
TRP	Technological Research Programme
VLF	Very Low Frequency
VLBI	Very Long Baseline Interferometry
VLTI	Very Large Telescope Interferometer

■ Physics of the Sun

EUV	Extreme Ultra-Violet
HEO	High Earth Orbit
LEO	Low Earth Orbit
LEST	Large Earth-based Solar Telescope

RC	Ritchey-Chrétien (telescope)
SOHO	Solar and Heliospheric Observatory
VLT	Very Large Telescope

■ Life Sciences

BLSS	Bioregenerative Life-Support System
CEBAS	Closed Equilibrated Biological Aquatic System
CELSS	Controlled Ecological Life-Support System
CIRCE	Computeur Intégrateur du Rayonnement Complexe dans l'Espace
COMET	Cosmic Dust Experiment on Salyut-7
COSPAR	Committee on Space Research
ERA	Exobiology Radiation Assembly
EURECA	European Retrievable Carrier
GCR	Galactic Cosmic Radiation
HZE	High Charge (Z) and Energy (E) Particles
LSS	Life-Support System
MELISSA	Microbiological Ecological Life-Support System Alternative
SCR	Solar Cosmic Radiation
SETI	Search for Extraterrestrial Intelligence
S/N	Signal-to-Noise Ratio