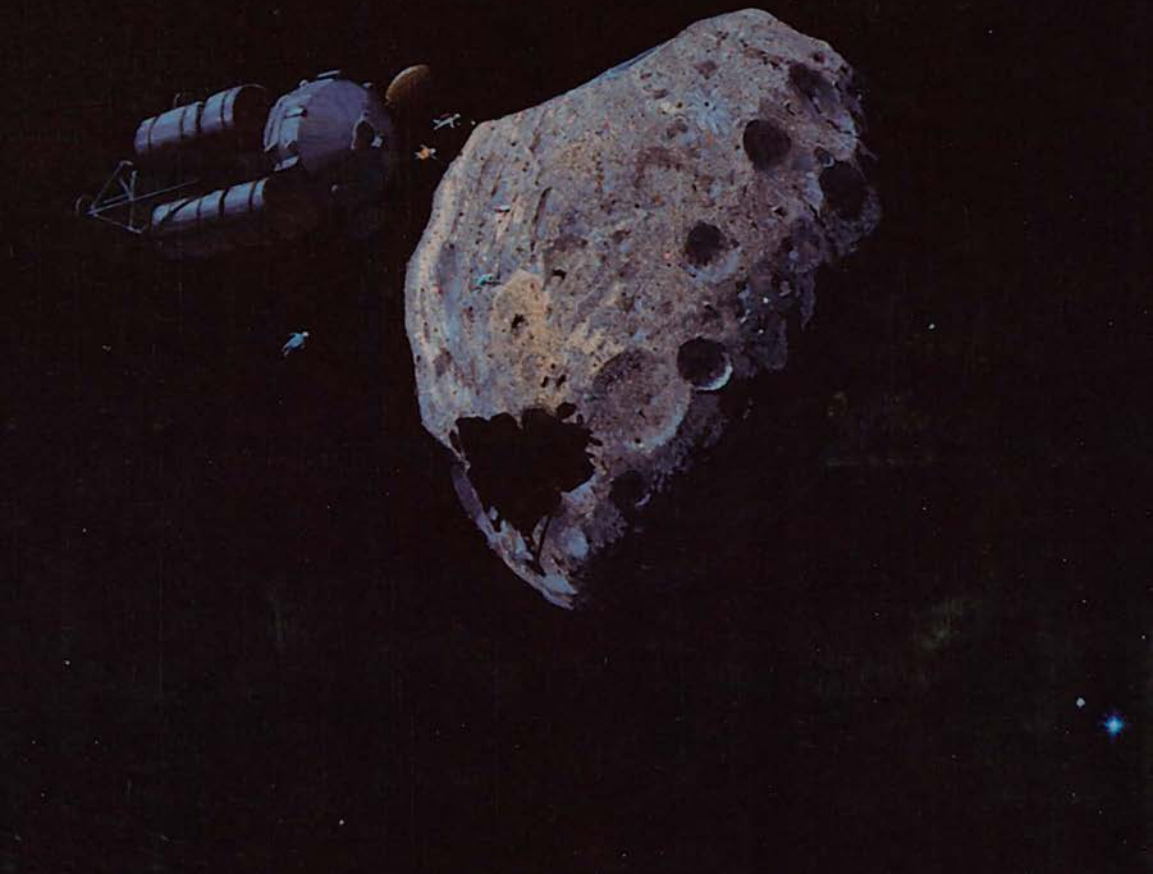
RESOURCES OF NEAR-EARTH SPACE

J. LEWIS
M. S. MATTHEWS
M. L. GUERRIERI
EDITORS



SPACE SCIENCE SERIES

Tom Gehrels, General Editor

Planets, Stars and Nebulae, Studied with Photopolarimetry
Tom Gehrels, editor, 1974, 1133 pages

Jupiter
Tom Gehrels, editor, 1976, 1254 pages

Planetary Satellites
Joseph A. Burns, editor, 1977, 598 pages

Protostars and Planets
Tom Gehrels, editor, 1978, 756 pages

Asteroids
Tom Gehrels, editor, 1979, 1181 pages

Comets
Laurel L. Wilkening, editor, 1982, 766 pages

Satellites of Jupiter
David Morrison, editor, 1982, 972 pages

Venus
D.M. Hunten, L. Colin, T.M. Donahue and V.I. Moroz, editors, 1983, 1143 pages

Saturn
Tom Gehrels and Mildred S. Matthews, editors, 1984, 968 pages

Planetary Rings
Richard Greenberg and André Brahic, editors, 1984, 784 pages

Protostars and Planets II
David C. Black and Mildred S. Matthews, editors, 1985, 1293 pages

Satellites
Joseph A. Burns and Mildred S. Matthews, editors, 1986, 1021 pages

The Galaxy and the Solar System
Roman Smoluchowski, John N. Bahcall and Mildred S. Matthews, editors, 1986, 485 pages

Meteorites and the Early Solar System
John F. Kerridge and Mildred S. Matthews, editors, 1988, 1269 pages

Mercury
Faith Vilas, Clark R. Chapman and Mildred S. Matthews, editors, 1988, 794 pages

Origin and Evolution of Planetary and Satellite Atmospheres
S.K. Atreya, J.B. Pollack and M.S. Matthews, editors, 1989, 881 pages

Asteroids II
Richard P. Binzel, Tom Gehrels and Mildred S. Matthews, editors, 1989, 1258 pages

Uranus
Jay T. Bergstralh, Ellis D. Miner and Mildred S. Matthews, editors, 1991, 1076 pages

The Sun in Time
C.P. Sonett, M.S. Giampapa and M.S. Matthews, editors, 1991, 996 pages

Solar Interior and Atmosphere
A.N. Cox, W.C. Livingston and M.S. Matthews, editors, 1991, 1414 pages

Mars
H.H. Kieffer, B.M. Jakosky, C.W. Snyder, and M.S. Matthews, editors, 1992, 1536 pages

Protostars & Planets III
E.H. Levy and J.I. Lunine, editors, 1993, 1596 pages

Resources of Near-Earth Space
J. Lewis, M.S. Matthews and M. Guerrieri, editors, 1993, 977 pages

RESOURCES
OF NEAR-EARTH SPACE

RESOURCES
OF NEAR-EARTH SPACE

John S. Lewis
Mildred S. Matthews
Mary L. Guerrieri

Editors

With 86 collaborating authors

THE UNIVERSITY OF ARIZONA PRESS
Tucson & London

About the cover:

The opportunity to acquire and utilize space resources in the next century is illustrated by this painting of a human expedition to an Earth-approaching asteroid. Astronauts float from the "parked" spacecraft to the surface of the asteroid to begin testing regolith/surface properties and techniques for resource recovery. The distant Earth-moon system is seen in the lower right. The shadow of the spacecraft, mimicking the shape of the Apollo lunar landing module, symbolized the continuity of space exploration from our generation to the next. Painting by William K. Hartmann.

The University of Arizona Press

♾ This book is printed on acid-free, archival-quality paper.
Manufactured in the United States of America.

98 97 96 95 94 93 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Resources of near-Earth space / John S. Lewis, Mildred S. Matthews, Mary L. Guerrieri.
p. cm. — (Space science series)
Includes bibliographical references and index.
ISBN 0-8165-1404-6 (alk. paper)
1. Space sciences. 2. Lunar geology. 3. Astrogeology. 4. Cosmochemistry. 5. Marks (Planet). I. Lewis, John S. II. Matthews, Mildred Shapley. III. Guerrieri, Mary L. IV. Series.
QB500.R38 1993
333.9'4—dc20 93-23753
CIP

British Library Cataloguing in Publication data are available.

CONTENTS

COLLABORATING AUTHORS

Sherwood, B., 199
Singer, R. B., 709
Snyder, G. C., 427
Spudis, P. D., 427
Staehle, R. L., 427
Stoker, C. R., 659
Swindle, T. D., 367
Taylor, L. A., 51, 69
Vaniman, D. T., 17
Waldron, R. D., 229, 257
Weissman, P. R., 569
Woo, H., 229
Woodcock, G. R., 199
Zent, A. P., 737
Zhao, Y., 149
Zurek, R. W., 845

PART I

Introduction

USING RESOURCES FROM NEAR-EARTH SPACE

JOHN S. LEWIS
University of Arizona

DAVID S. McKAY
NASA Johnson Space Center

and

BENTON C. CLARK
Martin Marietta Civil Space & Communications

The parts of the solar system that are most accessible from Earth (the Moon, the near-Earth asteroids, and Mars and its moons) are rich in materials of great potential value to humanity. Immediate uses of these resources to manufacture propellants, structural metals, refractories, life-support fluids and glass can support future large-scale space acitivites. In the longterm, non-terrestrial sources of rare materials and energy may be of great importance here on Earth.

I. INTRODUCTION

Space activities have become so much a part of modern life that they are almost invisible. We are aware how important communications satellite relays are for inexpensive long-distance telephone calls and television news broadcast, and we see the color images of Earth's cloud cover in the weather reports on TV. But we are usually not aware that the commercial airplane in which we fly or the ferry we ride locates itself by using navigation satellites 12,000 miles above our heads. We may never hear that the Soviet Union launched several satellites each year dedicated to assessing the American, Canadian, and Argentine wheat crops to help plan their grain purchases. We may be even more astonished to know that the truck we just passed on the highway is being tracked by a satellite, or that the strategic balance between the great powers has long been monitored based on photographic, electronic imaging, radar, infrared, and signal monitoring from space. Information from space gives us our first warning of preparations for war, the best guarantee against surprise attack. It also warns us of oil spills, crop disease, severe storms, ozone holes, forest fires, and climate change. Space is already worth roughly $100 billion per year to the global economy.

In addition, the exploration of space plays a fundamental role in modern culture. We daily glean new knowledge of the origin and evolution of planets, of the environment within which we live, and of the laws that govern global change. From our studies of other planets we bring home a greater understanding of Earth. We share in the excitement of discovery as human explorers explore alien landscapes. We share also in the traditions of western culture, in which experimentation and exploration constantly expand our horizons and test the assumptions and dogmas of the past. The urge to test ideas, hypotheses, and assertions against objective standards is not only congenital in our culture; it is contagious. Where such testing is absent the true cannot be distinguished from the false; the tyrant can stand unchallenged; obsolete habit can successfully withstand innovative insight.

In the emerging world of great-power cooperation and friendly competition, it is likely and highly desirable that one of the most important drivers of research and innovation, military spending, will diminish sharply in importance. But every highly industrialized country knows that technical innovation and improved productivity are the keys to economic health. Space stands out as one of very few areas of human endeavor that can not only satisfy the explorational urge and provide us the facts and insights we need to exercise our stewardship of Earth, but also keep the cutting edge of scientific and technological innovation razor-sharp.

It is in the best interests of humanity to lower the cost of activities in space, if only to make our accustomed services more affordable. This can be done in part by lowering launch costs from Earth, and in part by building spacecraft on assembly lines instead of one at a time. But there is more to the economics of space than simply continuing with business as usual. This book deals in a broad, timely way with the next giant step in human use of space: to harness the energy and material resources of nearby space, not just to lower the cost of present space activities, nor even simply to make future large-scale space activities much more affordable, but actually to make use of these resources in the service of the greatest material needs of humanity. Because we are concerned with the near future, we concentrate almost solely on the resource potential of nearby bodies in space. We consider the Moon, the near-Earth asteroids, and Mars and its two moons, Phobos and Deimos.

II. THE MOON

A total of 56 spacecraft (out of 75 attempts) were sent to the Moon or its vicinity between the first lunar mission in 1959 (the Soviet Luna 1 flyby) and the last in 1976 (the Soviet Luna 24 robotic sample return mission), a period of about 17 years. This golden age of lunar exploration matured through a number of stationary robotic landers in the Luna and Surveyor series, several orbital photographic mapping missions in the Lunar Orbiter and Luna programs, three robotic sample returns (Luna 16, 20 and 24), and two robotic rovers (Lunokhod 1 and 2), to six human landings in the Apollo

program, which returned a total of 381.7 kg of lunar materials. Despite the efforts of many to promote new flight programs, only a single lightweight mission, Japan's Muses A, has been sent specifically to the the Moon since 1976, a time interval roughly equal in duration to the entire golden age of lunar exploration in which those 56 missions were flown. The only large spacecraft to approach the Moon in these years is the Jupiter-bound Galileo mission, which flew by the Earth-Moon system in 1991 and late 1992 for reasons related to its trajectory to Jupiter, quite incidental to lunar science. What did we learn from the golden age of lunar exploration? Why did we quit going to the Moon for so long? What might justify a return to the Moon?

As the data from these early missions acccumulated, and as the returned samples were analysed in detail, it became clear that the Moon was the scientific key to much of the early and intermediate history of the Solar System; in particular, to how Earth formed and evolved. The lunar samples probably have been better studied than any samples from anywhere else, including Earth. Such studies have led to the hypothesis of the formation of a lunar core, mantle, and crust, and of an early "magma ocean" from which the lunar highlands formed. The Moon was found to have had a complex volcanic history, with many kinds of basalts and trace-element-rich differentiates forming and extruding or intruding, principally during the first two billion years of lunar history. Extrusive volcanism on the Moon tapered off and apparently ceased some time between roughly two and three billion years ago.

One crucial result of lunar sample studies is that we now know in great detail the physical properties and the chemical and mineralogical composition of a broadly diverse set of rocks and soils collected from nine sites on the near side of the Moon, a set that appears, based on remote sensing data, to be broadly representative of the entire near side.

We do not, however, have global coverage of the Moon. Only a rather narrow band, most of it close to the equator, was covered by the orbital geochemical mapping (gamma-ray spectroscopy) experiments carried out during the Apollo program. Although Earth-based multispectral mapping has continued to improve since the Apollo era, we still do not have direct chemical data on most of the Moon's surface. Thus the task begun by Apollo, to carry out complete geochemical mapping of the Moon, remains as unfinished business. This precise mission will be the central task of two proposed robotic orbiters which may fly as early as the mid-1990s. In addition, a series of robotic landers have been proposed for the middle to late 1990s. These landers will be used to confirm and calibrate the orbital remote-sensing data at specific landing sites. They will also provide detailed geochemical, mineralogical, and geophysical data on the subsurface, and help certify potential landing sites for human visits to follow around the end of the century.

Visits to the Moon during the Apollo program were short and intense. No effort was made to use lunar materials to help support the crew or to provide anything other than scientific samples for return to Earth. In fact, lunar dust

was a major annoyance in mission operations: dust contamination of the space suits was so severe that it would have seriously limited any additional lunar surface activities beyond those limited tasks actually accomplished by the Apollo landings. When we again resume flying crews to the Moon, unlike Apollo, we will make major efforts to take advantage of lunar materials in a variety of ways to support the needs of the lunar outpost and its transportation system. We may also find ways to use lunar resources to help us return useful products from the Moon to Earth.

Lunar materials can be used to support human activity at an outpost in a variety of ways. The earliest use of regolith or rocks may be for radiation shielding and for protection against blowing dust raised by ascent and descent rocket engines. Thermal isolation, heat storage, and ballast mass for cranes or trucks are other possible early uses of raw (unprocessed) lunar materials. Our early experiments with actual processing will probably be aimed at the extraction of volatiles, especially at oxygen production from lunar minerals.

For many years after resuming operations on the Moon, the use of lunar resources with the greatest economic significance may be the manufacture of rocket propellants. When a crew lands on the Moon, about half the mass landed on the surface will be the propellant intended for use on the return flight to Earth. If this propellant could be produced from local materials, half of that payload capacity could be saved. Either the mass of payload delivered to the Moon could be doubled, or the size of the vehicle departing from Earth could be halved. Even if the hydrogen needed for takeoff from the Moon had to be carried along from Earth, the use of locally extracted oxygen could reduce the amount of propellant that must be lifted from Earth by a factor of ten. Indeed, O:H mass ratios of up to 12:1 could be used by the lunar ascent engines with little performance penalty. If lunar oxygen can be produced reliably, it becomes feasible to operate a lunar-based lander, using lunar oxygen for both ascent and descent, to rendezvous in lunar orbit with a vehicle sent from Earth. The sooner a system can be set up on the lunar surface to produce and store lunar oxygen, the sooner these substantial savings on transportation costs can be realized. Eventually, perhaps even within a few years, it may be possible to export lunar-produced oxygen from the Moon to low Earth orbit, where it might be economically attractive to use it as propellant in support of lunar resupply operations or even missions to Mars.

Solar wind hydrogen extracted from the lunar regolith may eventually be used for propellant and to produce water for use on the Moon. Water itself may be used for human life support, growing crops, controlling dust in airlocks and inside habitats, and for protection against cosmic radiation. Interestingly, a given mass of water is a much better shield against energetic protons than an equal mass of lunar rock, dirt, or even metals. Water has the added advantage of flexibility, in that it can be poured or pumped from place to place, used inside hollow walls, or stored in bladders covering the roofs of habitats.

Other uses of lunar materials will evolve with time. We anticipate locally

produced and fabricated metal structural elements, since metals are byproducts of all schemes for oxygen production on the Moon. Metals can also be used in the construction of electrical distribution systems, at first for outpost utilities, but later for solar-produced electric power which may be exported to Earth. In addition to this beamed solar energy, ^{3}He is another potential export product which will be of great value if large-scale fusion power generation becomes a reality on Earth.

Lunar exploration may also reveal new and unexpected deposits, including water at the poles, water trapped from impacts of carbonaceous asteroids and comets, or local geochemical concentration of sulfur, potassium, chromium, and other potentially useful resources. If such resources are found, they may radically alter both our general plans and our specific timetable for lunar base development and resource utilization. Therefore an early and aggressive program of robotic exploration may be justified on the basis that it may have a major effect on lunar outpost planning. But even if major surprises do not appear, geochemical mapping from orbiter and lander missions will help us select the optimum outpost site, and will provide the scientific framework for future detailed science investigations. The scientific characterization of surface chemical and physical properties will be accompanied by complementary experiments on manipulation and processing of regolith materials and primary rocks.

The early years of a renewed lunar program will include a resource utilization demonstration on the first human flight, a small oxygen production plant (10 to 15 tonnes yr^{-1} capacity) within three or four years, and a much larger plant with 50 to 100 tonne yr^{-1} capacity within a few more years. Within the constraints of currently envisioned launch systems, cargo payloads will be limited to about 25 tonnes. However, preliminary calculations suggest that a 100 tonne yr^{-1} oxygen plant, including its own electric power system, may be possible within that mass limit.

The long-term purpose of our return to the Moon is to make lunar operations as self-sufficient as possible, and perhaps even to "show a profit" through the export of materials that are useful in space or on Earth. This goal can only be reached through an aggressive utilization program which promptly and greatly enhances the ease of delivery of payloads in the lunar transportation network. The best measure of the success of our renewed lunar program is whether we go to stay permanently: we shall do so only if we learn to use local resources in greater quantities and ever more efficiently to support ourselves. As we approach self-sufficiency, we also increase our ability to perform more and more sophisticated science. This science program will include not only lunar investigations, but also astronomy, radio astronomy, space physics, and life sciences. Much of this is new science made possible by having a base that can provide its own life support materials, and expand its own living space, laboratory space, and infrastructure through use of local materials.

III. THE NEAR-EARTH ASTEROIDS, PHOBOS, AND DEIMOS

The near-Earth asteroids (NEAs), with orbits that cross or graze Earth's, are hard to study from Earth, and have never been visited by a spacecraft. Indeed, our first (and only) asteroid flyby occurred when Galileo passed the main-belt asteroid Gaspra en route to Jupiter in 1991. The small asteroid-like Martian moons Phobos and Deimos have been studied by the Mars-orbiting Mariner 9 spacecraft. The ill-fated Soviet Phobos mission, launched in 1988, lost one spacecraft en route to Mars and the second shortly after the beginning of close scrutiny of Phobos. The surfaces of these bodies are very dark, like carbonaceous asteroids, but they lack a detectable absorption feature due to chemically bound water.

The 200 known NEAs have a size distribution generally similar to that in the main belt, with smaller bodies vastly more abundant than large ones. The largest NEA, 1036 Ganymed, has a diameter of 40 km, and the smallest known NEAs (1991 BA, 1991 VG, and 1992 DU) have diameters of 8 or 9 m. Forty-four NEAs have been studied photometrically to determine their compositional types: these include metallic, carbonaceous, primitive chondritic, and differentiated basaltic asteroids. A large fraction, perhaps as much as 60%, of the NEAs are extinct comet cores, possibly with massive ice cores covered by a few meters of dark, fluffy carbonaceous dust. Recently one NEA, 1979 VA, was found to be identical to comet Wilson-Harrington, which was observed as an active comet in 1949. The NEA population samples the asteroid belt very widely, and is very diverse: there is nearly as much compositional variety in the near-Earth population as in the entire asteroid belt.

Because of their planet-crossing orbital paths, the most likely long-term fate for the NEAs is to collide with Earth or Venus. In January of 1991 a tiny asteroid, 1991 BA, flew by Earth at less than half the distance of the Moon. Given its orbital velocity relative to Earth, the impact of 1991 BA on the Earth's surface would have caused an explosion with a yield equivalent to 200 kilotons of TNT, ten times the size of the Hiroshima atomic bomb. Random impacts of megaton size occur every few decades somewhere on Earth, and raise the fear of an accidental impact in a sensitive area triggering World War III. It is, however, technically possible to discover large numbers of NEAs and determine their orbits as a sort of "early warning" system. In recent years, photographic techniques have demonstrated a discovery rate of up to 15 NEAs per year. Now, with the Spacewatch program operational, it alone should be capable of finding 2 or 3 nearby asteroids per month.

Another consequence of their orbital paths is that NEAs are ideally situated to provide meteorites to Earth. It seems likely that the most common classes of meteorites falling on Earth are fragments knocked off some of the NEAs by a small number of recent collisions with other pieces of debris. These meteorites are our best source of information about processes in the early solar system, during the era of planetary formation. Cataloging and

analysing large numbers of new meteorite classes, and establishing confirmed relationships between particular meteorite types and particular asteroids, are both very helpful in unraveling the complex and important early history of the planets.

Because of their nearby orbits and their small size, many of the NEA bodies are energetically more accessible than the Moon. About 45 of the 200 known NEAs are easy enough so that a given booster rocket could soft-land a larger payload on them than on the Moon. The *return* trip, starting with takeoff from a body with a surface gravity of less than a thousandth of Earth's surface gravity, is even easier. It is possible to depart from one asteroid, 1982 DB, and return to intersection with Earth with a delta V of only 60 m s^{-1}. By comparison, the delta V to return to Earth from the Moon is about 6000 m s^{-1}. Because the payload mass decreases exponentially with the delta V requirement, it is clear that a given rocket engine can lift far more mass off an asteroid than it could lift off the Moon. Mission opportunities to a given NEA occur about every 2 to 4 years. The number of NEAs greater than 100 m in diameter is about 70,000: if we knew the orbits of this many NEAs, we could expect a launch window to open up about every 20 minutes.

Compared to the Moon and Mars, asteroids are clearly less traditional targets for manned exploration. Phobos and Deimos, lying as they do on the doorstep of Mars, and possibly containing great quantities of volatiles for use in manufacturing rocket propellants, are more obvious targets for manned exploration and exploitation. There are, however, good reasons to carry out an early test of the hardware for a Mars mission in a low-gravity environment, with full-duration engine burns and full Martian round-trip time, without the hazards of operating deep in Mars' gravitational field. In the long run, economic considerations may make manned visits to these asteroids desirable. But it is easier to foresee unmanned missions to the NEAs, not only to explore them and return samples to Earth, but also to retrieve very large masses of raw or processed asteroidal material to cislunar space. NEAs may be the most attractive source of shielding, propellants, metals, and refractories in orbits around the Earth. They might provide the building materials for solar power satellites, propellants for Mars missions, or life-support materials for use in a lunar base. The resource richness and diversity of these asteroids and their energetic accessibility favor their economic exploitation; the long trip times make them less attractive.

It is known that simple material return missions that do no processing at the asteroid can return, to low Earth orbit (LEO) or Earth, a mass of asteroidal material up to six times as large as the mass that had to be launched into LEO in order to carry out the mission; we say that the mass payback ratio (MPBR) is about 6. A comparable mission to the Moon would have an MPBR of about 0.1. If *in-situ* chemical propellant production (especially liquid oxygen) is allowed, and an aerobrake is carried to facilitate return to LEO, then the MPBR for the Moon rises to about 2.4 after repeated round trips. The mass of the aerobrake brought from Earth (or, alternatively, the mass of propellant

brought from Earth for use in achieving return to LEO) is large, and is the main factor limiting the MPBR of these missions. Extraction of water from a well-situated NEA, combined with either nuclear thermal or solar thermal propulsion (and without use of aerobraking upon return to Earth) permits an MPBR of about 100:1 for repeated round trips. Manufacturing an aerobrake in space, rather than bringing it from Earth, confers similar advantages on asteroidal return missions.

Such large MPBRs make it likely that the return of very large masses of nonterrestrial raw materials, such as metals or carbonaceous asteroid material, to selected Earth orbits can be achieved economically. These materials can then be processed into SPS construction members, propellants, life-support fluids, glasses, ceramics (for heat shields), and radiation shielding. The challenge is to devise a combination of resources, transportation systems, extraction and processing technologies, and fabrication and assembly techniques that can deliver useful products to the point of use at less cost than direct launch from Earth.

The feasibility of devising such end-to-end schemes for profitable use of NEA, Phobos, and Deimos materials depends on several scientific and technical developments. First, programs to discover and spectrally characterize small solar system bodies must be broadened and strengthened. Second, spacecraft missions to fly by and rendezvous with a variety of these bodies must be undertaken to give us an adequate understanding of the chemical and physical state of their surfaces. Third, we have seen that landing and sample return from these bodies is little more difficult than rendezvous. Thus the direct investigation of their surfaces by spacecraft and the thorough characterization of their materials in laboratories here on Earth are both surprisingly easy to accomplish. In parallel with these scientific investigations, mining and engineering research aimed at handling dirt in microgravity and extracting useful products from classes of NEA material that are found in our meteorite collections are also essential. The Space Station could be an excellent platform for such research.

IV. MARS

Even before the Space Age, the planets Mars and Venus held a special fascination. Here were the bodies upon which other life forms might exist; worlds that someday might be another home for an adventurous mankind. Mars held a special allure because its surface was not obscured with clouds as was Venus and there were tantalizing geographic features that seemed, to some astronomers, to evidence a regularity that could be explained only by an economically active civilization.

With the advent of interplanetary spacecraft, these views were radically changed. Conditions on the surface of Venus, as discovered by the Soviet Venera spacecraft, are extraordinarily hostile to living organisms: high pressures (100 times that of Earth's atmospheric pressure), very high temperatures

(sufficient to melt lead), and bone-dry humidity. Meanwhile, the conjectured canals of Mars turned out to be just that. Mars too was at first seen, as viewed by the flyby of Mariner IV, as inhospitable to life, a barren, cold desert with landscape remarkably similar to the scarred surface of our lifeless Moon. It was only with the subsequent missions of Mariners 6, 7, and most notably the orbiting spacecraft Mariner 9, that scientists could begin to appreciate the geological richness and complexity of the Martian environment. A variety of evidence indicates that the environment at one time was warm and wet enough that liquid water helped shape the Martian landscape. The winter polar caps, originally thought to be water ice in analogy with Earth's, were shown to contain carbon dioxide ice ("dry ice") by the Mariner sensors. Later, the extended Viking orbiter missions demonstrated both concepts were to some extent correct, with both water and CO_2 ice present, depending upon the season. Indeed, the permanent north polar cap which remains after the larger dry ice cap sublimates away during the local summertime is composed of H_2O.

Even though Mars is today a frozen desert, much water in the form of ice, water vapor, and adsorbed films on powdery soil is present and can be readily extracted by heating the soil or compressing and cooling the atmosphere.

The most important breakthroughs in our understanding of Mars came with the touchdown of the Viking Landers in 1976. Although the primary mission of these Landers was to detect biological activity and the nature of organic compounds in soil, much was learned about the physical and geological characteristics of the surface. In fact, neither microbial lifeforms nor organics were detected. Many aspects of the chemistry of the atmosphere and soil were determined, however. The atmosphere was found to be mainly CO_2 with small amounts of N_2 and Ar, and even lower quantities of CO, O_2, and H_2O vapor. The soil was found to be generally silicate in composition with a high iron-mineral content, including some magnetic material, but also including high concentrations of sulfate and chloride salts and even traces of bromides as well. Without the salt content, the residual elemental composition is strikingly similar to the composition of the Shergotty meteorite (now thought to have originated on Mars).

Even with these major advances by space probes, Mars remains a planet of many mysteries. Is Mars seismically active, and how did its geological history differ from its nearest neighbor, Earth? Why has it virtually no magnetic field? What happened to cause the climate to change from a more temperate one to its current frigid state? Where is the water now locked up? Are the giant volcanoes or the many smaller ones still active? Was life once present, and might it still be present in individual fumerolic oases, much as the specialized flora and fauna that inhabit individual volcanic vents in the deep oceans on Earth? What determines when and how violent the regional and global dust storms occur? The answers to these scientific puzzles are analogous to many of the leading scientific questions still being asked about Earth, and may shed light on how better to control the environment of our own planet.

Mars is also the one terrestrial planet where humans could live, work,

and develop a rewarding kind of self-sufficiency. The day/night cycle is admirably suited to the natural circadian rhythm of humans, although many other aspects of the environment are harsh, particularly the low temperatures and low atmospheric pressure. The latter is by far the most significant new challenge to adaptation and will require spacesuits for outdoor activities, although the low gravity level (about 38% that of Earth) makes such suits feasible without excessively burdening the astronauts. Perhaps the greatest discovery of the space age has been that Mars indeed is a volatile-rich planet, containing abundant quantities of light elements compared, for example, to our Moon. Light elements, such as hydrogen, carbon, nitrogen, and oxygen (H, C, N, O) are not only the key elements from which tissues and cells are made, but are also the critical ingredients of a majority of the manufactured goods which provide much of our quality of life. Even the chemical forms of these elements are directly useful. For example, it would be possible to raise crops using the abundant CO_2 in the atmosphere and H_2O extracted from the soil. Breathing oxygen could be chemically extracted from either of these substances, or provided as a byproduct of plant photosynthesis. Entire biospheres could be created in enclosed environments by piping in sunlight which, although somewhat weaker at the surface of Mars because of the greater distance from the Sun and the dusty atmosphere, is nonetheless quite adequate for plant growth.

Propellants such as hydrogen, methane, methanol, oxygen, and hydrogen peroxide can all be made in relatively direct chemical processes from these compounds, alleviating the supply line from Earth for return propellant and rover propulsion. Using a combination of biomass extraction, chemical synthesis, and photochemical processes a wide variety of organic materials could be produced. Metals, salts, sulfur, acids, and many other materials could be extracted from the soil minerals. Paving stones could be cast by wetting and compacting the salt-rich soil.

The realization that Mars is scientifically of highest importance and is a resource bonanza at the same time that it could be habitable for humans is responsible for the increased interest in future missions. The National Commission on Space singled out Mars exploration and eventual colonization as the major objective for the next 50 years in space. The Committee for the Future of the U. S. Space Program calls Mars "the ultimate goal."

There is currently no shortage of proposed methods for conducting the next phases of Mars missions. The official missions include the U.S. Mars Observer orbiter, launched in October of 1992, and a Soviet orbiter/lander mission which may include a rover and balloon experiment as well and is scheduled for a 1994 launch. Ranging from small cheap hard-landers to very large, long-range roving science platforms, the future exploration of Mars by unmanned robotic or automated missions can be very ambitious. However, the large distances between Earth and Mars limits roundtrip communications to 8 minutes at best, and up to 40 minutes at the worst. With these long delays, telerobotics becomes tedious and inefficient. Quasi-autonomous operation of

these vehicles will become necessary to allow any possibility of reasonable accomplishments. Yet, if human explorers were at the surface, they could make field sorties orders of magnitude faster and much more scientifically rewarding because of the possibility of serendipitous discoveries. The flexibility and cognitive functional capabilities of humans as compared to robots would be highly leveraging for understanding new and complex sites. Probably only human presence would permit the discovery of fossils or other major revelations which cannot reasonably be anticipated in advance.

Before humans should go to Mars, it will be important to characterize certain aspects more thoroughly than has so far been possible. Rovers will demonstrate trafficability through the loose, powdery soil and over the abundant angular rocks. A sample return mission should be of extremely high priority so that the soil may be studied in the finest laboratories on Earth and methods devised for the most efficient utilization of its chemical resources. In addition, there are those who fear the possible presence of chemical or biological activity in the soil which could have toxic or other deleterious effects on humans. With careful preservation of robotically returned samples, these properties could be measured and understood.

The reasons for sending humans to Mars are many. Science is one justification. Colonization can be another. Learning more about the most Earth-like body in the solar system has direct benefits in our future on our home planet. Technological developments in high-efficiency food production under adverse physical conditions, robotics, exploration scientific equipment, waste management and pollution control in confined quarters, advanced communications, expert control systems, and many other fields will accompany any project for human missions to the planets; and, although not likely by current understanding, there are nevertheless the results that can occur whenever probing the unknown, such as finding some natural material of extraordinary value on Mars which would economically justify the expense of bringing it back to Earth.

V. HOW THIS BOOK ADDRESSES THE ISSUES

In the next few years, for scientific or cultural reasons, humanity will decide to build a base on the Moon, and to send an expedition from Earth to Mars. How can we increase the scope and reduce the cost of these ambitious activities? In this book we discuss how we can greatly reduce the cost of a lunar base by using lunar regolith to shield it from energetic radiation. We can further reduce the cost by extracting liquid oxygen from lunar oxide minerals and using it to supply air for the base and rocket propellant for the return to Earth. On Mars, we could process the atmosphere to make liquid carbon monoxide and oxygen or other propellants for local transportation, or for takeoff on the return to Earth. We could use local resources to build metal structures, make refractory aerobrakes, build enclosed biospheres, and grow food. Space

activities can be made ever more autonomous, ever freer of the need to launch vast masses of material from Earth at great cost.

But there is another even more ambitious possibility: that energy and material resources might be profitably returned from nearby space to Earth. If we inventory our most pressing needs on Earth, we find an interesting, pervasive theme: whether we seek a source of fresh water in desert areas, a means to manufacture fertilizer and power farm equipment in underdeveloped countries, or a way to transport food to where it is most needed, the essential requirement is abundant, cheap energy. But, thanks largely to our ability to monitor global temperature, ozone, and atmospheric pollution from space, we are aware of a second, equally pressing, necessity: to find a source of abundant, cheap energy *with greatly reduced environmentally destructive side-effects*. We need energy without radioactive waste; without carbon dioxide emissions; without strip mining, smog, soot, and massive oil spills. We need to find a source of cheap, clean energy. Perhaps the most promising prospects for achieving this goal are the construction, using nonterrestrial materials, of solar power satellites in high orbits around the Earth or on the lunar surface, beaming power down to receiving antennas on the ground, and clean fusion of nonterrestrial ^{3}He with terrestrial deuterium.

The first step in making these dreams come true is to build up our base of scientific knowledge of what is available in space; to discover and characterize the resources. Next, we need to develop the technology to mine, beneficiate, extract, and process these resources, and fabricate them into useful products. We then need to understand the transportation and logistics system necessary to go where the resources are, deliver them to a processing plant, and carry the products from the factory to their site of use. Finally, we need to analyze the economics of this proposed scheme in order to see whether it is worth doing. This volume summarizes the present state of the art in each of these activities. Our purpose is to present a broad, up-to-date survey of a young and rapidly evolving field. It is intended as a technical introduction to the use of nonterrestrial materials for scientists, engineers, and industrial and governmental project managers who seek to make space more accessible.

Can we design a future in which space activities defray their own costs? Or can they be made wholly self-supporting, undertaken for the economic benefit of humanity? Just because this field is young and rapidly evolving, we cannot yet answer these questions with confidence, but there are so many attractive prospects that it would be irresponsible not to seek the answers diligently. The stakes could not be higher.

PART II

The Moon

A GEOCHEMICAL ASSESSMENT OF POSSIBLE LUNAR ORE FORMATION

LARRY A. HASKIN and RUSSELL O. COLSON
Washington University, St. Louis

DAVID T. VANIMAN
Los Alamos National Laboratory

and

STEPHEN L. GILLETT
Mackay School of Mines

From studies of lunar samples we know that extensive chemical fractionation occurred during the Moon's igneous differentiation, and from remote sensing studies we know that the Moon's outer crust is laterally and vertically heterogeneous on large and small scales. Here, we review current knowledge of chemical compositions of lunar materials in the context of known or suspected lunar geochemical processes. We speculate on how various elements might have been concentrated into potential ore deposits.

I. INTRODUCTION

We must seek the economic high ground of near-Earth space, a major arena of international exploration and speculation. This means learning what is in near-Earth space and how we might use it. This in turn requires careful assessment of material resources of the Moon.

Of course, we can obtain from Earth any material we might wish to use in near-Earth space or on the Moon. However, the escape velocity from Earth is 11.2 km s^{-1}, whereas the escape velocity from the lower-gravity Moon is only 2.4 km s^{-1}, so much less energy is required for liftoff into Earth-Moon space. Thus, there may be combinations of materials and transportation costs that will make lunar products more economical for use in space than equivalent materials lifted from Earth, particularly when large masses of material are required.

Apart from the Moon, the nearest bodies that might serve as resources for use in near-Earth space are the Earth-orbit-crossing asteroids (see, e.g., Lewis and Lewis 1987). These objects have almost negligible liftoff energies. Unfortunately, current knowledge of the presence of regolith that might be

easily mined and of chemical compositions rests on spectroscopic evidence and is unconfirmed by surface sampling. Also, each asteroid approaches Earth only infrequently. Furthermore, although the near absence of gravity may make liftoff energy negligible, it may make materials extraction more difficult.

Many suggestions about mining extraterrestrial bodies to provide material for use on Earth justly suffer economic disrepute. Most plausible uses proposed for extraterrestrial materials are for use in space. The only currently fashionable candidate for use of an extraterrestrial material on Earth is solar-wind-implanted ^{3}He, a promising fuel for commercial fusion power. Mining it from the Moon's regolith is being considered seriously (Wittenberg et al. 1986). Costs associated with a cleaner environment on Earth may eventually make extraterrestrial production of electrical power and some manufactured goods economically competitive with terrestrial production, but competing innovations in terrestrial technology may prevent this.

To estimate the value of the Moon's resources, we must know the chemical and mineralogical compositions and physical states of the surface materials of the Moon, and we must understand the geochemical processes that produced them. Ores may be produced by the principal processes of planetary differentiation operating locally to extremes and by minor processes that locally concentrate elements. Most economic concentrations on Earth result from local, unusual circumstances. Our present knowledge of lunar soil and rock compositions has been summarized recently (Heiken et al. 1991). That knowledge is based on a meager sampling of the Moon's surface and estimates from remote sensing measurements. It has not yet led us to a quantitative understanding of the major processes of lunar chemical differentiation and has provided only hints about minor ones. Thus, for accurate resource forecasting, we need to identify and quantify the pertinent chemical and physical processes that formed the early lunar crust and those that have changed it since. For this, as well as for direct observational discovery of potential ores, we need much more detailed mapping and sampling of the lunar surface. The more we know about lunar surface compositions and the better we understand lunar geochemical processes, the more confidently we can predict what kinds of mineral deposits formed on the Moon and where we should look for them. The first purpose of this study is to indicate our present level of knowledge and to suggest possible directions for further work. In this, we summarize and build on earlier discussions of geochemical processes that might have produced concentrated lunar resources (Gillett 1983,1990*a*,1991; Haskin 1983,1984,1985; Taylor 1990*a*, *b*).

The dry state of the Moon's interior and surface complicates our perception of lunar ore potential. Most ores on Earth depended on internal or external water for their formation. The Moon apparently has no appreciable internal or surface water, so we may not expect to find most types of terrestrial ores on the Moon. Nevertheless, there is abundant evidence that substantial geochemical separations occurred during the formation of the Moon's crust.

Moreover, the Moon is laterally heterogeneous, and recent Galileo data indicate this heterogeneity extends to the farside (see, e.g., Pieters et al. 1991). Thus, the second purpose of this study is to consider the present evidence for the types of geochemical separations that occurred on the Moon during its development. We then speculate whether those processes might have been capable of producing concentrated mineral deposits that might someday serve as ores.

Knowledge of mineral deposits is only part of the story. At any given time, it is the combination of resource character, state of mining and processing technology development, and product need that will determine what deposits on the Moon might be regarded as ores (i.e., economical sources of material). We must, therefore, devise, test and develop technologies that could extract needed materials from potential lunar ores. These technologies must be appropriate to lunar surface conditions (Haskin 1985). We are adapted to exploiting the Earth, with its wide variety of familiar raw materials, including abundant and cheap oxygen and water, and an elegant and intertwined network of technologies, most of which use oxygen and water freely. Because the Moon has no air, liquid water, or industrial infrastructure, meaningful economic analysis of the value of lunar materials (and, similarly, other extraterrestrial materials) is especially difficult.

This discussion is based on the classification of chemical elements into six categories (Haskin and Warren 1991). From a resource point of view, different geochemical processes are responsible for the behaviors of different element groups. The highest known concentrations of elements in three of the element groups appear to result from processes acting on atoms originally in the lunar interior; these groups are the major elements (ME, those such as oxygen and silicon that comprise the main mass of the lunar crust), the minor elements (MIE, those such as phosphorus that are abundant enough to form minerals in which they are essential components but that constitute $< 1\%$ of the crust), and the incompatible trace elements (ITE, elements of intrinsically low abundance but concentrated in the lunar crust relative to the lunar interior). The highest concentrations of two groups of elements on the lunar surface result from processes acting on matter of external origin, i.e., atoms introduced to the Moon after it had developed a solid surface. These groups are the siderophile elements (SE, those that readily form alloys with Fe metal, and whose main lunar source is meteorites), and the solar-wind-implanted elements (SWIE, elements that are very volatile or whose compounds were very volatile at the time the Moon formed and that have collected in the lunar soil through capture of solar-wind ions). The sixth group comes from both internal and external sources. These are the vapor-mobilized elements (VME, trace elements whose compounds vaporize at relatively low temperatures and that appear to have moved within the Moon by magmatic heating and on its surface by impact heating, and whose external source is meteorites). From our point of view as intended users of resources, a major concern is the extent to which the Moon may have developed exploitable mineral deposits of the various chemical

elements. Below, we speculate on how elements might have been gathered into concentrations that could serve as ores.

II. THE ORIGIN OF THE MOON AND ITS DIFFERENTATION INTO MANTLE AND CRUST

Two conditions must be satisfied before a mineral body that might serve as an ore can be produced. A natural mechanism that can concentrate a desired chemical element must operate, and the initial abundance of the element must be high enough that the mechanism can concentrate it to a useful extent. It turns out that the Moon's average surface concentrations for many chemical elements are similar to Earth's. This is reasonable, as both planets appear to have formed in the same area of the solar system and both underwent extensive internal chemical separation to concentrate most chemical elements other than Fe and Mg into their crusts (see, e.g., Anders, 1977; Drake 1986; S. R. Taylor 1982,1986; Larimer 1986). However, there are obvious and important differences between the two planets. Thus, it is useful first to provide a brief framework for understanding the behaviors of different groups of elements during formation of the Moon.

The best model devised so far for the origin of the Earth-Moon system (Hartmann and Davis 1975; Hartmann 1986; Stevenson 1987; Melosh 1989) in outline is roughly as follows: in our general region of the solar nebula, a large proto-Earth formed and underwent internal chemical separation to produce a metallic core and a silicate mantle. A Mars-sized object formed simultaneously and also produced a core. The two objects collided. Their cores merged and their mantles partially merged but partly vaporized to form a ring around the Earth, from which material condensed and then accreted to form the Moon. In this scenario, the Moon has no appreciable core, because its core-forming metal was lost to the Earth. The composition of the Moon's mantle resembles that of Earth's mantle because both of the colliding objects had accreted in the same region of the nebula. The Moon has low internal concentrations of relatively volatile elements because they evaporated into space instead of condensing after the collision. Alternatively, the highly volatile elements may not have accreted originally, but may have been added as a late veneer after the Moon and Earth had formed (see, e.g., Smith 1982; Carr and Wänke 1991). The Moon may have been too small to retain them. After formation, the mantles of both the Earth and the Moon differentiated chemically to form Mg-Fe-silicate mantles and aluminous, trace-element-enriched crusts.

From the >4.4 Gyr ages of some highland rocks, we know that the Moon differentiated so rapidly that at least its outer few hundred kilometers must have melted extensively. This information has led to the "magma-ocean" model for lunar differentiation (see review by Warren 1985*b*), which we introduce here in simplified form. In this model, which resembles that for a classical "layered intrusion," but on a planetary scale, a global ocean of magma

once covered the Moon; this ocean leaked its heat into space. As it cooled, it initially crystallized only dense Mg- and Fe-rich minerals, which, because of their high densities, settled toward the ocean floor. As the composition of the residual magma evolved, ilmenite, clinopyroxene and feldspar also crystallized. The feldspar, less dense than the residual magma, floated to form a crust. The last of the residual liquid remained beneath the crust; it was rich in incompatible trace elements (ITE). Later events released some of the residual liquid to the surface. Later still, remelting of deeply buried Fe-Mg-rich minerals produced fresh basaltic magma that rose to form the lavas of the lunar maria. About 20% of the lunar surface area consists of mare lavas, which lie as a thin veneer on the thick highland crust. Arguments continue over the nature and extent of the Moon's crustal differentiation and how feldspar-rich the crust is (see, e.g., Korotev and Haskin 1988). Large-scale convection and ongoing giant meteoroid impacts undoubtedly complicated the crystallization of the ocean. In any event, the Moon underwent protracted, large-scale igneous fractionation early in its history that left it laterally and vertically heterogeneous.

As they formed, the Moon and Earth were continually struck by residual planetesimals and smaller objects. This bombardment continues today, but its intensity has tapered off as the population of impactors decreased (see, e.g., *Basaltic Volcanism Study Project* 1981, Ch. 8; S. R. Taylor 1982). Bombardment produced the extensively cratered surface of the Moon and it excavated the great basins that contain the maria, perhaps in an unusually catastrophic set of collisions as recently as ~4 Gyr ago (Tera et al. 1974; Ryder 1990). Impacts had shattered and partly mixed the igneous rocks of the highlands, so few intact samples were found. Most lunar surface materials are soils (complex products of impact pulverizing, mixing and melting processes; see, e.g., McKay et al. 1991) and breccias (rocks composed of sintered fragments of earlier rocks and soils). The bulk of the soil particles is very small. Some 50% by weight of a typical soil will pass through a 70 μm sieve. Together, soils and breccias along with some fragments of unbrecciated rock, make up the regolith. Remote sensing measurements receive their signals from this mostly fine-grained material. The flooding of the maria with lava post-dated the heaviest bombardment, and many samples of well-preserved basaltic lavas have been collected from regoliths covering the maria.

The impact processes of heating, excavation and pulverization may also have created conditions for ore formation. However, through dispersion of their target material, impacts destroy ore bodies as well.

III. ELEMENTS CONCENTRATED MAINLY FROM EXTRALUNAR SOURCES

A. Solar-Wind-Implanted Elements

The solar-wind-implanted elements (SWIE) group consists of H, He, C, N, Ne, Ar, Kr and Xe. It has been argued informally that the Moon should

be written off as a practical source of material or as a possible location for settlement because it has no water or carbon and is virtually incapable of life support (e.g., an "...airless, waterless, dead world"; Sagan 1989). In fact, this surmise is incorrect (Haskin 1992); total lunar surface abundances of these elements are high, although concentrations are low. The Moon appears to have little, if any, *internal* water, carbon, or nitrogen. Typical concentrations of H in lunar basalts are $\sim 1\,\mu$ g g^{-1} or less, and typical concentrations of C and N are $<100\,\mu$ g g^{-1} and $<50\,\mu$ g g^{-1}. In contrast, terrestrial basalts commonly contain ~0.05 to 1% by weight H_2O and ~0.01 to 0.11% C (*Basaltic Volcanism Study Project* 1981, Ch. 1). The inferred low concentrations in the lunar interior are consistent with the impact scenario for lunar origin (see, e.g., Wänke and Dreibus 1986). Even if these elements were initially present within the Moon, on exhalation from the interior they could have been ionized and swept into space by the magnetic fields of the solar wind (see, e.g., Lindeman et al. 1973; Manka and Michel 1973).

A more productive approach to assessing the Moon's water potential than a direct search for water may be to consider the constituents of water, H and O. Hydrogen ions are the principal constituent of the solar wind. The solar-wind plasma, traveling outward from the Sun at speeds of the order of km s^{-1}, strikes the lunar surface without significant impedance from a lunar atmosphere or significant repulsion by the Moon's weak, local magnetic fields. The ions imbed themselves in soil grains to depths of a few atomic diameters. The soils have been so well stirred by meteoroid impacts to a depth of at least 2 to 3 m for such long times that most grains have had substantial exposure to the solar wind. Hydrogen and the other SWIE can be extracted by heating the soil to ~700° C.

Figure 1 is a histogram of H concentrations in mare materials. (Note that for this and all histograms in this chapter, statistics depend on the unfortunately biased combination of what was sampled, what was analyzed, and what was collected into the database.) Soils and regolith breccias typically contain H atoms in bulk concentrations of $\sim 50\,\mu$ g g^{-1} (see, for recent reviews Carter 1985; Haskin 1992; L. A. Taylor 1990*a*). By Earth standards, this is a low concentration; pure water contains 11% by weight H, some 2200 times more than lunar soil. Nevertheless, the concentration of H in the lunar soil is the equivalent of ~0.5 liter of water per m^3, or $\sim 10^6$ gallons per square mile to a depth of ~2 yards (Haskin 1992). As the most abundant element in the soil is oxygen (~45%), the constituents of water are plentiful on the Moon. Whether they are economically valuable depends on need and the relative costs of furnishing them from lunar and other sources. Such mining would be large scale and would require an extensive infrastructure.

Next to H, He is the most abundant element in the solar wind. Solar-wind He has a much higher proportion of ^{3}He (atomic ratio $^3He/^4He \sim 4.8 \times 10^{-4}$) than terrestrial He (atomic ratio $^3He/^4He = 1.4 \times 10^{-6}$); on Earth, most He in the atmosphere is ^{4}He from radioactive decay of U and Th series elements, and the atmosphere is shielded from the solar wind. Also, the total quantity

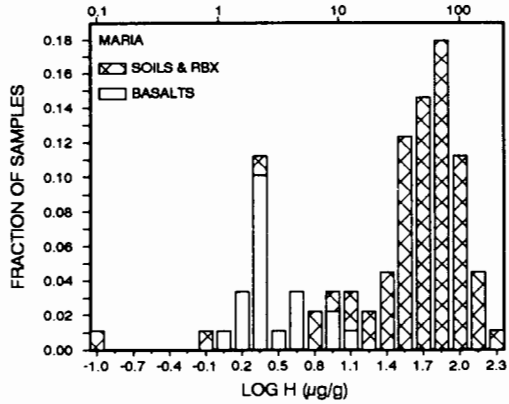


Figure 1. Abundances of H in lunar mare materials. Mare basalts have very low H concentrations; the values shown for basalts are higher than would be expected in the original lavas because cosmic-ray spallation reactions during exposure on the lunar surface have produced substantial H. Soils and regolith breccias (RBX) are enriched compared to mare basalts to the extent of $\sim 50\,\mu$ g g^{-1} H (average) because their grain surfaces contain solar-wind H. The same is true of highland soils. Soils with low H concentrations are "immature," i.e., most of their grains have not resided on the lunar surface long enough to approach saturation levels of H. Note that for all histograms in this chapter, statistics depend on the unfortunately biased combination of what was sampled, what was analyzed, and what was collected into the database.

of ^{3}He in lunar soils is much greater than in the Earth's atmosphere. For use in commercial fusion reactors, should those be successfully developed, ^{3}He has advantages over ^{3}H, the customarily considered fuel. It is not radioactive itself, and its principal fusion products with ^{2}H are ^{4}He and a proton, rather than ^{4}He and a reactor-damaging neutron. The potential for use of lunar ^{3}He in commercial reactors is under serious consideration (Wittenberg et al. 1986). Mining for lunar ^{3}He to fuel fusion reactors would automatically render extraction of H and other SWIE economical because the SWIE would be a copious by-product. The quantities of the SWIE in lunar soils are reasonably well established (Haskin and Warren 1991). The economic value of the SWIE depends on need, technology, and availability and price from alternate sources.

Abundant terrestrial experience shows the importance of ore grade to economical extraction, but speculation enters when we consider the possibility of higher grade deposits on the Moon. There are at least two possibilities for higher concentrations of SWIE elements: fumarolic activity and collisions with comets and carbonaceous meteorites. Several transient events of supposed gas emission, possibly fumarolic activity, have been reported. The Apollo surface mass spectrometer recorded a burst of ions at masses 14, 28

and 32, very tentatively interpreted as evidence for volcanic emission of perhaps N, N_2, and O_2 (Hoffman et al. 1973). Another suggested example of fumarolic activity was the observation from lunar orbit of temporal variations in ^{222}Rn emission from lunar "hot spots" (Gorenstein et al. 1973). The rest of the examples involve telescope observations (see, e.g., Middlehurst 1967; Cameron 1977). It is difficult to accept these observations optimistically as indicating sources of abundant internal H, C, or N because no clear, direct evidence for significant concentrations of these gases in the lunar interior is available from the collected samples of lunar igneous rocks. We have found no hydrous lunar minerals and no evidence of aqueous alteration of igneous minerals; minerals in 3.9 Gyr old lunar basalts are better preserved than those from the freshest terrestrial lavas. The average residual C concentration in erupted basalt does not appear to exceed $\sim 25\ \mu$ g g^{-1} (Petrowski et al. 1974). Nevertheless, most lunar basalts are vesicular, and some gas had to make the vesicles. Sato et al. (1973) and Sato (1976) suggest that CO from oxidation of graphite may have been responsible, and indicate that low total C concentrations ($<100\ \mu$ g g^{-1}) in lunar magma might suffice. Thus, although it is good to keep an open mind about the possibility of emission of SWIE from the lunar interior, there does not yet seem strong reason to expect reliable, concentrated sources associated with lunar outgassing.

Also somewhat speculative is the possibility of external sources other than the solar wind. During its lifetime, the Moon has been struck repeatedly by meteoroids and fragments of comets that would have released water and carbonaceous materials into its atmosphere. It has been proposed that residues of such bombardment may be cold-trapped in permanently shaded areas in craters at the lunar poles (Watson et al. 1975; Arnold 1979). It has also been argued that such cold-trapped materials would have been lost by sputtering and other processes (Lanzerotti and Brown 1981). If present, large, near-surface concentrations of H as water, as hydrous minerals, or as hydrocarbons could be observed by remote neutron spectroscopy. Cosmic-ray spallation-derived neutrons are captured by H nuclei. A high ore grade of, say, 0.1% water could be a factor in siting a lunar base. Whether it would change the economics sufficiently to favor a polar location relative to one nearer the lunar equator, where SWIE H and rock-bound O would be the only sources of water, remains speculative.

B. Siderophile Elements

The siderophile element (SE) group includes Fe, Co, Ni, Ge, Mo, Ru, Rh, Pd, Sb, W, Re, Os, Ir, Pt and Au. According to the collisional scenario for formation of the Moon, an Fe-rich core had already extracted SE from the silicate material that vaporized to provide the material from which the Moon formed. Indeed, concentrations of most SE in lunar materials are very low.

Siderophile behavior occurs for each SE only at low enough oxygen and sulfur fugacity for that element to be in the metallic state. The most abundant SE is Fe, but most of the Moon's Fe acts as a major element (ME) in the

2+ oxidation state. A small fraction acts as an SE where there is insufficient oxygen to keep it all oxidized. Most SE are more easily reduced than Fe. Much of the Moon's CO is also in the 2+ oxidation state, but most of its Ni was never oxidized. It is unclear whether the Moon has developed a small metallic core or other concentrations of Fe alloy that would harbor substantial quantities of the SE (Haskin and Warren 1991). Small quantities of Fe metal occur in lunar basalts, which contain Fe^{2+} as a major constituent, suggesting that some metallic Fe is present in the lunar interior; however, this metal may have formed by reduction during ascent toward eruption (Sato 1976). Were this the case, Fe metal might conceivably have been transported in the vapor phase as, for example, $Fe(CO)_5$. This gas would form at high pressure, but dissociate as it reached the lunar surface (Colson 1991). It is highly speculative, but there might be Fe metal deposits at lunar fumaroles, if they exist. Additional metallic Fe is produced within the lunar soils, apparently through reduction by solar-wind H when soils are melted on meteorite impact (Carter and McKay 1972; Housley et al. 1973,1974; Morris 1980).

The most likely source of most SE for possible practical use would be the meteoritic metal from bodies that have impacted onto the lunar surface. Much of the metallic Fe, Ni, Pt-Pd-group metals, and Au in the regolith is meteoritic in origin, and the inter-element ratios of this metal are similar to those found in chondritic meteorites. Some older regolith metals have inter-element SE ratios that differ from those of any known meteorite, and some investigators ascribe this to ancient impactors of a variety not sampled by modern meteorite falls (see, e.g., Morgan et al. 1972,1974; Anders et al. 1973; Hertogen et al. 1977; Korotev 1987*a*, *b*). Other investigators attribute these unusual inter-element ratios to modification of indigenous lunar metal by equilibration with SE from ordinary meteorites (see, e.g., Ringwood and Wänke 1990).

Figure 2 shows the distribution of Ir concentrations; Ir is a typical SE, and there is a substantial database for it for lunar materials. There are insufficient data available for a similar diagram for Ni, an element of perhaps greater resource interest. Concentrations of Ni in most highland and mare igneous rocks, although $\sim 10^6$ times their Ir concentrations, are below the detection limits of the analytical procedures in common use. The concentration level of Ni in high-Ti basalts and highland monomict rocks (HMCT; see caption to Fig. 2) is $< 10\ \mu g\ g^{-1}$; that for low-K basalts is in the range ~ 30 to $50\ \mu g\ g^{-1}$. This contrasts sharply with terrestrial basalts, many of which have Ni concentrations as high as several hundred $\mu g\ g^{-1}$. The reasons why concentrations of most SE are substantially lower in lunar basalts than in terrestrial basalts are not known. The low concentrations in lunar basalts have been taken to reflect expected values for silicate melts equilibrated with metallic Fe at low oxygen fugacities. Extraction of the SE into dense, immiscible sulfide is another possibility. Recent experimental evidence suggests that the higher SE concentrations found in terrestrial basalts could correspond to full equilibration between the Earth's mantle and core (Colson 1990). Thus, lunar silicate material may either have undergone extraction of metal and sulfide since the

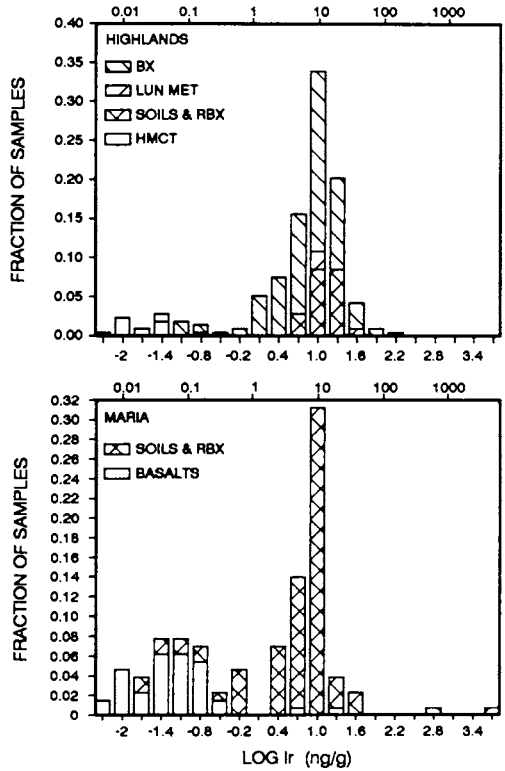


Figure 2. Histograms of Ir concentrations in highland and mare materials. Soils, regolith breccias (RBX), and other breccias (BX) that developed from igneous rocks mixed with regolith or impactor materials have relatively high Ir concentrations owing to additions of meteoritic elements. The types of fill in the histogram blocks correspond to HMCT (highland monomict rocks, believed to be original, unadulterated igneous rocks of the highland crust), soils and RBX (soils and regolith breccias, pulverized mixtures of earlier highland or mare rocks plus meteoritic material), LUN MET (lunar meteorites, which are regolith breccias from unknown regions of the Moon but regions far from the ITE-rich regions sampled by Apollo missions, and found as meteorites in the Antarctic), BX (breccias, coherent rocks consisting of fragments of earlier rocks that have undergone various extents of melting; at one extreme, a breccia may derive almost entirely from a single igneous rock formation, at the other extreme from melting of regolith), basalts (mare basalts, essentially uncontaminated by highland material or meteoritic material).

Moon formed, or the absence of a SE-rich core for the Moon may have denied the lunar silicates this source of SE as the Moon evolved.

Because of meteoritic contributions, concentrations of SE in the lunar soils substantially exceed those in the igneous rocks (see, e.g., Keays et al. 1970; Hertogen et al. 1977). Meteoritic metal makes up ~0.1% of most lunar soils (Anders et al. 1973). It can be somewhat concentrated magnetically, but not into pure form. On impact, most of an infalling meteorite vaporizes.

Small fragments of metal may survive, but gram to kilogram sized pieces might do so only rarely. Also, one of the principal constituents of mature lunar soils (those with long surface exposures, which is true of most soils) is agglutinate. Agglutinates are μm-sized particles composed of small rock and mineral fragments bonded together by copious glass produced by impact heating. The agglutinates contain much of the Fe metal that is in the soils, and separating this metal from the agglutinates would be difficult. Some types of Apollo 16 melt breccias contain 0.5 to 1% metal.

Because they accumulate meteoritic debris, the lunar soils remain the most promising source for most SE. Meteorites of $>10^6$ tonnes mass (and one >50 tonnes) have been found on Earth; such a find on the Moon could be a valuable source of SE (or of other elements such as the SWI, depending on the type of meteorite). Some SE may be locally concentrated in sulfides (e.g., FeS, troilite). Fe metal can also be obtained in unlimited quantities by reductive extraction from mare basalt. Concentrations of all other siderophile elements are low in lunar igneous rocks. Since the Moon accreted, additional separation and sinking of metallic iron may have occurred. Metal deep within the Moon's interior is metal lost from potential use.

IV. ELEMENTS CONCENTRATED MAINLY FROM INTERNAL LUNAR SOURCES

Before discussing the element groups that have been concentrated from indigenous lunar sources, it is instructive to consider their behavior during the main processes of lunar chemical differentiation.

A. The Major Elements

Magma-ocean and alternative scenarios for chemical differentiation all depend on the operation of certain geochemical processes. These processes can also operate on a local as well as a global scale. Geochemical processes that bring about partial separations of major elements (ME) from each other include fractional crystallization and partial melting (commonly with density separation of crystals from residual melt), and liquid-liquid immiscibility. Processes that recombine major elements include mixing of magmas and assimilation of solid material by a magma. Combinations of processes, such as assimilation plus fractional crystallization, or fractional crystallization plus liquid-liquid immiscibility, can cause substantial geochemical differentiation. The processes that produced the lunar crust and mantle are believed to have been equilibrium processes, at least on a local scale, and to have been controlled by temperature, pressure and ME composition.

The elements that make up most of the Earth also make up most of the Moon. These are O, Si, Al, Fe, Mg, Ca and Na, the most abundant elements that could form fairly refractory compounds in the solar nebula. On Earth, K is usually regarded as a major element, but it is rarer on the Moon. On the Moon, Ti is a major element. Except for Ti and the most volatile ME,

Na, lunar and terrestrial abundances are similar (Drake 1986). This is not particularly surprising if the Earth and Moon formed in the same region of the solar nebula. In and on the Moon, the ME mostly form the following minerals: calcic plagioclase feldspar ($CaAl_2Si_2O_8$ with some $NaAlSi_3O_8$ in solid solution), low-Ca pyroxene (a solid solution of $MgSiO_3$ and $FeSiO_3$), high-Ca pyroxene (a solid solution of $CaSiO_3$, $MgSiO_3$, and $FeSiO_3$), olivine (a solid solution of Mg_2SiO_4 and Fe_2SiO_4), and ilmenite ($FeTiO_3$).

The compositional evolution of a cooling magma ocean would be complicated (see, e.g., Longhi 1977; Longhi and Boudreau 1979). Simple fractional crystallization of magma with the composition of the early lunar mantle begins with precipitation of olivine or low-Ca pyroxene, or both. The composition of the remaining magma would evolve continuously, and the initial minerals would be joined and eventually replaced by other minerals, especially high-Ca pyroxene, ilmenite and plagioclase feldspar. Compositions of the minerals would also evolve continuously, with olivine and the pyroxenes becoming more Fe rich and Mg poor, and feldspar becoming more Na rich and Ca poor. On the Moon, Na is so low in abundance that most feldspar shows little compositional evolution. Where crystallization proceeded far enough, minor minerals would join the remaining major minerals, and the liquid might separate into two immiscible portions.

Melts continued to rise to the surface even after a thick lunar crust had formed. Batches of melt intruded into the crust, where they solidified slowly as plutonic rocks (Warren 1985*a*). Some melts spilled out onto the surface as lavas. These later melts are presumed to be partial melts from the lunar mantle. Although a magma-ocean-derived mantle in purest form would consist almost entirely of olivine or low-Ca pyroxene, at least some regions of the actual lunar mantle retained some high-Ca pyroxene, plagioclase feldspar, ilmenite and possibly garnet. The most common liquid formed by partial melting of such material has basaltic composition (similar to the compositions of the mare lavas or the noritic or gabbroic plutonic rocks of the highlands).

Once crystallized, such liquid consists mainly of feldspar and pyroxene, plus minor amounts of ilmenite and sometimes olivine. Thus, overall, the crust became enriched in Al, Ca, Na and Ti and depleted in Fe and Mg, relative to the undifferentiated Moon. Although the melts that produced the plutonic rocks of the highlands derived from mantle that still contained abundant Ca, Al and trace elements, those that produced the mare basalts derived from mantle that was partially depleted in Al, Ca, Na and trace constituents; therefore the mare basalts are enriched in Fe and, in some cases, Ti, compared with highland materials. The residues of partial melting, like the early solids from the magma ocean, would consist mainly or almost entirely of olivine and low-Ca pyroxene and would have the compositions of dunite or pyroxenite rock. The early dunite and pyroxenite produced by early magma-ocean crystallization would lie deep below the crust. Dunite found in the Apollo collection is believed to be a product of fractional crystallization of the later melts that intruded into the crust to shallower depths that could be reached by impact

cratering.

On Earth, water dissolved in melts dominates extreme magmatic fractionation. Water eventually exsolves as a vapor phase into which many trace elements partition, commonly as chloride complexes. This is the origin of the classical hydrothermal system and is the source of many economic deposits on Earth (see, e.g., Holland 1972; Burnham 1979). On the Moon, the occurrence of products of this type of extreme fractionation is highly conjectural, as no hydrous vapor is likely. Nevertheless, compositions of lunar materials indicate that extensive chemical fractionation has occurred, especially for highland materials, as is evident from Fig. 3. When crystallization proceeds far enough in a dry system, minor minerals may precipitate, and the melt may even separate into two immiscible liquids. Immiscible sulfides may form, and unmixing of "granitic" and ferroan-phosphatic melts has been suggested to account for some observed separations (see, e.g., Rutherford et al. 1976; Neal and Taylor 1989). Both processes are known to have occurred on a small scale in lunar materials. More exotic unmixed melts are conceivable (e.g., chlorides; Gillett 1990*b*) but are unlikely as they require extreme compositions. Extreme fractionation in very dry magmatic systems is also inhibited by the high viscosities of the strongly polymerized silicate melts that form as fractionation progresses; another important effect of dissolved water in terrestrial magmatic systems is to break down these polymers.

Figure 3 shows the maximum and minimum values of ME concentrations found so far in materials of the highlands and maria. Because mare lavas all have basaltic composition, their range of ME concentration is limited to less than a factor of 10, except for Ti. The range for any ME in highland materials depends on whether the element is an essential element of all common rocks, such as Si, or not essential to some, such as Mg and Ti that are almost absent in anorthosite (nearly pure plagioclase). The broad ranges in highland ME compositions reflect the presence of rocks that consist mainly of a single mineral.

All of the ME except oxygen are cations, and in most minerals they are combined with oxygen. Because of this common combination, the weights of ME in rocks are usually reported as oxides rather than elements. The highest concentrations of Al (~19%, or 36% as Al_2O_3) and Ca (~14%, or 20% as CaO) are found in anorthosites, which are rocks that consist almost entirely of feldspar. Al concentrations can be determined by remote sensing of X rays from lunar orbit (see, e.g., Adler et al. 1973). Several large (>1 kg) samples of anorthosite were obtained at Apollo 16 and one (>300 g) sample was found at Apollo 15, indicating that there are deposits of anorthosite in some regions of the highlands. Anorthosite is such an extreme of composition that we cannot expect to obtain a material richer in Al or Ca. From a resources point of view, the main problem would be to locate a body of anorthosite or a very anorthosite-rich soil. An anorthosite-rich soil (15.5% Al, or 29% as Al_2O_3; ~80% plagioclase) was found at station 11, Apollo 16. Earth-based and Galileo mission remote sensing of highland terrain, including some crater

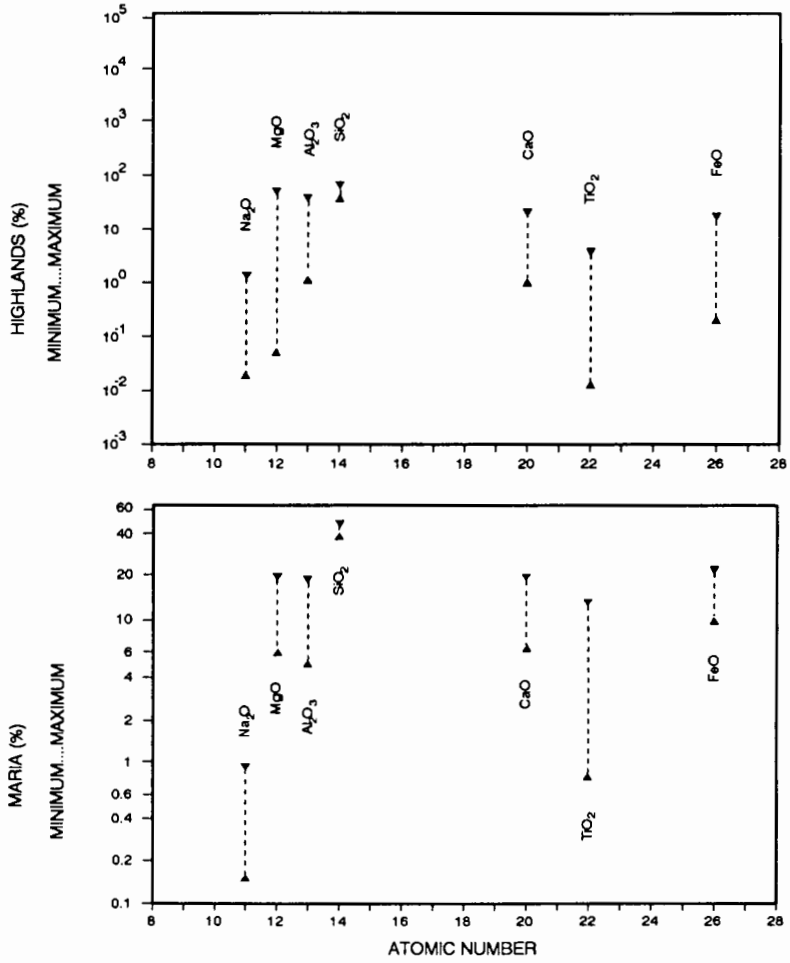


Figure 3. Graph of maximum and minimum observed major element (ME) concentrations in highland and mare materials. The major element cations are positioned along the horizontal axis in sequence of atomic number. Values for highland materials are plotted on the same vertical scale as used in Fig. 4, for comparison with the range of variation of the incompatible trace elements. All values are for the elements expressed here in wt% as oxides, according to petrologic convention. The highest and lowest practical concentrations for ME are found in the major rock-forming minerals, which can in principle be separated from their host rocks. Concentrated natural deposits of individual minerals would be the best bet for ME ores; alternatively, mineral separates from rocks or soils would give high concentrations for some ME. Some highland rocks are nearly monomineralic in composition. Anorthosites consist of nearly pure plagioclase feldspar and have high Ca, Al and Na concentrations, but low Fe, Mg and Ti concentrations. Dunite consists of nearly pure olivine and has a high Mg concentration, but low Ca, Al, Ti and Na concentrations. In the maria, no monomineralic rocks are known, but all mare basalts have high Fe concentrations, and some have high Ti concentrations. All soils are intermediate in concentrations to the extreme rocks, being mixtures of them and consequently of the major rock-forming minerals.

walls and central peaks where slopes are steep enough to minimize obscuring regolith, found deposits of anorthosite, but suggest that such deposits may be rare (Pieters 1986; Spudis et al. 1989; Pieters et al. 1991). Aluminum ore on Earth is bauxite, a mixture of hydrous oxides produced by intense weathering in tropical climates. Obviously, we do not expect such deposits on the Moon.

The highest concentrations of Ti and Fe are in ilmenite-rich basalts; samples with as much as ~17% Fe (22% as FeO) and ~8% Ti (13% as TiO_2) were collected at the Apollo 17 site, and similar basalts were also sampled at the Apollo 11 site. Relative concentrations of Ti have been mapped on a broad scale by reflectance methods (Pieters 1978). All known mare basalts have high concentrations of Fe (~15%, or 19% as FeO); Haskin and Warren 1991). Ilmenite concentrates are the starting point for some proposed methods for extracting oxygen from lunar material (see, e.g., Gibson and Knudsen 1985). However, obtaining ilmenite from mare lavas and lava-derived soils, where it is intimately combined with other minerals, may not be easy (see, e.g., Heiken and Vaniman 1991). Also, separating Fe from the residuum left after O_2 extraction will be difficult.

Dunites and pyroxenites are likely to have the highest Mg concentrations. Small quantities of dunite, a dense, almost purely olivine rock, were found at the Apollo 17 site. These are believed to be products of fractional crystallization in plutons at relatively shallow depths. Some crater central peaks appear to be rich in olivine, with varying proportions of (presumably) plagioclase (e.g., Copernicus Crater; Pieters 1982). On a smaller scale, we might expect to find dunite as a fractional crystallization product of particularly Mg-rich magmas in magma lakes, which might form where surface lava flowed into previously formed impact craters (see, e.g., Haskin et al. 1985). Discovering such pools will be challenging, as the surfaces may have a veneer of ordinary lava composition. Similar arguments may be made for finding Mg-rich pyroxenite, the analogous deposit of pyroxene. Ilmenite and even Fe metal might also form enriched deposits in exhumed plutons or magma lakes. It might seem that the Moon's low gravity would not favor production of deposits by settling of crystals. However, the low viscosities of the Moon's alkali-element-poor magmas more than make up for the Moon's lower gravitational potential (Taylor and Lu 1992).

The concentration of Si is nearly constant at ~21% (~45% as SiO_2) in common lunar soils. The concentration of oxygen in soils is also essentially constant at ~45%. Resource considerations will depend mainly on which soil compositions are most suitable for preferred methods of separating these elements.

On the basis of what we know from acquired lunar samples and remote sensing data, we may be confident of finding materials with high concentrations of major elements. Local vagaries in concentration of a desired element or mineral will be important. If the desired material is at all specialized, e.g., an ilmenite-rich soil as feedstock for a particular process, *in situ* ore-body verification will be necessary. This is because remote sensing techniques are

surface techniques that see only to depths of a few μm to 10 to 20 cm. The Apollo 16 site, for example, has a nearly constant surface soil composition, but more variable compositions at depths of a few centimeters (Korotev 1981).

B. The Incompatible Trace Elements

Elements that do not enter the crystal structure of a mineral as well as they enter a liquid in equilibrium with that mineral are said to be "incompatible" with that mineral; i.e., their partition coefficients between solid and melt phases have values less than one. Incompatible trace elements (ITE) on the Earth and Moon are those whose partition coefficients between the minerals olivine, pyroxene, and feldspar and their equilibrium melts are small (usually <0.01 to ~ 0.1). This large group includes the alkali elements (except that Na is abundant enough to be an essential constituent of plagioclase feldspar and may thus behave as a major element), the alkaline-earth elements (except Ca), the Rare-Earth Elements (REE: Y and the lanthanides), B, Zr, Nb, Sn, Hf, Ta, U and Th. When present in melts, many of the siderophile and vapor-mobilized elements will also act as incompatible elements.

Figure 4 shows the maximum and minimum observed concentrations for ITE in highland and mare materials. When an element is not an essential constituent of the principal minerals of a rock, the restrictions on their concentrations are eased. Thus, the spread of ITE concentrations in highland materials is 4 to 5 orders of magnitude. The spread in mare materials is closer to 1 to 2 orders of magnitude, because these materials derive largely from mare basalts, a closely knit family of rocks from a relatively restricted type of mantle source.

During partial melting of mantle material, ITE tend to concentrate into the first liquid produced. If 1% of a typical portion of mantle melts, the concentrations of ITE in the liquid may reach 100 times that of the mantle. As on Earth, however, few lavas have the compositions expected for low-degree partial melts. Most appear to have evolved from higher-degree melts by fractional crystallization. Melts rise through gradients of decreasing temperature as they approach the surface, and they may be impeded and stored for long periods along the way. They lose heat and partially crystallize. As crystals form and separate, the bulk of the ITE remain with the melt and become even further enriched in it.

Figures 5 and 6 are histograms of La and K, which serve as examples of ITE. Comparisons with terrestrial shales, as approximations of average terrestrial continental surface levels, and several suites of basalts, for comparison with the mare data, are included. Note that the La concentration ranges are comparable for both planets. This is consistent with differentiation of silicate materials of roughly comparable ME and ITE composition to form olivine or low-Ca pyroxene mantles and feldspar-enriched crusts. In contrast, K concentration levels are substantially lower on the Moon than in the roughly comparable materials on Earth. Compounds of K are more volatile than those of most ITE, and K and other alkali elements were presumably lost along with

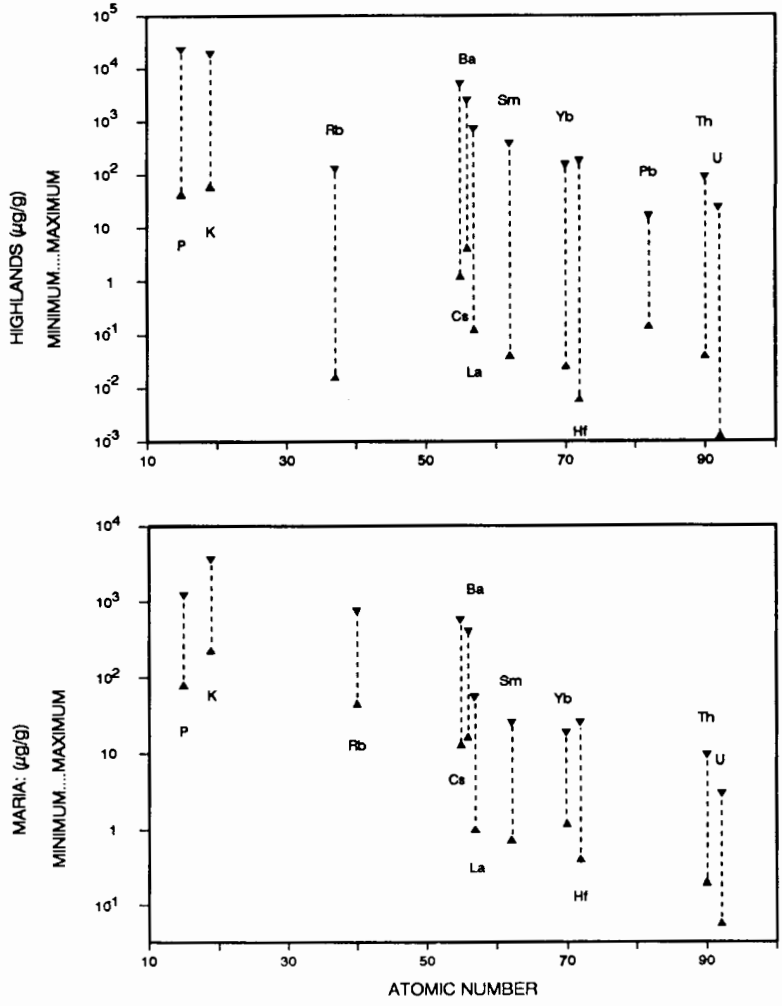


Figure 4. Graph of maximum and minimum incompatible trace element (ITE) concentrations for lunar mare and highland materials. Highest ITE concentrations are in specialized highland monomict (HMCT) rocks (felsite, P-rich alkali anorthosite, quartz monzodiorite) that are extreme products of fractional crystallization. Lowest values are for other HMCT, those that are almost monomineralic products of fractional crystallization. In contrast to the ME in HMCT, whose values of max/min seldom exceed 50, those for most ITE cover 4 to 5 orders of magnitude.

VME (vapor-mobilized elements; see below) when the Moon formed.

Some highland materials are very enriched in ITE relative to the bulk Moon. A glassy highland material called KREEP (after K, rare-earth elements, and P) was found in Apollo 12 soil (Hubbard et al. 1971). Since that time, KREEPy materials have been found at all Apollo sites. The lunar near side in the region of Mare Imbrium and its surroundings may be the Moon's region

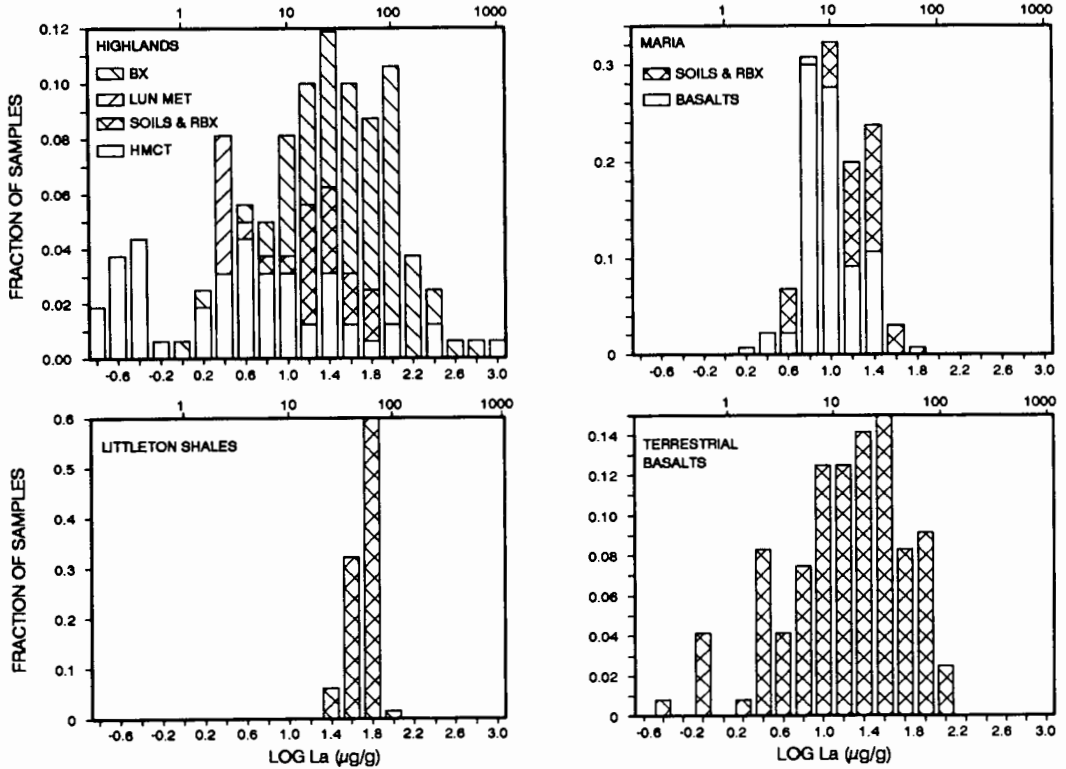


Figure 5. Histogram of La concentrations in highland materials (see caption of Fig. 2 for key), terrestrial shales of the "Littleton" formation (as typical terrestrial continental surface material, for comparison) (B. Moss, unpublished data), lunar mare basalts, and terrestrial basalts for comparison (*Basaltic Volcanism Study Project* 1981). Note the large spread in highland monomict (HMCT) concentrations, as discussed in the caption for Fig. 4. The concentration peak for LUN MET is less than that for SOILS and RBX which come from the ITE-rich near side. Concentrations for BX span nearly the same range as HMCT data because some BX are igneous rocks minimally contaminated with ITE-richer material. The lunar mare basalts have a substantial range of concentrations (although less of a range than the more thoroughly sampled terrestrial basalts). This range presumably results from differences in the ITE concentrations of mantle source regions of the different basalts as well as differences in the extent of melting of the sources and in compositional modification by processes such as fractional crystallization.

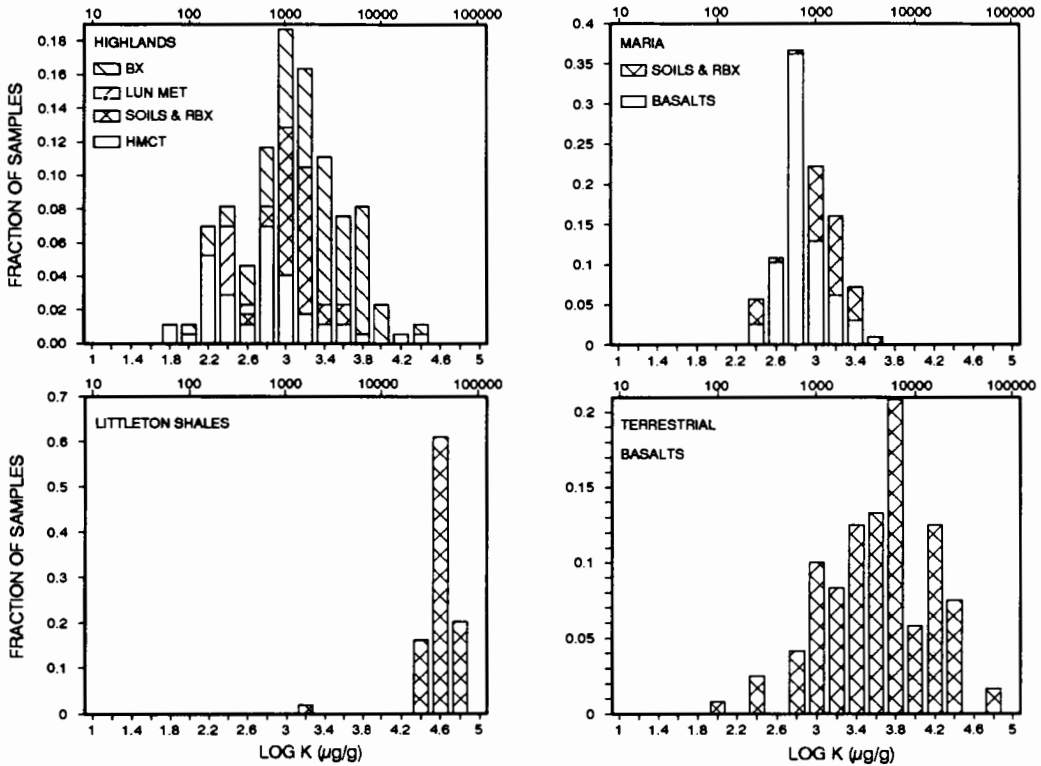


Figure 6. Histogram of K concentrations. See Fig. 5 for a general explanation. Note, however, that unlike La, the general level of K concentrations in lunar material is substantially less than in typical terrestrial materials.

of principal ITE concentration, as indicated by sampling and remote sensing of Th gamma rays (Metzger et al. 1977). Remotely sensed data from the far-side highlands indicate lower concentrations of Th, and the lunar meteorites of highland origin, which are breccias of compacted regolith of unknown provenance, have no significant KREEPy component. Although the relative abundances of ITE in all KREEPy materials are similar, subtle differences are found between KREEPy materials from different areas.

The mechanism by which the KREEPy ITE-rich materials have been brought to the lunar surface is not known. On the basis of their ME compositions, the ITE-rich components of some KREEPy materials may have been highland basaltic lavas before they become incorporated into breccias and soils. Two small samples of KREEP basalt were found at Apollo 15, and numerous fragments are in the soil. Other KREEPy material has been found as intrusive rocks. The residual liquid from crust formation, according to the magma ocean hypothesis, would be highly enriched in ITE, and processes might have brought some of it or, more likely, a derivative of this primordial ITE-rich melt, to the surface (urKREEP; Warren 1988).

The highest ITE concentrations found in lunar materials do not have the relative ITE abundances of KREEP. They appear to have developed through extensive fractional crystallization of KREEPy material. A recent study of highly ITE enriched materials from Apollo 14 soil particles concluded that crystallization of plagioclase and pyroxene first enriched a residual liquid in ITE concentrations. Then whitlockite (Ca phosphate) and zircon (Zr silicate) precipitated along with the plagioclase and pyroxene, producing crystals rich in the ITE Zr, Hf and REE, leaving a more evolved residual liquid rich in the ITE Rb, Cs, Ba, Ta, Th and U. The solid product, quartz monzodiorite, has high concentrations of Zr (up to 4240 μg g^{-1}) and La (up to 696 μg g^{-1}). The liquid product then split into two immiscible liquids, although the two did not separate macroscopically from each other (Jolliff 1991). The pair of liquids solidified together to yield a lunar felsite, with K-rich feldspar and high concentrations of Ba (up to 2290 μg g^{-1}). It has been suggested (Neal and Taylor 1989) that large-scale unmixing of felsitic and ferroan-phosphatic melts such as these may have occurred deep within the lunar crust as the final phase of KREEP differentiation.

C. The Minor Elements

The minor elements (MIE) as defined in *The Lunar Sourcebook* (Haskin and Warren 1991, Ch. 8) include P, Sc, V, Cr, Mn, Ga and Sr. Those of probable economic interest include P, Mn and Cr. Different MIE behaved differently as the Moon differentiated into mantle and crust. Some (P, Cr, Ga and Sr) were incompatible during silicate differentiation to form the mantle and crust and acted almost like incompatible trace elements (ITE). Others, compatible with pyroxene, (Sc, V and Mn) became less concentrated in the crust. Geochemical behavior depends, of course, on oxidation state; in oxidized form, the SE Co is compatible with olivine and pyroxene and behaves as a MIE.

The bulk of the P in highland materials occurs as the trace mineral schreibersite, Fe_3P, according to Hunter and Taylor (1981). The highest concentrations of P are associated with the highest ITE concentrations. These rocks contain accumulated whitlockite (or merrillite), $Ca_3(PO_4)_2$, or, less frequently, apatite, mainly fluorapatite, $Ca_5(PO_4)_3F$. Concentrations of 0.1 to 0.3% P are common in KREEPy ITE-rich breccias. The samples richest in P are small (50 mg) fragments of alkali anorthosite and quartz monzodiorite that contain >2% P (Haskin et al. 1973; Jolliff 1991). Large samples of such P-rich material were not collected, but this only indicates that such materials are relatively rare, and does not signify that no large deposits formed. The association of substantial deposits of apatite (and ilmenite) with anorthosite rocks is known on Earth (see, e.g., Kolker 1982), and unmixing of felsitic and ferroan-phosphatic liquids is one suggestion to account for this (Ryder et al. 1975).

Figure 7 is a histogram of P abundances in lunar and terrestrial materials. The highest lunar P concentrations overlap the most common terrestrial concentrations. The most common value for lunar highland materials is only slightly less than the most common value for the Littleton shales. However, the lunar database contains mainly analytical data for Apollo materials, which are from the KREEP-rich lunar near side. Most HMCT have very low P concentrations; exceptions are the ITE-rich alkali anorthosite and quartz monzodiorite discussed earlier. The typical concentration in the lunar highlands is probably about half an order of magnitude lower than the peak of the histogram would suggest. There is a difference of about that much between mare basalts and the most common terrestrial basalts. Nevertheless, the general levels of P in the lunar and terrestrial crusts are similar.

Concentrations of Cr are about an order of magnitude higher in lunar basalts than in common terrestrial ones. Because the interior of the Moon is more reduced than that of the Earth, lunar Cr is present mainly as Cr^{2+}. On Earth, garnet and aluminous Cr(III) spinel apparently retain Cr as Cr^{3+} in the mantle; this does not occur on the Moon (see, e.g., Schreiber and Haskin 1976). Also, the partition coefficient for Cr^{2+} in olivine is low (~0.6–0.9), and only a little higher in the pyroxenes (both Ca-poor and Ca-rich pyroxenes have partition coefficients for Cr^{2+} of ~1; see, e.g., Schreiber and Haskin 1976). Some highland plutonic rocks contain minor amounts of spinel (in this case, the Fe-Mg-Al spinel pleonaste) with ~3 to 6% Cr, at least partly as Cr^{3+} (Weiblen et al. 1974). Perhaps local conditions may have led to concentrated deposits of Cr in the lunar crust, perhaps even chromite ($FeCrO_4$). It is speculated that the oxidation of Cr^{2+} to Cr^{3+} may have reduced small amounts of Fe^{2+} to the metallic state.

Figure 8 is a histogram of Cr concentrations in lunar and terrestrial materials. Concentrations of Cr are higher in lunar materials than in their approximate terrestrial counterparts, in line with the probable dearth of chromium spinel in the Moon's mantle.

Concentrations of Mn in lunar rocks reach ~0.25%, but most Mn con-

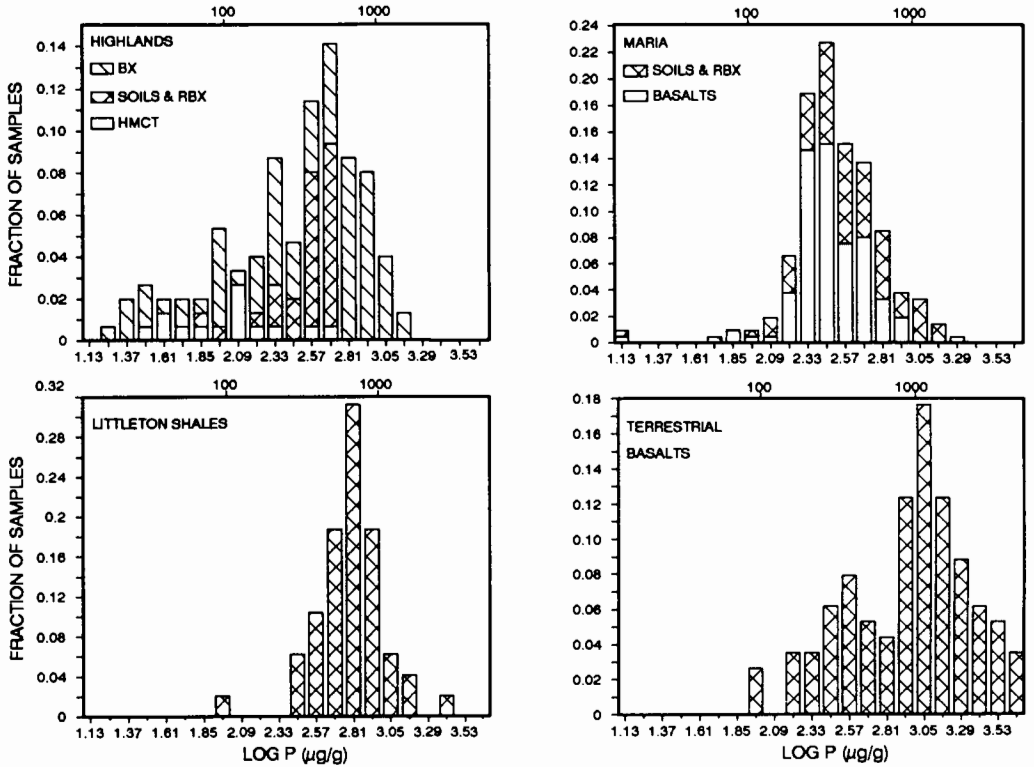


Figure 7. Histogram of P concentrations. See Fig. 2 for key. High P concentrations tend to accompany high ITE concentrations. Most P in the highlands is apparently in the mineral schreibersite (Fe_3P), but the highest known P concentrations are in phosphate-mineral-bearing alkali anorthosites and quartz monzodiorites, extreme products of fractional crystallization. Concentrations of P in the ITE-rich highland materials of the lunar near side are nearly as high as P concentrations for the Littleton shales. Those for lunar mare basalts are lower than found in terrestrial basalts.

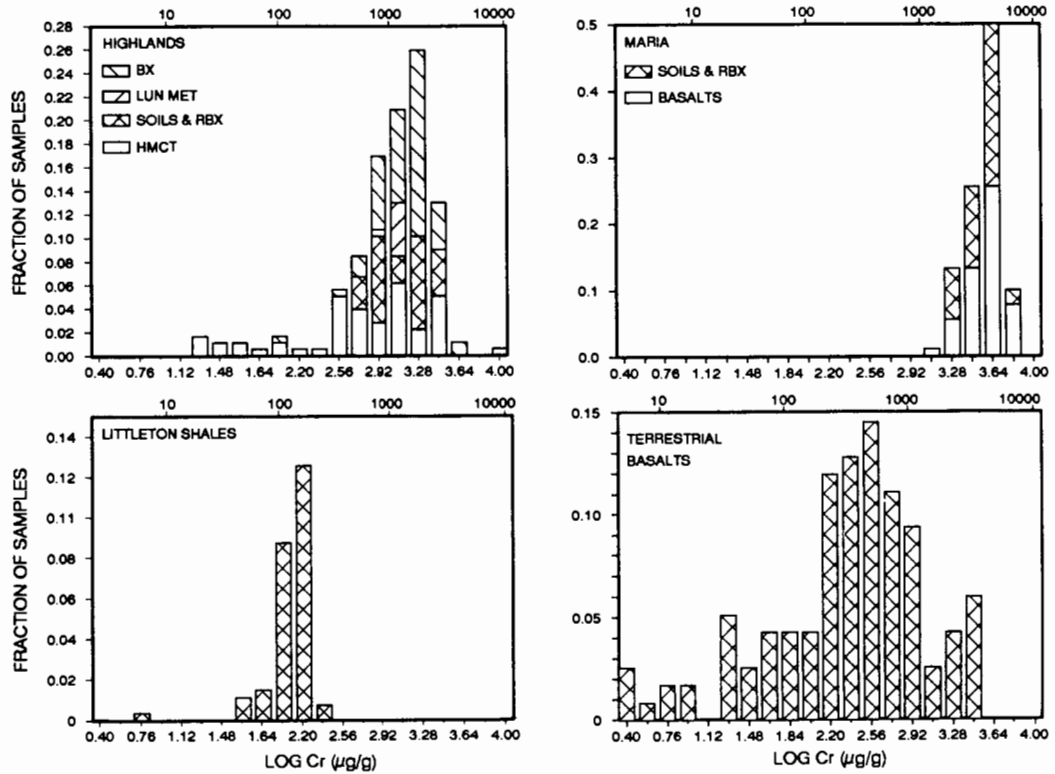


Figure 8. Histogram of Cr concentrations. See Fig. 2 for key. Cr concentrations in lunar materials are high compared with typical materials of the Earth's crust, presumably because spinel retains Cr in the Earth's mantle but not in the Moon's. Note in particular that Cr concentrations in mare basalts are $\sim$ 10 times higher than those in common terrestrial basalts.

centrations are closely related to those of Fe^{2+}, and so far no independent Mn minerals have been found. The lunar Fe/Mn ratio remains approximately constant at a value of ~80 (Laul et al. 1972). On Earth, Mn is oxidized to the 4+ oxidation state and becomes concentrated in the sedimentary cycle with Fe^{3+}. On the Moon, this oxidation does not occur, and there is no equivalent mechanism for separating Mn from the ME. We expect the highest Mn concentrations in the Apollo samples would be in ilmenite in mare basalts, and may not exceed ~0.5%. Accumulations of ilmenite in lava ponds would be a richer source of Mn than basalts.

Similarly, lunar Sc and V are dispersed at low concentration in ilmenite and clinopyroxene of mare basalts, and Ga and Sr are dispersed into lunar feldspar. Fractional crystallization does not substantially concentrate any of these four elements and, so far, high concentrations have not been observed. On Earth, V is mobilized and separated by oxidation and aqueous transport.

V. ELEMENTS OF COMBINED INTERNAL AND EXTERNAL ORIGIN: THE VAPOR-MOBILIZED ELEMENTS

The vapor-mobilized elements (VME) include S, the halogens, Cu, Zn, As, Se, Ag, Cd, In, Te, Hg, Tl, Pb and Bi. The group includes many elements valued for their industrial use. The elements of this group or their compounds vaporize at relatively low temperatures (a few hundred °C) and many of them occur in meteorites and terrestrial ores as sulfides. The VME behave as incompatible in most melts; F probably behaves only as an ITE, but is discussed here with the other halogens. The VME occur in very low concentrations in lunar basalts, our best probes of the composition of the lunar interior.

Figure 9 is a histogram of lunar Cd abundances. The materials with the highest Cd concentrations are soils and regolith breccias, and concentration levels are roughly the same for both mare and highland soils, consistent with a meteoritic origin for much of the regolith Cd. Most mare basalts contain an order of magnitude less Cd than soils. The data for highland monomict (HMCT) rocks are more scattered.

As was the case for the SWIE, the overall lunar VME concentrations are lower than typical terrestrial concentrations, and presumably for the same reason. As was the case for the SE , the bulk of these elements in the lunar soils probably comes from impacting carbonaceous meteorites. The relative abundances of the VME in the lunar soils only roughly resemble those of the carbonaceous C1 meteorites, which contain significant quantities of VME and are suspected to be the most abundant meteorite type, but are much more similar to those of the C1 meteorites than to those of any other known material.

There are at least two reasons why the relative abundances may differ from those of C1 meteorites even if the VME relative abundances of the contributing impactors closely resemble those of C1 meteorites. The first is possible additions to the regolith of VME from the lunar interior, which may

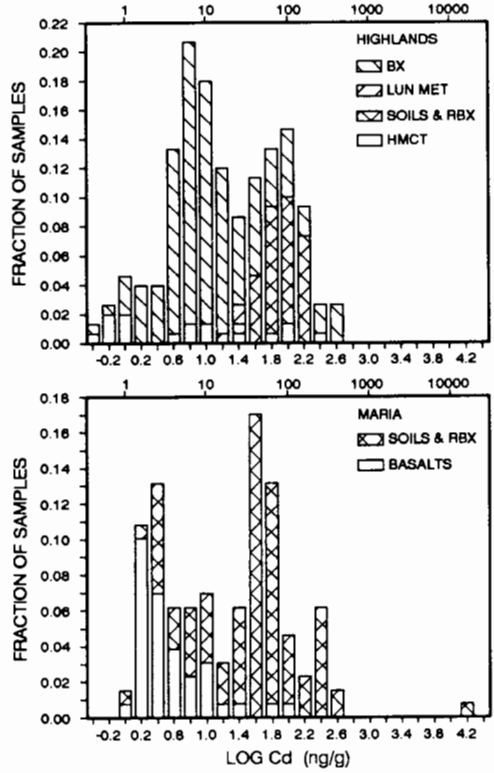


Figure 9. Histogram of Cd concentrations. See Fig. 2 for key. Note the low Cd concentrations in the highland monomict and mare basalts, reflecting the low concentrations indigenous to the Moon. Concentrations in SOILS and RBX and some BX are higher because of meteoritic additions to the regolith. Note the spread of nearly 6 orders of magnitude, and the very high concentrations in a few samples of soil. Especially high concentrations of Cd, Ag, and Bi were found in the core tube from Apollo 12, and are believed not to be inadvertent contamination by terrestrial material (Laul et al. 1971). This might be a result of vapor mobilization and concentration within the regolith.

not resemble C1 meteorites in VME relative abundances. The second is lunar chemical fractionation among the members of the group.

There is good evidence that VME are present in the lunar interior. Members of the group are associated with lunar volcanic ash. Our main samples of such ash are green glass spherules from Apollo 15 and orange glass spherules from Apollo 17, which are almost surely of igneous origin from deep within the lunar mantle (see, e.g., Delano 1979). These glasses are coated with S, Zn, Cd, Pb and presumably other VME. It is believed that these coatings formed when vaporized VME condensed onto the spherules as the erupted cloud of

gas and ash cooled before falling to the Moon's surface (see, e.g., McKay et al. 1973). Concentrations of VME in the coatings are not well characterized but are high enough that the bulk concentrations in the spherules are among the highest for lunar materials (Morgan et al. 1974; Haskin and Warren 1991). Some breccias acquired during the Apollo 16 mission were found to have partial coatings of rusty material when examined on Earth. These coatings (akaganeite, a hydrous ferric oxide) apparently resulted from reaction of lawrencite, $FeCl_2$, when exposed to terrestrial or spacecraft water and air (see, e.g., L. A. Taylor et al. 1973). Presumably, $FeCl_2$ vapor had permeated the breccias and condensed in cracks. One of the consequences of impacts into the lunar crust is heating of material ejected from craters and heating of material beneath craters (see, e.g., Bratt et al. 1985). Blankets of ejecta on crater rims may be largely unconsolidated and porous. They are excellent insulators, so temperatures may remain high enough (several hundred °C) and for long enough (years) to mobilize halides and other VME compounds. At high enough pressures and low enough temperatures, crystallization of VME compounds may take place from a vapor state. Colson (1991) has shown that a gas similar to terrestrial fumarolic gas, with $pCl_2 = 4 \times 10^{-9}$ atm and in equilibrium with basaltic melt, could carry sufficient $FeCl_2$ for significant mass transport (see also Chapter by Fegley and Swindle). Such a gas, on cooling to ~920 to 950°C, would precipitate most of its $FeCl_2$.

Lead is also moved as vapor in the regolith. Some samples contain more Pb than their parent U and Th could have furnished, whereas others contain less (see, e.g., Silver 1970; Nunes and Tatsumoto 1973). There is evidence that Hg evaporates from surface soils to a depth of ~10 cm owing to the heating of the soil during the lunar day (Jovanovic and Reed 1979; Reed and Jovanovic 1979). Concentration ratios among VME are not constant in lunar regolith materials, suggesting some mechanism for their chemical fractionation, and vaporization and condensation may be the main process. Some fluctuation may stem from contributions from internal sources, especially for Zn. It is difficult to assess the importance of contributions from the lunar interior. Volcanic emissions of VME would not necessarily cause those elements to be lost to space, in contrast to the case for the far more volatile SWIE. The VME probably behave like ITE elements during igneous differentiation, up to the point where temperature and pressure conditions allow them to exsolve as vapor or perhaps form an immiscible sulfide melt.

Sulfur seems surprisingly abundant in lunar surface materials, considering that other VME are so low in general lunar abundance. For instance, lunar basalts contain ~0.05 to 0.25% and typical terrestrial basalts range from ~0.06 to 0.1% (see, e.g., Gibson et al. 1975). Sulfur also moves within breccias. In at least one sample, blebs and veins of FeS (troilite) are observed (Norman 1981; Lindstrom and Salpas 1983). The troilite appears to result from reaction of some form of S, perhaps S_2 vapor, with the Fe component of olivine, producing FeS and pyroxene. How extensive such S metasomatism may be, and whether sulfurous or troilitic liquids extract and concentrate other

VME is not known. The appearance of such troilite may involve reaction of S_2 gas with the olivine to produce O_2, in which case it would be sensitive to local oxygen pressure (Colson 1991). In any event, the process has been observed on a small scale and offers the possibility of concentrated ores of S and perhaps other VME. Alternatively, a troilite-rich liquid might exsolve and extract other VME into it. Troilite is a dense phase that might be concentrated by gravitational as well as vapor-transport mechanisms. Sulfide immiscibility has led to ores on Earth (see, e.g., Naldrett and Macdonald 1979). Figure 10 is a histogram of S concentrations in lunar materials. Mare basalts have slightly higher concentrations than highland materials. Most samples fall within an order of magnitude of each other, similar to the pattern of the ME.

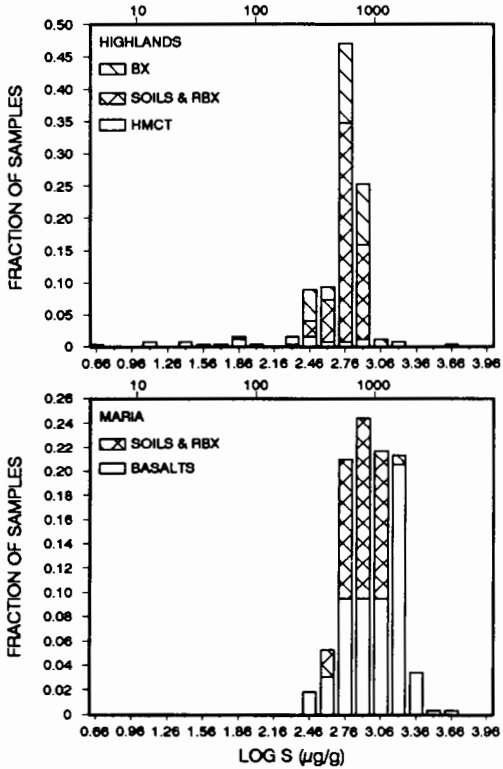


Figure 10. Histogram of S abundances. See Fig. 2 for key. Concentrations of S in mare basalts are in the range found for terrestrial basalts, suggesting the presence of substantial S in the Moon's mantle. Blebs and veins of troilite (FeS) have been observed in at least one breccia.

VI. CONCLUSIONS

As long as only small quantities of material are needed in space, Earth will

remain the sole sensible source. As soon as relatively unspecialized materials in quantities of hundreds of tonnes per year are needed in space, lunar sources have improved chances of being competitive. Any major, non-human rated structures such as solar power stations would require such material; so would radiation shielding of human-rated structures outside of low-Earth orbit. Lunar oxygen and hydrogen for fuel could become the standard source of energy for moving materials in near-Earth space and beyond. The specific locations of major structures also affect the economics of materials in space. If a facility is to be used on the surface of the Moon, use of local material saves both its energy of liftoff from the Earth and braking on the Moon. Partly offsetting the potential saving of transportation cost is the cost of developing processes for extracting and using local material and hauling the necessary equipment from the Earth. Exploration for ore bodies will be an additional cost, although such exploration can largely overlap with scientific investigations.

Eventually, more specialized materials may be desired. Geochemical assessment of the Moon's resources is barely underway. The first iteration of sampling in the context of global remote sensing analysis cannot yet take place because there still has been no global remote sensing. Nor has there been adequate, detailed, theoretical consideration of the data we have so far, particularly in the context of ore formation. Innovations in extraterrestrial mining, processing, and manufacture have not received sufficient attention and, particularly, laboratory-scale testing and development. Lacking this knowledge, economic analysis is speculative, except perhaps for processing of the most common soils.

We are still learning about the Moon from the materials in hand—the Apollo and Luna samples and the Antarctic meteorites. We need global geochemical maps, a better knowledge of the nature of lunar crust, and a more detailed understanding of lunar stratigraphy and regolith formation if we are to predict and assess ore potential accurately. This is an iterative process that requires orbital remote sensing, detailed exploration from an outpost, and additional sample collection and analysis. Geophysical exploration could also yield an unexpected boon, in the discovery of large meteorite bodies rich in the SE or SWIE that are otherwise dispersed in the lunar surface.

Global mapping and sample collection and analysis alone are not adequate to characterize the economic potential of the Moon accurately. Not evident in the economic equation are the dawning and development of ideas that might lead to practical use of the Moon. There is probably no efficient means of introducing these into economic analysis other than our gaining practice in living in space and on the Moon. Investment in this is a gamble; we cannot be certain what rate of return it will bring. We can be fairly certain that the return will be low, or nil, if we do not make a substantial investment, however. We need practice at living and working on the lunar surface, and part of the activity there should be demonstrating the use of local resources. Such demonstration requires prior development and testing of appropriate processes on Earth. Knowledge of suitable processes enables more intelligent

assessment of lunar ore potential. We do not know what specialized materials the Moon may offer, but we already know that the Moon's resources could be valuable in support of a lunar base or settlement.

REFERENCES

Adler, I., Trombka, J. I., Schmadebeck, R., Lowman, P., Blodget, H., Yin, L., Eller, E., Podwysocki, J. R., Wiedner, J. R., Bickel, A. L., Lum, T. K. L., Gerard, J., Gorenstein, P., Bjorkholm, P., and Harris, B. 1973. Results of Apollo 15 and 16 X-ray experiment. *Proc. Lunar Sci. Conf.* 4:2783–2791.

Anders, E. 1977. Chemical compositions of the Moon, Earth, and eucrite parent body. *Phil. Trans. Roy. Soc. London* 285:23–40.

Anders, E., Ganapathy, R., Krähenbühl, U., and Morgan, J. W. 1973. Meteoritic material on the Moon. *The Moon* 8:3–24.

Arnold, J. R. 1979. Ice in the lunar polar regions. *J. Geophys. Res.* 84:5659–5668.

Basaltic Volcanism Study Project. 1981 (New York: Pergamon).

Bratt, S. R., Solomon, S. C., and Head, J. W. 1985. The evolution of impact basins: Cooling, subsidence, and thermal stress. *J. Geophys. Res.* 90:12415–12433.

Burnham, C. W. 1979. Magmas and hydrothermal fluids. In *Geochemistry of Hydrothermal Ore Deposits*, ed. H. L. Barnes (New York: Wiley), pp. 71–136.

Cameron, W. S. 1977. Lunar transient phenomena (LTP): manifestations, site distribution, correlations, and possible causes. *PEPI* 14:194–216.

Carr, M., and Wänke, H. 1991. Water on Mars. *Lunar Planet. Sci.* XXII:181–182 (abstract).

Carter, J. L. 1985. Lunar regolith fines: A source of hydrogen. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 571–581.

Carter, J. L., and McKay, D. S. 1972. Metallic mounds produced by reduction of material of simulated lunar composition and implications on the origin of metallic mounds on lunar glasses. *Proc. Lunar Sci. Conf.* 3:953–970.

Colson, R. O. 1990. Solubility of Ni° in silicate melts and implications for metal/melt and crystal/melt partitioning. *Geol. Soc. of Amer. Annual Meeting Abstracts with Programs* 22:A164.

Colson, R. O. 1991. Mineralization on the Moon? Theoretical considerations of Apollo 16 "rusty rocks," sulfide replacement in 67016, and surface-correlated volatiles on lunar volcanic glass. *Proc. Lunar Planet. Sci. Conf.* 22:427–436.

Delano, J. W. 1979. Apollo 15 green glass: Chemistry and possible origin. *Proc. Lunar Planet. Sci. Conf.* 10:275–300.

Drake, M. J. 1986. Is lunar bulk material similar to Earth's mantle? In *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips and G. J. Taylor (Houston: Lunar and Planetary Inst.), pp. 105–124.

Gibson, E. K., Jr., Chang, S., Lennon, K., Moore, G. W., and Pearce, G. W. 1975. Sulfur abundances and distributions in mare basalts and their source regions. *Proc. Lunar Sci. Conf.* 6:1287–1301.

Gibson, M. A., and Knudsen, C. W. 1985. Lunar oxygen production from ilmenite. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 543–550.

Gillett, S. L. 1983. Lunar ores. In *Space Manufacturing 1983*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 277–296.

Gillett, S. L. 1990*a*. Lunar ores from magmatic processes: A speculative assessment. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 88–97.

Gillett, S. L. 1990*b*. Chloride lavas: A possible magmatic differentiation product under anhydrous conditions? *Lunar Sci. Conf.* XXI:417–418 (abstract).

Gillett, S. L. 1991. Lunar resources: Thoughts of an economic geologist. *Space Power* 10:3–17.

Gorenstein, P., Golub, G., and Bjorkholm, P. J. 1973. Spatial features and temporal variability in the emission of radon from the moon: An interpretation of results from the alpha particle spectrometer. *Proc. Lunar Sci. Conf.* 4:2803–2809.

Hartmann, W. K. 1986. Moon origin: The impact trigger hypothesis. In *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips and G. J. Taylor (Houston: Lunar and Planetary Inst.), pp. 579–608.

Hartmann, W. K., and Davis, D. T. 1975. Satellite-sized planetesimals and the lunar origin. *Icarus* 24:504–515.

Haskin, L. A. 1983. Material resources of the Moon. *Lunar Planet. Sci.* XIV:23–24 (abstract).

Haskin, L. A. 1984. The resources of space. *Planetary Report* 4:16–17.

Haskin, L. A. 1985. Toward a spartan scenario for use of lunar materials. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 435–443.

Haskin, L. A. 1992. Water and cheese from the lunar desert: Abundances and accessibility of H, C, and N on the Moon. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, April 5–7, 1988, Houston, Tex., ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 393–396.

Haskin, L. A., Helmke, P. A., Blanchard, D. P., Jacobs, J. W., and Telander, K. 1973. Major and trace element abundances in samples from the lunar highlands. *Proc. Lunar Sci. Conf.* 4:1275–1296.

Haskin, L. A., Korotev, R. L., Lindstrom, D. J., and Lindstrom, M. M. 1985. Geochemical and petrological sampling and studies at the first moon base. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.) pp. 199–210.

Haskin, L. A., and Warren, P. H. 1991. Chemistry. In *The Lunar Sourcebook*, eds. G. Heiken, D. Vaniman and B. French (Cambridge: Cambridge Univ. Press), pp. 357–474.

Heiken, G. H., and Vaniman, D. T. 1991. Characterization of lunar ilmenite sources. *Proc. Lunar Planet. Sci. Conf.* 20:239–247.

Heiken, G. H., Vaniman, D. T., and French, B. M., eds. 1991. *The Lunar Sourcebook* (Cambridge: Cambridge Univ. Press).

Hertogen, J., Janssens, M.-J., Takahashi, H., Palme, H., and Anders, E. 1977. Lunar basins and craters: Evidence for systematic compositional changes of bombarding population. *Proc. Lunar Sci. Conf.* 8:17–45.

Hoffman, J. H., Hodges, R. R., Jr., Johnson, F. S., and Evans, D. E. 1973. Lunar atmospheric composition results from Apollo 17. *Proc. Lunar Sci. Conf.* 4:2865–2875.

Holland, H. D. 1972. Granites, solutions, and base metal deposits. *Econ. Geol.* 67:281–301.

Housley, R. M., Grant, R. W., and Paton, N. E. 1973. Origin and characteristics of excess Fe metal in lunar glass welded aggregates. *Proc. Lunar Sci. Conf.* 4:2737–2749.

Housley, R. M., Cirlin, E. H., Paton, N. E., and Goldberg, I. B. 1974. Solar wind and micrometeorite alteration of the lunar regolith. *Proc. Lunar Sci. Conf.* 5:2623–2642.

Hubbard, N. J., Meyer, C., Jr., Gast, P. W., and Wiesmann, H. 1971. The composition and derivation of Apollo 12 soils. *Earth Planet. Sci. Lett.* 10:341–350.

Hunter, R. H., and Taylor, L. A. 1981. Rust and schreibersite in Apollo 16 highland rocks: Manifestations of volatile-element mobility. *Proc. Lunar Planet. Sci. Conf.* 12:253–259.

Jolliff, B. L. 1991. Fragments of quartz monzodiorite and felsite in Apollo 14 soil particles. *Proc. Lunar Planet. Sci. Conf.* 21:101–118.

Jovanovic, S., and Reed, G. W., Jr. 1979. Regolith layering processes based on studies of low-temperature volatile elements in Apollo core samples. *Proc. Lunar Planet. Sci. Conf.* 10:1425–1435.

Keays R. R., Ganapathy, R., Laul, J. C., Anders, A., Herzog, G. F., and Jeffrey, P. M. 1970. Trace elements and radioactivity in lunar rocks: Implications for meteorite infall, solar-wind flux, and formation conditions of Moon. *Science* 167:490–493.

Kolker, A. 1982. Mineralogy and geochemistry of Fe-Ti oxide and apatite (nelsonite) deposits and evaluation of the liquid immiscibility hypothesis. *Econ. Geol.* 77:1146–1158.

Korotev, R. L. 1981. Compositional trends in Apollo 16 soils. *Proc. Lunar Planet. Sci. Conf.* 12:577–605.

Korotev, R. L. 1987*a*. The meteoritic component of Apollo 16 noritic impact melt breccias. *Proc. Lunar Planet. Sci. Conf.* 17, *J. Geophys. Res. Suppl.* 91:E491–E512.

Korotev, R. L. 1987*b*. The nature of the meteoritic component of Apollo 16 soil, as inferred from correlations of iron, cobalt, iridium, and gold with nickel. *Proc. Lunar Planet. Sci. Conf.* 17, *J. Geophys. Res. Suppl.* 91:E447–E461.

Korotev, R. L., and Haskin, L. A. 1988. Europium mass balance in polymict samples and implications for plutonic rocks of the lunar crust. *Revs. Mineral.* 21:227–258.

Lanzerotti, L. J., and Brown, W. L. 1981. Ice in the polar regions of the moon. *J. Geophys. Res.* 86:3949–3950.

Larimer, J. W. 1986. Nebular chemistry and theories of lunar origin. In *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips and G. J. Taylor (Houston: Lunar and Planetary Inst.), pp. 145–171.

Laul, J. C., Morgan, J. W., Ganapathy, R., and Anders, E. 1971. Meteoritic material in lunar samples: Characterization from trace elements. *Proc. Lunar Sci. Conf.* 2:1139–1158.

Laul, J. C., Wakita, H., Showalter, D. L., Boynton, W. V., and Schmitt, R. A. 1972. Bulk, rare earth, and other trace elements in Apollo 14 and 15 and Lunar 16 samples. *Proc. Lunar Sci. Conf.* 3:1181–1200.

Lewis, J. S., and Lewis, R. A. 1987. *Space Resources* (New York: Columbia Univ. Press).

Lindeman, R., Freeman, J. W., Jr., and Vondrak, R. R. 1973. Ions from the lunar atmosphere. *Proc. Lunar Sci. Conf.* 4:2889–2896.

Lindstrom, M. M., and Salpas, P. A. 1983. Geochemical studies of feldspathic fragmental breccias and the nature of North Ray crater ejecta. *Proc. Lunar Planet. Sci. Conf.* 13, *J. Geophys. Res. Suppl.* 88:A671–A683.

Longhi, J. 1977. Magma oceanography 2: Chemical evolution and crustal formation. *Proc. Lunar Sci. Conf.* 8:601–612.

Longhi, J., and Boudreau, A. E. 1979. Complex igneous processes and the formation of the primitive crustal rocks. *Proc. Lunar Planet. Sci. Conf.* 10:2085–2105.

Manka, R. H., and Michel, F. C. 1973. Lunar ion spectra and surface potential. *Proc. Lunar Sci. Conf.* 4:2897–2908.

McKay, D. S., Clanton, U. S., and Ladle, G. 1973. Scanning electron microscope study of Apollo 15 green glass. *Proc. Lunar Sci. Conf.* 4:225–238.

McKay, D. S., Heiken, G. H., Basu, A., Blanford, G., Simon S., Reedy, R., French B. M., and Papike J. J. 1991. The lunar regolith. In *The Lunar Sourcebook,* eds. G. Heiken, D. Vaniman and B. French (Cambridge: Cambridge Univ. Press), pp. 285–356.

Melosh, H. J. 1989. *Impact Cratering, a Geologic Process* (Cambridge: Oxford Univ. Press).

Metzger, A. E., Haines, E. L., Parker, R. E., and Radocinski, R. G. 1977. Thorium concentrations in the lunar surface. I. Regional values and crustal content. *Proc. Lunar Sci. Conf.* 8:949–1000.

Middlehurst, B. M. 1967. An analysis of lunar events. *Rev. Geophys.* 5:173–189.

Morgan, J. W., Laul, J. C., Krähenbühl, R., Ganapathy, R., and Anders, E. 1972. Major impacts on the moon: Characterization from trace elements in Apollo 12 and 14 samples. *Proc. Lunar Planet. Conf.* 3:1377–1395.

Morgan, J. W., Ganapathy, R., Higuchi, H., Krähenbühl, U., and Anders, E. 1974. Lunar basins: Tentative characteristics of projectiles, from meteoritic elements in Apollo 17 boulders. *Proc. Lunar Sci. Conf.* 5:1703–1736.

Morris, R. V. 1980. Origin and size distribution of metallic particles in the lunar regolith. *Proc. Lunar Planet. Sci. Conf.* 11:1697–1712.

Naldrett, A. J., and Macdonald, A. J. 1980. Tectonic settings of Ni-Cu sulphide ores: Their importance in genesis and exploration. In *The Continental Crust and Its Mineral Deposits*, eds. D. W. Strangway and J. Tuzo; also published as GAC Spec. Paper 20 (Waterloo, Ont.: Geological Assoc. of Canada).

Neal, C. R., and Taylor, L. A. 1989. Metasomatic products of the lunar magma ocean: The role of KREEP dissemination. *Geochim. Cosmochim. Acta* 53:529–541.

Norman, M. D. 1981. Petrology of suevitic lunar breccia 67016. *Proc. Lunar Planet. Sci. Conf.* 12B:235–252.

Nunes, P. D., and Tatsumoto, M. 1973. Excess lead in "rusty rock" 66095 and implications for an early lunar differentiation. *Science* 182:916–920.

Petrowski, C., Kerridge, J. F., and Kaplan, I. R. 1974. Light element geochemistry of the Apollo 17 site. *Proc. Lunar Sci. Conf.* 5:1939–1948.

Pieters, C. M. 1978. Mare basalt types on the front side of the moon: A summary of spectral reflectance data. *Proc. Lunar Planet. Sci. Conf.* 9:2825–2849.

Pieters, C. M. 1982. Copernicus crater central peak: Lunar mountain of unique composition. *Science* 215:59–61.

Pieters, C. M. 1986. Composition of the lunar highlands crust from near-infrared spectroscopy. *Rev. Geophys.* 24:557–578.

Pieters, C. M., Belton, M., Becker, T., Carr, M., Chapman, C., Fanale, F., Fischer, E., Gaddis, L., Greeley, R., Greenberg, R., Hoffmann, H., Head, J., Helfenstein, P., Jaumann, R., Johnson, T. V., Klaasen, K., McEwen, A., Murchie, S., Neukum, G., Oberst, J., Pilcher, C., Plutchak, J., Pratt, S., Robinson, M., Sullivan, R., Sunshine, J., and Veverka, J. 1991. Crustal heterogeneity of the Moon viewed from the Galileo SSI camera: Lunar sample calibrations and compositional implications. *Lunar Planet. Sci.* XXII:1067–1068 (abstract).

Reed, G. W., Jr., and Jovanovic, S. 1979. Near-surface daytime thermal conductivity in the lunar regolith. *Proc. Lunar Planet. Sci. Conf.* 10:1637–1647.

Ringwood, A. E., and Wänke, H. 1990. Cobalt and nickel concentrations in the "komatiite" component of Apollo 16 polymict samples—reply to R. L. Korotev. *Earth Planet. Sci. Lett.* 96:490–498.

Rutherford, M. J., Hess, P. C., Ryerson, F. J., Campbell, H. W., and Dick, P. A. 1976. The chemistry, origin and petrogenetic implications of lunar granite and monzonite. *Proc. Lunar Sci. Conf.* 7:1723–1740.

Ryder, G. 1990. Lunar samples, lunar accretion and the early bombardment of the Moon. *EOS: Trans. AGU* 71:313–323.

Ryder, G., Malcuit, R. J., and Vogel, T. A. 1975. Differentiation of an andesine-type anorthosite: The San Gabriel Suite. *Lithos* 8:311–316.

Sagan, C. 1989. The gift of Apollo. *Parade Magazine*, July 16, pp. 4–7.

Sato, M. 1976. Oxygen fugacity and other thermochemical parameters of Apollo 17 high-Ti basalts and their implications on the reduction mechanism. *Proc. Lunar Sci. Conf.* 7:1323–1344.

Sato, M., Hickling, N. L., and McLane, J. E. 1973. Oxygen fugacity of Apollo 12, 14, and 15 lunar samples and reduced state of lunar magmas. *Proc. Lunar Sci. Conf.* 4:1061–1079.

Schreiber, H. D., and Haskin, L. A. 1976. Chromium in basalts: Experimental determination of redox states and partitioning among synthetic silicates. *Proc. Lunar Sci. Conf.* 7:1221–1259.

Silver, L. T. 1970. Uranium-thorium-lead isotopes in some Tranquillity Base samples and their implication for lunar history. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1533–1574.

Smith, J. V. 1982. Heterogeneous growth of meteorites and planets, especially the Earth and Moon. *J. Geol.* 90:1–48.

Spudis, P. D., Hawke, B. R., and Lucey, P. G. 1989. Geology and deposits of the lunar Nectaris Basin. *Proc. Lunar Planet. Sci. Conf.* 19:51–59.

Stevenson, D. J. 1987. Origin of the Moon—the collision hypothesis. *Ann. Rev. Earth Planet. Sci.* 15:271–315.

Taylor, L. A. 1990*a*. Hydrogen, helium, and other solar-wind components in lunar soil: abundances and predictions. *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 68–77.

Taylor, L. A. 1990*b*. Rocks and minerals in the regolith of the Moon: Resources for a lunar base. In *Advanced Materials: Applications of Mineral and Metallurgical Processing Principles*, ed. V. I. Lakshmanan (Littleton, Co.: Soc. for Mining, Metallurgy, and Exploration), pp. 29–47.

Taylor, L. A. and Lu, F. 1992. The formation of ore mineral deposits on the moon: A feasibility study. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 379–383.

Taylor, L. A., Mao, H. K., and Bell, P. M. 1973. "Rust" in the Apollo 16 rocks. *Proc. Lunar Sci. Conf.* 4:829–839.

Taylor, S. R. 1982. *Planetary Science: A Lunar Perspective* (Houston: Lunar and Planetary Inst.).

Taylor, S. R. 1986. The origin of the Moon: Geochemical considerations. In *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips, and G. J. Taylor (Houston: Lunar and Planetary Inst.), pp. 125–143.

Tera, F., Papanastassiou, D., and Wasserburg, G. W. 1974. Isotopic evidence for a terminal lunar cataclysm. *Earth Planet. Sci. Lett.* 22:1–21.

Wänke, H., and Dreibus, G. 1986. Geochemical evidence for formation of the moon by impact-produced fission of the proto-earth. In *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips and G. J. Taylor (Houston: Lunar and Planetary Inst.), pp. 649–672.

Warren, P. H. 1985*a*. Anorthosite assimilation and the origin of the Mg/Fe bimodality of pristine lunar rocks: Support for the magmasphere hypothesis. *Proc. Lunar Planet. Sci. Conf.* 16, *J. Geophys. Res. Suppl.* 90:D331–D343.

Warren, P. H. 1985*b*. The magma ocean concept and lunar evolution. *Ann. Rev. Earth Planet. Sci.* 13:201–240.

Warren, P. H. 1988. The origin of pristine KREEP: Effects of mixing between

UrKREEP and the magmas parental to the Mg-rich cumulates. *Proc. Lunar Planet. Sci. Conf.* 18:233–241.

Watson, K., Murray, B. C., and Brown, H. 1975. The behavior of volatiles on the lunar surface. *The Moon* 13:121.

Weiblen, P. W., Powell, B. N., and Aitken, F. K. 1974. Spinel-bearing feldspathic-lithic fragments in Apollo 16 and 17 soil samples: Clues to processes of early lunar evolution. *Proc. Lunar Sci. Conf.* 5:749–767.

Wittenberg, L. J., Sarantius, J. F., and Kulcinski, G. L. 1986. Lunar source of ^{3}He for commercial fusion. *Fusion Tech.* 10:167–178.

A REVIEW OF POSSIBLE MINING APPLICATIONS IN SPACE

PETER G. CHAMBERLAIN
U. S. Bureau of Mines, Washington Office

LAWRENCE A. TAYLOR
University of Tennessee

EGONS R. PODNIEKS
U. S. Bureau of Mines, Twin Cities Research Center

and

RUSSELL J. MILLER
Colorado School of Mines, Center for Space Mining

Successful exploration of Mars and outer space may require base stations strategically located on the Moon. Such bases must develop a certain self-sufficiency, particularly in the critical life support materials, fuel components, and construction materials. This chapter reviews technology for the first steps in lunar resource recovery—those of mining. The topic is covered in three main categories: engineering properties of lunar regolith, surface mining and excavation, and underground mining. The chapter also contains a brief discussion of *in-situ* processes. The text describes mining technology ranging from simple digging and hauling vehicles (the "strawman") to more specialized technology including underground methods. *In- situ* processes—chemical and thermal—are identified to stimulate further thinking by future researchers.

I. INTRODUCTION

A. Needs

Certain fundamental requirements can be identified for facilities constructed in space. Most of these requirements to some extent involve excavation or materials handling. Shelter, for example, is required for systems and inhabitants operating in a nonterrestrial environment. The shelter must protect the inhabitants and equipment from radiation, particle impacts, temperature variations, and atmospheric extremes. Shelters can be prefabricated on Earth and erected on site or made with local materials. Even protective structures made from imported materials will require foundations and support from

local materials. Most construction materials would be prepared from local materials more economically than imported from Earth.

For manned operations, life support materials such as oxygen and water may be provided by indigenous resources. Fuel from local energy sources may provide facility power and power for space vehicle propulsion.

The extremely high costs for transporting supplies from Earth into space drive the need for mining and processing in space. The first base stations in space will probably appear on the Moon. A permanent base on the Moon must use indigenous resources to be cost effective. A successful base must develop a certain amount of self-sufficiency, particularly in the critical life-support materials and in readily available construction materials. Site preparation, excavation, mining, and waste disposal activities will all require mining and earth-moving equipment. Therefore, mining and processing capabilities must be developed for the Moon, asteroids, Mars, and other more distant planetary bodies.

B. The Problem

Mining and processing equipment and techniques presently used on Earth will not work in space without significant changes. Although terrestrial concepts may provide a good starting point for space operations, the following problems preclude direct application of terrestrial technology:

1. Present mining and processing methods are energy intensive; large quantities of energy will be more difficult and expensive to produce in space.
2. Large labor crews used to mine on Earth will not be available in space; automation and teleoperation are critical requirements.
3. Terrestrial equipment is too heavy and bulky to be transported to the Moon.
4. The environment in space imposes severe restrictions on equipment design and presents major operational difficulties. These include: (a) dust will hamper visibility, coat lenses and mirrors, and clog moving parts. Also, the large component of soils made up of glass particles will have a very abrasive effect on friction surfaces such as joints; (b) vacuum (1 lunar atmosphere = 10^{-9} to 10^{-12} torr) will create severe lubrication problems (most lubricants are volatile and will outgas in vacuum), which will restrict component movement and prevent some mechanisms from functioning; (c) tremendous temperature fluctuations between lunar night (−170°C) and day (125°C) will greatly restrict equipment designs. Earth mining and processing equipment can not work in an environment as hot as the lunar day nor as cold as the lunar night; (d) radiation hazards and micrometeorite bombardment (velocities up to 200,000 km hr^{-1}) dictate that equipment, electronic components, and operators must be rigorously protected; automation, teleoperation and computer assistance are paramount to minimize operator exposure; (e) low gravity (160 cm s^{-2}) presents stability and traction problems that must be overcome.

5. New maintenance concepts must be conceived; replacement parts transported from the Earth or made from indigenous materials would be costly and would require a very long lead time. Replacing parts will be considerably more difficult on the Moon than on Earth.
6. Little or no geotechnical engineering data exists for potential mining and processing sites. Technology development must be closely linked to data gathering activities such as a lunar orbiter, robotic or teleoperated surface exploration, and manned visits to the Moon.

C. Solution: New Mining Technology

Clearly, the Nation will need new technology to exploit lunar and other space resources. Any research and development program leading to use of local resources on the Moon or other planetary bodies must fit the overall timetable established for Space Exploration Initiative mission objectives. Such a comprehensive program of research and development should start with a broad consideration of concepts and narrow them to a few candidates based on rigorous scientific and engineering evaluation. The issues that seem most pervasive for all components of a lunar resource recovery system include the following:

1. energy consumption,
2. equipment mass and size,
3. automation and teleoperation potential,
4. complexity, ease of operation,
5. fabricability, ease of assembly in space,
6. construction potential with indigenous materials,
7. maintenance requirements—durability,
8. resupply of consumables,
9. compatibility with lunar environment.

The technology for exploiting lunar resources must be considered as a system. Without using a systems approach, developers can not optimize the overall method(s). Furthermore, excavation components of the system must be flexible enough to perform a variety of construction and mining tasks. The early systems will be small and emphasize simplicity and reliance. As mining evolves on the Moon, more units will accomplish any necessary production increases. Eventually, mining equipment will evolve into larger, more efficient, and more specialized systems to support larger base activities.

The leading candidate mining systems for lunar-base operations consists of simple digging and hauling equipment to provide regolith for producing oxygen by an ilmenite reduction process (NASA 1990; Siekmeier and Podnieks 1990) as shown in Fig. 1. Figure 1 also depicts the portal of an underground tunnel. At some future time, underground tunneling may receive strong consideration for providing safe, comfortable habitat and other facilities as well as raw materials for further processing.

The following sections of this chapter discuss major mining components

Figure 1. Lunar mining scenario (figure after Sickmeier and Podnieks 1990).

of a resource recovery system—a description of each, the leading concepts, and key issues.

II. ENGINEERING PROPERTIES OF LUNAR REGOLITH

The lunar regolith, a several-meter-thick layer of debris, was formed from lithic sources by the destructive processes of comminution and constructive processes of agglutinate formation, all brought about by meteorite impacts. Vertical mixing through several meters and lateral mixing over kilometers also contribute significantly to the character of the soil. The individual soil particles consist of fragments of rocks (e.g., basalt, anorthosite, breccias), minerals (e.g., olivine, pyroxene, plagioclase), and agglutinates (aggregates of pieces of rocks and minerals bonded together by impact-produced glass). It is possible that this regolith is the material which will provide the initial and major resource for lunar resource exploitation.

The depths of regolith are largely a function of the age of the locale. The regolith at the older highland regions are considerably deeper (Apollo 16 = 10–15 m; Apollo 14 = 7–10 m), whereas regolith on the younger mare is appreciably less (Apollo 11 = 3–6 m; Apollo 12 = 3–4 m; Apollo 15 = 3–5 m; Apollo 17 = 3–5 m). Most scenarios for lunar manufacturing will need considerable quantities of this regolith, and detailed site evaluations will have to be conducted before any mining commences.

In order to consider mining operations on the Moon, it is necessary to deal with the engineering properties of the lunar regolith (the term soil refers to the <1 cm portion of the regolith). These properties include density, compressibility, shear strength, cohesion, permeability, diffusivity, bearing capacity, slope stability and trafficability. A thorough discussion of these properties can be found in Carrier et al. (1991).

The surface soil is slightly cohesive and well sorted. In general, it is a well-graded, silty sand to sandy silt—SW-SM to ML in the Unified Soil Classification System. The median particles size is 40 to 130 μm, with an average of 70 μm (Carrier and Mitchell 1990). Some 50% of the soil is <44 μm, 20% <20 μm. It is variable with depth, and the complex stratigraphy encompasses layers that are highly erratic in relative position. Density can vary over short distances, both vertically and horizontally. The unique and complex nature of lunar soil has been imposed largely by the presence of abundant agglutinates (up to 50%) and makes terrestrial analogs deceptive. For example, the average specific surface area of sub-mm lunar solid samples was determined as 0.5 $m^2\ g^{-1}$, whereas a soil consisting of spheres with the same particle size distribution would have a specific surface area of only 0.065 $m^2\ g^{-1}$. This necessitated the definition of a new shape parameter, called equivalent surface area ratio (Carrier and Mitchell 1990).

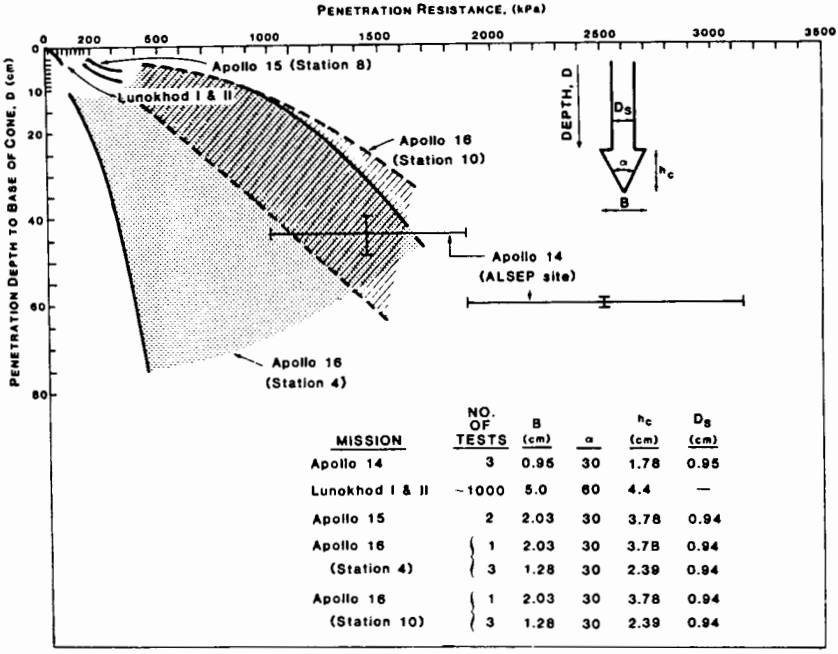
Probably the most unusual factor about the lunar soil is its density. Just a few centimeters beneath the surface, the soil is extremely dense, even denser than can be obtained with heavy compaction equipment on Earth (Carrier

and Mitchell 1990). At the surface, the relative density is 0 to 30%. Just 5 to 10 cm beneath the surface, it is 60 to 65%; at 30 cm, it is 90%, very dense. The average relative density for the upper 60 cm of the soil is 83±3% (Mitchell et al. 1974; Houston et al. 1974). This unique property is a result of static vertical stress and dynamic vibration brought about by the myriad of meteorite and micrometeorite impacts that the soil has experienced. Soil density varies from about 1.58 g cm^{-3} at the surface to >1.74 g cm^{-3} at 30 to 60 cm depth at mare sites.

Carrier and Mitchell (1990) illustrate this high relative density with discussion of the problems encountered by Apollo 15 astronauts when they attempted to drill hollow stems into the lunar soil. Because of the discontinuous nature of the helical auger flights (ribbing) on the drill stems, the cuttings rode up the ribs to a joint with no place to go except into the high relative density surrounding soil. The auger flight could not be cleared, and the bore stem became stuck in the ground at a depth of 1.5 m. However, there are several practical consequences of the high relative density, besides the need for continuous auger flights (Carrier and Mitchell 1990). Trafficability should not be a problem. Normal foundation loads can be easily supported. Excavated soil will bulk at least 10 to 15% even with extensive recompaction. Thin-walled core tubes manually can be driven only 0.7 m. It should be possible to excavate vertical walls to a depth of at least 2 m.

The property of *compressibility* describes the densification, or volume change, that occurs when confining stress is applied. The lunar soil has irregular fragile glassy particles called agglutinates that crush at relatively low confining stresses, which causes this soil to be more compressible than almost all terrestrial soils. The compression index, C_c, of lunar soil varies from 0.002 to 0.3. The compression behavior to a depth of >30 m (100 kPa) is provided in Mitchell et al. (1974).

The *shear strength* τ is expressed in terms of friction angle and cohesion and the Mohr-Coulomb strength theory. It is equal to the sum of the cohesive component c and a frictional component, $\sigma \tan \phi$: $\tau = c + \sigma \tan \phi$, where σ is the stress perpendicular to the surface of failure. The cohesive component c for lunar soils varies from 0.1 to 1.0 kPa, and the soil friction angle ϕ varies from 30 to 50 deg. The ultimate bearing capacity, slope stability, and trafficability of lunar soils are all governed by the soil shear strength. A self-recording penetrometer was used by the Apollo 15 and 16 astronauts to perform penetration tests with three 30° cones with base diameters 1.28 to 2.03 cm and one flat plate 2.54 cm×12.7 cm. Overall, there were a total of 17 cone and plate tests performed. Figure 2 depicts the penetration resistance measured during Apollo missions. Figure 3 shows the shear strength parameters for different depth ranges in lunar soil. In the top 60 cm of soil, average friction angle is 49 deg, and the cohesion for this interval is 1.6 kPa.



MISSION	NO. OF TESTS	B (cm)	α	h_c (cm)	D_s (cm)
Apollo 14	3	0.95	30	1.78	0.95
Lunokhod I & II	~1000	5.0	60	4.4	—
Apollo 15	2	2.03	30	3.78	0.94
Apollo 16	1	2.03	30	3.78	0.94
(Station 4)	3	1.28	30	2.39	0.94
Apollo 16	1	2.03	30	3.78	0.94
(Station 10)	3	1.28	30	2.39	0.94

Figure 2. Penetration resistance of lunar soil (figure after Mitchell et al. 1974).

III. SURFACE MINING AND EXCAVATION

A. Background

Base construction and resource recovery activities on extraterrestrial bodies can be on the surface, underground, or a combination of both. The relative benefits of surface or underground facilities is an important issue and will be discussed in more detail in Sec. IV. However, early operations (including mining) will start out on the surface. For operations on the Moon, this means excavating the regolith. As previously mentioned, this does not imply that Earth systems can be directly adopted for use in space. Nevertheless, terrestrial systems described below provide a starting point for developing space mining and excavating methods.

It is important to note that it would be premature to select one mining and excavation method at this time. Only after mission planners resolve mission objectives and define engineering constraints can the scope of mining activities be determined. The required quantities of materials and production rates have not yet be stated. Engineers can not develop realistic mining scenarios until those decisions are made and the limited resource data are augmented with detailed geotechnical information. We must match methods to deposit characteristics. The following sections will simply describe candidate systems

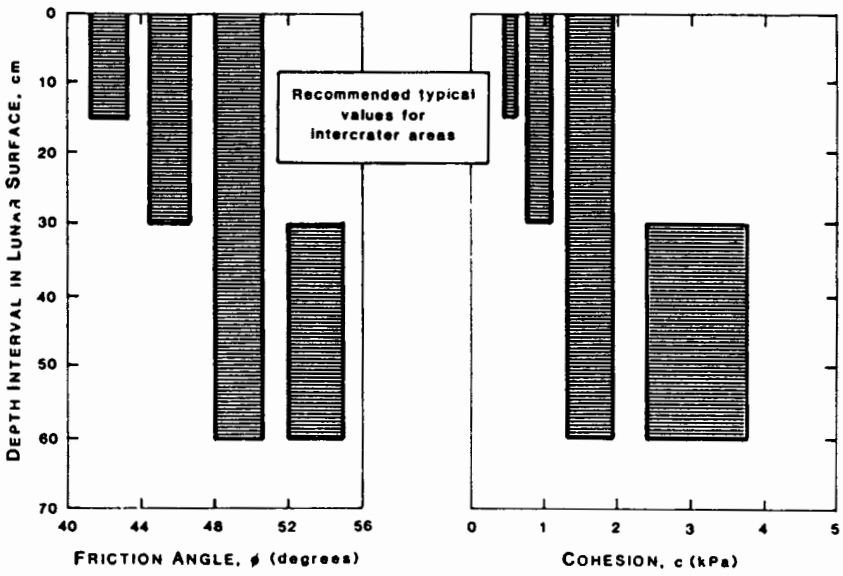


Figure 3. Shear strength parameters characteristic of different depth ranges of lunar soil (figure after Carrier and Mitchell 1990).

in very general terms to familiarize the reader with concepts that may form the base for lunar technology.

B. Concepts

Early space mining/excavation activities likely will feature small, simple machines suitable for excavating the lunar regolith, as well as for numerous other construction activities (Fig. 1). The leading candidates include (1) a mobile machine that digs, loads and hauls the regolith; (2) a mobile, multi-machine system wherein one machine digs and loads while another hauls; and (3) a system consisting of one fixed digging and loading machine plus a separate haulage vehicle (NASA 1990; Siekmeier and Podnieks 1990; Gothard and King 1989).

The first of these systems is based on the load-haul-dump machine presently used in mines. This machine scoops loose ore, loads it into an internal carrying compartment, and hauls it away. For the expected densely compacted lunar regolith, this machine would probably be fitted with a ripper blade to loosen the regolith for easier scooping. Explosive and vibrational loosening are also options.

With the ripper blade but without the hauling capability, the machine becomes the Ripper-Excavator-Loader discussed in Siekmeier and Podnieks (1990). This is one of the components of the second system, the mobile,

multi-machine system. The other component would be a simple haulage vehicle.

The third systems features a fixed excavation machine that is moved only when the pit advances.

Although several digging mechanisms could be integrated into such a machine, the articulated arm familiar in construction backhoes is a likely example. A separate haulage vehicle would transport the regolith to a processing site.

A coarse screen (grizzly, in mining terminology) would be located at the mine site to remove coarse-grained material and rocks from the mined regolith before haulage to the processing site (Fig. 1).

These concepts form the "strawmen" for evaluating other techniques. New innovative ideas should, however, be encouraged at this early time. Alternative excavation methods that have been proposed (for instance, at the First (NASA 1990) and Second Workshops on Construction and Mining in Space) include:

1. Front-end loader (used in mining and construction),
2. Clamshell (used to dig large holes),
3. Dozer (commonly used to shove soil),
4. Continuous drum-type mining machine (major coal mining machine),
5. Scraper (used for shallow digging and hauling),
6. Slusher (drum-wound cable and scoop arrangement for ore transportation),
7. Backhoe (commonly used to dig small holes and trenches),
8. Bucket-wheel excavator (overburden removal in large coal mines),
9. Dragline (large surface excavations),
10. Explosive casting (used to move overburden in some coal mines,
11. Auger (as on a snow thrower),
12. Rotating brush (as on a road sweeper).

Proposed alternate haulage methods include:

1. Conveyor system,
2. Cable tram,
3. Rail tram,
4. Pipeline,
5. Magnetically levitated containers,
6. Ballistic throwing,
7. Electrostatic transport. Each of the strawman methods and alternative technologies have many pros and cons that must be evaluated in rigorous scientific and engineering studies before firm recommendations can be made.

C. Issues

Equipment designed to work in space must operate under constraints much different from those in terrestrial environment. Temperatures and temperature changes cause problems. Day/night and sun/shade differences will severely impact equipment and operations.

Gravity allows equipment to develop adequate reaction forces for excavating and handling materials on Earth, but low lunar gravity will require new concepts for reaction forces.

Vacuum presents a difficult work environment for personnel and equipment. Mobile life support and/or remote operation will be required. Machine components must be designed and built for operation in vacuum. Equipment must be developed with minimum exposed seals or bearing surfaces. Cable-driven systems such as draglines and slushers will experience difficulties caused by friction and dust accumulation. Triboelectric charging may cause hazards.

Electromagnetic radiation and particle bombardment will be a major concern in the design and operation of surface mining equipment in space. Any surface equipment must be designed to function in the severe radiation environment. Dust also becomes a more severe problem for space operations where conditions result in ubiquitous clinging of dust to exposed surfaces. This is particularly critical because of the abrasive nature of the glassy particles in the dust.

Design criteria for surface mining and construction equipment frequently are in conflict. Designing such equipment will require compromise and considerable testing to gather pertinent information. The key factors to be considered in designing this equipment include:

1. Simplicity,
2. Ruggedness and robustness,
3. Flexibility,
4. Availability of proven advanced technology,
5. Low energy requirements,
6. Low machine mass,
7. Automation and teleoperation potential,
8. Tribology (bearings and seals),
9. Availability of advanced fabrication materials.

These factors must be considered for excavating a soil that is much denser and more difficult to dig than any in terrestrial applications.

IV. UNDERGROUND MINING

A. Background

Although early lunar resource recovery will feature surface mining, underground mining merits consideration for mature base operations. Surface

mining operations exposed to the lunar environment will require protection for equipment and operators. Underground mining could provide a sheltered mining environment and the mined-out openings could also shelter processing facilities, operations centers, and habitats.

Although still very controversial, a key issue may be the desirability of processing large rock chunks for oxygen rather than fine-particle soil. The relative recovery of oxygen may favor processing rock. The mare soil appears to contain less ilmenite than the rock from which it was formed. This is caused by the soil formation process whereby micrometeorite impacts melt some minerals that form glasses upon cooling. The glass welds minerals and rock fragments together into aggregates called "agglutinates" (Taylor 1990). As a soil matures, it therefore contains more glass and less mineral. Experiments have shown that the oxygen recovery from the underground rock is twice that from soil material (Heiken and Vaniman 1989). And it may be more effective to obtain rock from an underground mine than from scattered boulders in a surface mine.

The issue of efficient oxygen recovery is not that simple, however. As demonstrated on lunar samples, fine grained feedstock can be concentrated by magnetic beneficiation (Taylor and Oder 1990). This may mean that it is more energy efficient to upgrade and process lunar soil than crush, grind, and process the higher-grade lunar rock. An evaluation of this trade-off is a critical requirement of lunar planning process. So also is an evaluation of the relative merits of alternate processing strategies. Molten rock electrolysis, for example, may favor processing large rock chunks rather than soil, again with attendant advantages of underground mining.

The possible merits of underground mining include:

1. Underground mining simultaneously provides ore and working and living space for humans;
2. An enclosed pressurized underground complex allows "shirt-sleeve" environment for human habitation and normal operating conditions and maintenance for the equipment;
3. The reliability and expected usable lifetime of the equipment will be considerably longer when operating in the sheltered underground environment protected from vacuum outgassing, thermal variations, radiation, and dust adhesion;
4. The crew exposure to EVA will be reduced to a minimum.

B. Underground Mining Methods

Two general categories of underground mining systems are used on Earth: drill-blast-muck and mechanical. These and several novel approaches, such as microwave fragmentation or laser cutting, should be considered for space mining.

The drill-blast-muck system is the most common excavation method in terrestrial hard rock underground mining operations. Holes are drilled in a

rock face, an explosive is placed in the bottom of the hole, and the blast is detonated. The most common blasting agent is a mixture of ammonium nitrate and fuel oil, packaged as pellets. It is generally the method of choice on Earth because of its flexibility, low unit-energy cost, and minimal excavation equipment requirements.

For space applications, some of these benefits vanish. Because lunar underground mining operations may be close to other underground operations and habitat, blasting may be risky. It also would be important to develop blasting agents from reactive indigenous resources rather than transporting large amounts of explosives from Earth.

Mechanical and/or novel mining systems may be better suited for lunar operations. Four general types of mechanical mining systems are available on Earth: tunnel boring machines, drum-type continuous miners, roadheaders, and rock splitters.

(1) Tunnel boring machines (TBMs) were developed for applications requiring long, straight tunnels such as a railway tunnel. The machines are massive (on the order of hundreds of tons) and rather inflexible. Their limited versatility and large mass would seem to preclude their use for lunar excavations.

(2) Drum-type continuous miners developed for mining coal, can mine soft to medium hard material at the rate of 200 to 700 mt hr^{-1}. An integral loading system deposits the excavated rock on a belt conveyor or haulage vehicle. They are moderately flexible; they can be used to either drive a heading or excavate a room. Their drawbacks for lunar base excavation are: (a) dependence on machine weight to counteract the cutting forces; (b) massive size (50–100 mt); and (3) high-power requirements (250 to 300 kw). A benefit may be the present Bureau of Mines' research on automation and teleoperation of continuous miners (Schnakenberg 1990).

(3) The roadheader, originally intended to enlarge access headings in coal mines, has developed into a versatile machine for a wide variety of mining and excavation applications. It can cut soft- to medium-hard rock. However, it depends on the weight of the machine to counteract the cutting forces. The roadheader's main attraction is its boom-mounted cutterhead, which allows an operator to cut and trim an opening of virtually any shape. Roadheaders could be effective for newly established small- to medium-scale operations. Canadian potash mining companies have developed teleoperated roadheaders.

(4) The hydraulic rock splitter has been used extensively for secondary breakage. It is well suited for lunar application because it generates the tensile breaking force by pushing against its own anchoring system, requiring no external thrust (Fig. 4). The Bureau of Mines used the rock splitting principle to develop a unique version called the radial-axial splitter. This crawler-mounted excavation machine consists of a percussive drill, drill feed, radial-axial splitter and loader. For lunar application, the hydraulic percussive drill and radial-axial splitter can be mounted on a single boom; the system would weigh 5.5 mt with power requirement of 50 kw and a production rate

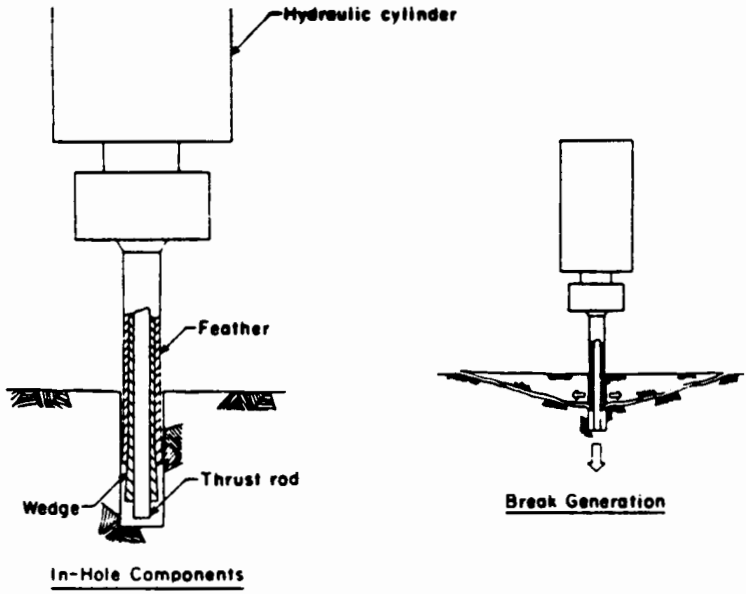


Figure 4. Rock splitter principle of operation (figure after Anderson 1990).

of 10 mt hr^{-1} (Anderson 1990).

Novel excavation methods that have been uneconomical on Earth may find application in a lunar setting. Such methods include thermal fragmentation with electromagnetic energy in the form of microwaves or laser beams. Sonic vibration, and electromagnetically accelerated projectiles have also been suggested, but too little evidence is presently available to warrant further discussion.

Research engineers have fragmented rock in terrestrial and in simulated lunar environments with *electromagnetic energy*. For lunar applications, this fragmentation form might require relatively low input energy. Electromagnetic methods can also be integrated with processing operations to recover certain elements by direct pyrolysis.

Bureau of Mines researchers conducted preliminary tests on fragmenting igneous rocks using three forms of electromagnetic energy: (1) CO_2 laser; (2) microwave; and (3) solar (Lindroth and Podnieks 1988). All three methods effectively fragmented the rock with thermal stresses or melted it, depending on the rock properties and operating parameters. The most energy-efficient method was to fracture the rock by inducing a "hot spot" inside it. This zone expands and shatters the surrounding rock material. The frequency of the microwave energy and dielectric constants of the rock control the fragmentation. In a mining application, a microwave device could be pointed against a rock face in an underground opening to fracture the rock. The

fractured rock would then be transported to a processing site.

Electromagnetic waves can also melt and disintegrate the rock. An interesting possible application might be to direct a laser beam to kerf cut a large rock block from the face, perhaps assisted with simple mechanical wedges. The block could then be hauled to a processing site via a conveyor or even magnetically levitated carts.

At this time, the most attractive underground excavation systems appear to be the radial-axial splitter and the roadheader sized down to lunar application. Considering the weight and energy restrictions, especially for the initial lunar mining operations, the splitter method may be the best choice.

Of the methods considered for underground excavation, single-level open stoping would be the most desirable. The low lunar gravity may minimize ground control problems. If necessary for ground control, a room-and-pillar method can be substituted. Mining configurations are described in detail in *SME Mining Engineering Handbook* (Cummins and Given 1973).

To create a terrestrial working environment and habitat in the underground excavation, the opening must be sealed (Fig. 5) (Siekmeier and Podnieks 1990). To maintain an "atmospheric" pressure inside the opening, the surfaces of the excavation also must be sealed. Researchers have glazed basaltic rocks using different electromagnetic power (microwave) sources to seal the cracks (Lindroth and Podnieks 1988). A suitably sized microwave generator would produce 50 kW power at 2.45 GHz frequency and use about 80 kW electric power. Such a unit would melt and seal approximately 1 m^2 hr^{-1} of the underground excavated surface. This rate would increase if experiments show that the wall would adequately seal at shallower depths of melted rock.

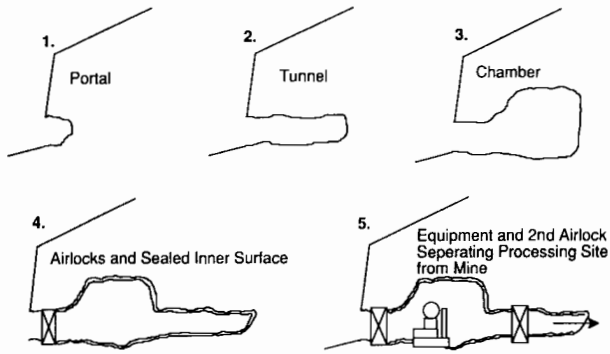


Figure 5. Underground mining and sealing concept.

C. Issues

The harsh lunar environment will require the human habitats to provide safe shelter from intense radiation, temperature extremes and vacuum. Studies

on above-ground habitats indicate difficulty in constructing a safe and efficient habitat with an inside environment of terrestrial atmospheric pressure and temperature. At the Extraterrestrial Mining and Construction Workshop in 1989 (NASA 1990), participants recommended constructing underground habitats for superior safety. There are numerous examples on Earth where underground mined-out space is used for a variety of commercial and defense activities. Considering these factors, underground habitats and service facilities may provide an excellent alternative to surface structures.

Despite the possible advantages of underground lunar mining, surface mining will also be necessary for lunar-base development. Hydrogen and helium occur primarily in the upper several meters of the regolith. These resources will have to be mined with surface methods. Therefore, surface and underground mines will likely be operated jointly in order to support the lunar base.

It seems probable that lunar construction will at some point feature underground tunneling and excavation. A key decision would be whether to use the excavated rock as feedstock for oxygen production or to simply dump the excavated rock as roadfill or waste. The issue arises because excavated rock would have to be crushed and ground before processing for most of the proposed oxygen recovery processes.

Engineers must evaluate the trade-offs in benefits of higher-grade underground ore versus the increased energy consumption and equipment requirements for crushing.

A related issue is the need to consider the entire mining, beneficiation, and processing system as a whole when evaluating energy trade-offs. Some of the oxygen and construction metal processing strategies (for example, direct pyrolysis) would look more promising in context of underground mining. Underground mining can provide large blocks that, especially in vacuum, may be more efficiently heated than crushed material. Direct pyrolysis of walls in an underground chamber sounds far-fetched now but could be particularly intriguing for a mature lunar base. The point is, we now have an opportunity to consider and systematically evaluate many innovative mining and processing systems before becoming locked into one.

Underground openings must be sealed to provide terrestrial atmosphere in the underground opening. Technology must be developed to effectively install an airlock in a mine opening. The further development on the underground mine will require effective sealing of any cracks in the rock mass to maintain the atmospheric pressure in the mined-out openings used for processing and base control facilities, and living quarters for the base personnel.

Climate control in the underground complex will require effective temperature-control systems and air recirculation, including capabilities for filtering out dust and other impurities. Climate control technologies for sealed installations are available but must be adapted to lunar underground space applications.

V. IN-SITU PROCESSING

A. Definition and Purpose

In place or *in-situ* processing is a technology whereby a mineral or element is recovered without excavating the ore for processing in surface plants. Terrestrial technology features chemical solvents injected into rock underground to dissolve the target minerals. The resultant solutions are then pumped back to the surface for treatment to recover the desired commodity, i.e., copper, uranium, salt, potash, sulfur, etc. *In-situ* processing replaces the mining, beneficiation and smelting components of a resource recovery system. On the Moon, oxygen or construction materials could be recovered with *in-situ* processes. It should be noted that these processes would not permit recovery of solar-wind-deposited gases, which would require separate recovery operations.

B. In-Situ Processing Concepts

On Earth, *in-situ* processing by chemical leaching is commercially applied to copper and uranium recovery (Ahlness and Pojar 1983; Larson 1978). Wells are drilled into the zone of interest in a pattern of injection wells and recovery wells (Fig. 6). Solvents (usually dilute sulfuric acid or ammonia) are pumped down the injection wells and forced through the formation to adjacent recovery wells. As the solution moves through fractures and interconnected pores, it dissolves the metallic minerals. When these solutions reach the recovery wells, they are pumped to the surface. The target metals are then precipitated or plated from solution.

This process works well on Earth because the metals of interest readily dissolve at ambient conditions. However, none of the leading candidate processes for lunar oxygen recovery would appear to be practical at the ambient lunar rock temperatures of $-23°$ C (see Chapter by Taylor and Carrier).

A more promising application on the Moon would be thermal *in-situ* processing. Such a concept would be based on operations in underground tunnels and chambers. Holes could be drilled into the walls for probes for electrolysis or direct pyrolysis. Microwave heating would seem to represent an interesting potential heat source for direct pyrolysis.

C. Issues

The central issue in chemical *in-situ* processing is the identification of processes that would work under lunar conditions to recover oxygen and/or useful construction metals such as aluminum, iron and titanium. Even identifying such a process, researchers must perform considerable experimentation to determine the kinetics of the process, the technology for operating a well field under lunar conditions, and information on potential recovery rates. Another difficulty to overcome would be containing volatile solvents within a fractured rock mass exposed to the lunar ultra-high vacuum environment.

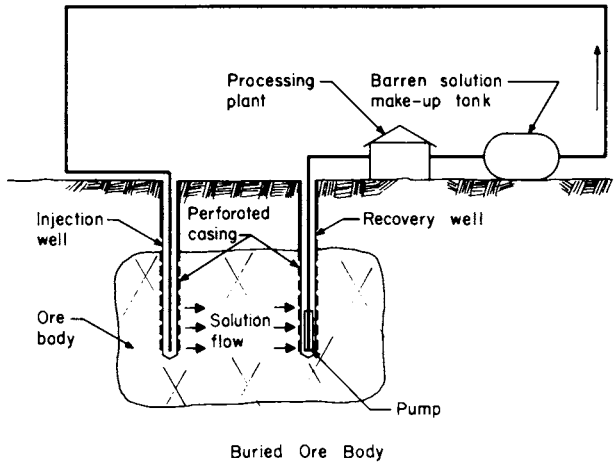


Figure 6. *In-situ* leach processing concept.

Although underground excavation with direct pyrolysis or electrolysis methods of *in-situ* processing appear to have a better chance of success than chemical *in-situ* processes, considerable development work must be completed to provide equipment that will perform under lunar conditions with low-mass and low-energy requirements. The energy requirements to mine and process material *in-situ* must be compared with the combined energy requirements for surface mining, underground mining, beneficiation and processing.

VI. SUMMARY

Lunar resources can provide many life-support elements for a lunar base and fuel for space vehicles. Construction materials for bases in space can also be obtained from indigenous resources.

Mining and processing equipment and techniques presently used on Earth will not work in space without significant changes. Although terrestrial concepts described in this chapter provide a good starting point for space-oriented research and development, a rigorous program of scientific and engineering study must be conducted to provide the necessary technology.

Before embarking on such a program, it is critical to define space mission requirements. Engineers also believe that we must obtain considerable data on the character of base mining and construction sites before singling out one or two mining methods to study.

REFERENCES

Ahlness, J. K., and Pojar, M. G. 1983. In Situ Copper Leaching in the United States: Case Histories of Operations. Bureau of Mines IC 8961 (Pittsburgh: Bureau of Mines).

Anderson, S. J. 1990. Drill-split mining with radial-axial loading splitters. In *Rock Mechanics Contribution and Challenges: Proc. 31st U. S. Symp.*, Golden, Co., June 18–20 (Rotterdam: Bulkema), pp. 511–518.

Carrier, W. D., III, and Mitchell, J. K. 1990. Geotechnical engineering on the Moon. *de Mello Volume: A Tribute to Prof. Dr. Victor F. B. de Mello*, ed. E. Blucher (São Paulo: Editora Edgard Blücher), pp. 51–58.

Carrier, W. D. III, Olhoeft, G. R., and Mendell, W. 1991. Physical properties of the lunar surface. In *The Lunar Sourcebook,* eds. G. Heiken, D. Vaniman and B. French (Cambridge: Cambridge Univ. Press), pp. 475–594.

Cummins, A. B., and Given, I. A., eds. 1973. *SME Mining Engineering Handbook*, vol. 1 (New York: Soc. of Mining Engineers, American Inst. of Mining, Metallurgical, and Petroleum Engineers).

Gothard, B., and King, B. 1989. An extraterrestrial excavator robot. Space Mining and Manufacturing: Proc. First Annual Symp. UA/NASA SERC, Tucson, Ariz., Oct. 24–26, pp. IV-21–IV-43.

Heiken, G. H., and Vaniman, D. T. 1989. Petrography of lunar ilmenite resources. *Lunar Planet. Sci.* XX:400–401 (abstract).

Houston, W. N., Mitchell, J. K., and Carrier, W. D., III. 1974. Lunar soil density and porosity. *Proc. Lunar Planet. Sci. Conf.* 5:2361–2364.

Larson, W. C. 1978. Uranium In Situ Leach Mining in the United States. Bureau of Mines IC 8777 (Pittsburgh: Bureau of Mines).

Lindroth, D. P., and Podnieks, E. R. 1988. Electromagnetic energy applications in lunar mining and construction. *Proc. Lunar Planet. Sci. Conf.* 18:365–373.

Mitchell, J. K., Houston, W. N., Carrier, W. D., III, and Costes, N. C. 1974. Apollo Soil Mechanics Experiment S-200, Final Report. NASA Contract NAS 9—266, Space Sci. Lab Series 15, Issue 7 (Berkeley: Univ. of California).

NASA. 1990. *Proc. of the Workshop on Extraterrestrial Mining and Construction*, ed. B. M. Register, Golden, Co., May 2–4, 1989 (Houston: Lockheed, E. S. C.).

Schnakenberg, G. H., Jr. 1990. U. S. Bureau of Mines coal mining automation: Research update. In *Mining Automation*, ed. J. N. Wilson (Saskatoon: Saskatchewan Research Council), pp. 179–191.

Siekmeier, J. A., and Podnieks, E. R. 1990. Mining Technology for Lunar Resource Utilization. BOM-NASA Project Report, W-17.036 (NASA).

Taylor, L. A. 1990. Rocks and minerals in the regolith of the Moon: Resources for a lunar base. In *Advanced Materials: Applications of Mineral and Metallurgical Processing Principles,* ed. V. I. Lakshmanan (Littleton, Co.: Soc. for Mining, Metallurgy and Exploration), pp. 29–47.

Taylor, L. A., and Oder, R. R. 1990. Magnetic beneficiation of highland and Hi-Ti mare soils: Rock, mineral, and glassy components. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, vol. 1, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 143–152.

OXYGEN PRODUCTION ON THE MOON: AN OVERVIEW AND EVALUATION

LAWRENCE A. TAYLOR
University of Tennessee

and

W. DAVID CARRIER, III
Bromwell and Carrier, Inc.

The production of oxygen on the Moon utilizing indigenous resources and materials is paramount to a successful permanent habitation on the lunar surface. At least 20 different processes have been put forth to accomplish this. The two lunar liquid oxygen generation schemes which have received the most study to date are those involving: (1) reduction of ilmenite ($FeTiO_3$) with H_2, but also by CO and CH_4; and (2) molten silicate (magma) electrolysis, both direct and fluoride-fluxed. Several other processes, including glass reduction with H_2, vapor phase pyrolysis, ion (plasma) pyrolysis, carbochlorination, HF acid leaching, and fluorine extraction, also have received significant study. However, all processes should be addressed at this stage in our considerations. There is an obvious need for considerably more experimentation and study. Some of these requisite studies are in progress. This chapter reviews 20 processes for the production of oxygen on the Moon, including an evaluation of the perceived feasibility for each.

I. INTRODUCTION

The *in situ* utilization of resources is requisite to the settling and evolution of an autonomous colony on the Moon. The production of lunar liquid oxygen (LLOX) could result in tremendous cost savings on propellant for effective transportation systems, and oxygen also is needed to support human existence. With such high costs for bringing oxygen (or other supplies) from Earth to the Moon, it is prudent that indigenous lunar sources be sought for any materials needed for lunar bases. In addition, exporting of LLOX to low-Earth orbit (LEO) could also effect a highly significant economic savings, compared to an Earth to LEO transport.

Extensive engineering/science studies are necessary for an orderly preparation in anticipation of a return to the Moon. Industry, government, and academia must combine their wealth of engineering experience in order to

develop concepts and designs for effective resource utilization (see, e.g., Simon 1985; Christiansen et al. 1988; Sherwood and Woodcock 1991; Sullivan 1990). Specifically, there is a wealth of data in the literature on the subject of oxygen production on the Moon; unfortunately, much of it is in "gray literature" virtually hidden in reports in both government and industry. Particularly noteworthy are the pioneering studies of Rosenberg et al. (1966) in the mid 1960s and the more recent contributions of Waldron (1985,1989). The most recent compilation of various engineering aspects of LLOX production can be found in an Eagle Engineering report to NASA by Christiansen et al. (1988). We have drawn liberally from this excellent paper. Other sources which we have used extensively are those of L. A. Taylor (1990*b*,1992), L. A. Taylor et al. (1990), and L. A. Taylor and Carrier (1991). Although we have tried to reference extensively the ideas presented in this chapter, we apologize for any lack of referencing that may have occurred. Lastly, the *Lunar Sourcebook* (Heiken et al. 1991) is a thorough condensation of lunar scientific and engineering data which will be invaluable for lunar process design, especially with regards to the types and nature of lunar resources available.

The goal of this chapter is to review many of the various process concepts which have been proposed for the production of oxygen on the Moon. Aspects of the production schemes are addressed which bear on their feasibilities for winning oxygen from lunar materials. These processes are in various stages of development ranging all the way from early engineering concepts to proven schemes. This is the reason for the apparent unbalanced treatment of the various processes in this chapter. In addition, although largely a subjective effort, we have attempted to rank these processes according to our perceived potential for success as *the chosen process*. This is not intended to be anything but a paradigm; others may rank the processes differently.

II. FACTORS FOR CONSIDERATION

Thorough systems definitions of all aspects of oxygen generation processes are necessary. Energy requirements are obviously of great importance, as is consideration of solar versus nuclear-electric power sources. The overall simplicity of the process is also of great importance. This includes not only the number, complexity and efficiency of steps in the entire process, but also the feedstock requirements and costs of importing consumable reactants.

The sensitivity of process mass and power to both feedstock and production rate are paramount. In the evaluation of the lunar feedstock, we must determine whether sites in maria or highlands are to be occupied and whether rocks or soils are to be mined. The highlands are composed of rocks and regolith containing appreciable anorthosite, with large amounts (>70 wt%) of plagioclase feldspar, thereby making them rich in CaO, Al_2O_3 and SiO_2. In contrast, the maria consist of basalts with lesser amounts of plagioclase, but with appreciable quantities of olivine, pyroxene, and oxide minerals (ilmenite and spinels), making the maria compositions considerably richer in

FeO, MgO, and TiO_2. Does the feedstock require extensive beneficiation or can the lunar "ore" be used largely untreated, except maybe for coarse sizing? The largely unconsolidated nature of the lunar regolith makes it the material of choice for most processes, but certain schemes for oxygen production (e.g., ilmenite reduction by hydrogen) require sizing, possible grinding, and magnetic/electrostatic beneficiation.

There are certain minor phases in the rocks and soils that may have deleterious effects upon various steps of the overall oxygen-production processes. For example, troilite (FeS) is a common minor phase (<1 wt%), particularly in mare basalts, and it is not easily removed, especially from its typical attachment to ilmenite. Upon roasting, particularly in hydrogen or oxygen, it readily decomposes, liberating sulfur in the form of H_2S and SO_2, respectively. It may seem possible to use bulk, unbeneficiated mare soil for a particular oxygen process, but the presence of sulfur may not be tolerated in the process. Such minor yet potentially harmful components of the indigenous materials must be taken into account in the evaluation of feedstock requirements for these processes.

Once the processes are well understood in the laboratory and, more importantly, in pilot plant operation, it will be possible to scale the plant mass, power requirements, feedstock requirements, etc. for the actual design of the production plant. At that time, comparisons should be made between delivering a series of small self-contained modular production plants to increase oxygen production versus constructing a single, large plant.

III. POTENTIAL PROCESSES FOR OXYGEN PRODUCTION

There are over 20 different process concepts which have been proposed for the production of oxygen from lunar materials. In this chapter, we review most of these. Table I lists the particular processes, with appropriate references, which are reviewed. It is possible to classify the schemes for oxygen generation according to the nature of the chemical process involved (see, e.g., Waldron 1989). Our classification is based upon the manner in which the reactants are processed. Thus, we recognize divisions of gas/solid interaction, silicate/oxide melts, pyrolysis, aqueous solutions, and co-product recovery (Table I). The subdivisions further address specific processes. Although the processes ilmenite reduction with hydrogen, molten silicate electrolysis, fluxed molten silicate electrolysis, and vapor-phase pyrolysis have received considerable study to date and are the most commonly referred-to processes, some lesser known processes are presented in abbreviated form.

Many of these processes have a common thread in requiring a step involving electrolysis or thermochemical reduction to split H_2O, CO_2, etc. in order to release the oxygen for LLOX production. This will entail modification of existing conventional techniques. Alternatively, it may be possible to develop a process using a ceramic solid electrolyte at elevated temperatures (600–800°C), an important consideration for energy conservation.

TABLE I
Processes for Oxygen Production on the Moon

Processes	References
Solid/Gas Interaction	
Ilmenite reduction with hydrogen	Gibson and Knudsen (1988*a*)
Ilmenite reduction with C/CO	Chang (1959); Zhao and Shadman (1990)
Ilmenite reduction with methane	Friedlander (1985)
Glass reduction with hydrogen	Mckay et al. (1991)
Reduction with hydrogen sulfide	Dalton and Hohman (1972)
Extraction with fluorine	Burt (1988); Seboldt et al. (1991)
Carbochlorination	Lynch (1989)
Chlorine plasma reduction	Lynch (1989)
Silicate/Oxide Melt	
Molten silicate electrolysis	Haskin (1985); Colson and Haskin (1990)
Fluxed molten silicate electrolysis	Keller (1986); Keller and Taberaux (1991)
Caustic dissolution and electrolysis	Dalton and Hohman (1972)
Carbothermal reduction	Rosenberg et al. (1966); Cutler and Krag (1985)
Magma partial oxidation	Waldron (1989)
Li or Na reduction of ilmenite	Semkow and Sammells (1987)
Pyrolysis	
Vapor phase reduction	Steurer and Nerad (1983); Chapter by Senior
Ion (plasma) separation	Steurer and Nerad (1983)
Plasma reduction of ilmenite	Allen et al. (1988)
Aqueous Solutions	
HF acid dissolution	Waldron (1985)
H_2SO_4 acid dissolution	Sullivan (1990)
Co-Product Recovery	
Hydrogen/helium water production	Christiansen et al. (1988)

IV. SOLID/GAS INTERACTION

Lunar minerals can be reacted with gases in order to produce oxygen. This division addresses the different gases that are interacted with the various feedstocks. Several of these processes involve the mineral ilmenite, $FeTiO_3$. Basically, the reduction of ilmenite can be considered as the reduction of the

FeO portion of the ilmenite, leaving a product of TiO_2 and Fe°, with the release of oxygen. These are not necessarily new processes. In fact, a proven industrial process for the reduction of ilmenite with various gas phases (e.g., H_2, CO) was included in the patent application of Chang (1959), who was interested in the recovery of the metals. However, for a lunar application, it is the liberated oxygen which is of greatest value. Therefore, several of the oxygen-production processes discussed below are based upon these initial ideas.

A. Ilmenite Reduction with Hydrogen

Williams et al. (1979) and Williams and Mullins (1983) first speculated upon the reduction of ilmenite for the production of lunar oxygen. Ilmenite, $FeTiO_3$, with up to 8 wt% MgO, makes up 10 to 20% by volume of some mare basalt rocks and lesser amounts of mare soils (L. A. Taylor 1990*b*; S. R. Taylor 1975,1982). This feedstock can be readily beneficiated from the mare basalt soils (Taylor and Oder 1990) and rocks (Vaniman and Heiken 1990). Furthermore, the kinetics of its reduction are considerably faster than for silicate minerals (Hammond and Taylor 1982) The actual reaction of interest for the reduction of ilmenite is:

$$FeTiO_3 \Longleftrightarrow Fe + TiO_2 + 1/2O_2. \quad (1)$$

Oxygen constitutes 10.4 wt% of the products (right side of the equation). The position of this reaction curve as a function of temperature and oxygen partial pressure (fO_2) was determined by L. A. Taylor et al. (1972). The reduction of ilmenite can be brought about by driving the equation to the right through reduction. Choosing a reductant is important to the efficiency and complexity of the process. The reduction of ilmenite by hydrogen involves the following chemical reaction:

$$FeTiO_3 + H_2 \Longleftrightarrow Fe + TiO_2 + H_2O$$

$$H_2O \Longleftrightarrow H_2 + 1/2O_2. \quad (2)$$

Figure 1 depicts a flow diagram for the ilmenite reduction process. Although this diagram refers specifically to reduction by hydrogen, it illustrates the system applicable to CO and CH_4 as well. The overall oxygen-production process using hydrogen has been studied extensively by Gibson and Knudsen (1988*a*, *b*), Gibson et al. (1990) and Williams (1985). Reaction temperatures of 700 to 1000°C are needed in order to obtain suitable reaction rates and acceptable conversion efficiencies. The water produced is electrolyzed into H_2 and O_2, and the hydrogen is recycled.

The concept behind this process is simple, and ilmenite is abundant in many mare rocks and soils. Carbotek, Inc. (Houston, Texas) has patented an ilmenite, hydrogen-reduction technique involving a fluidized-bed process for the production of LLOX (Gibson et al. 1990; see Fig. 2). The reactor consists

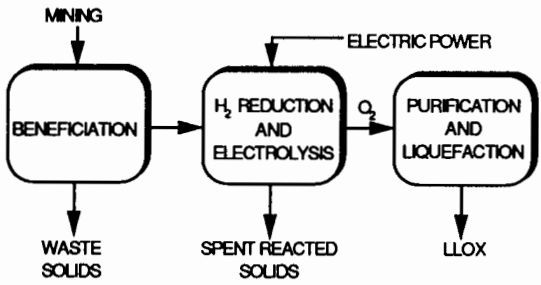


Figure 1. Flow diagram for the production of LLOX by ilmenite reduction with hydrogen.

of three stacked fluidized beds with countercurrent flow of gas and solids. The hydrogen passes upward from the bottom of the reactor, whereas the ilmenite feedstock passes downward. In this design, the top bed is for preheating the incoming feed solids against hot, recycled hydrogen, and the bottom bed is to extract otherwise wasted heat from the spent solids. For such a continuous process, the crucial feedstock parameters are: ilmenite content, non-ilmenite constituents, particle size and density, and amounts of troilite (FeS), with its corrosive sulfur.

Evaluation of H_2 Reduction of Ilmenite Process. The chemistry of this process is not complicated; however, lunar ilmenite commonly contains chemical impurities of Mg (up to 8% MgO) and Cr (up to 2.5% Cr_2O_3). The effects of these upon the kinetics of ilmenite reduction with hydrogen, or any other gas, remains an unknown. This hydrogen reduction process has been demonstrated in the laboratory with pure $FeTiO_3$. A continuous process has been designed—e.g., the fluidized bed process of Carbotek; however, the practicality of this process at 1/6 terrestrial gravity remains to be demonstrated. The process temperatures are relatively low (700 to 1000°C), thereby reducing heat exchange and energy requirements.

Work to date would seem to indicate a low equilibrium per-pass conversion of H_2 to H_2O. As per-pass conversions decrease, particularly at lower temperatures, the gas-flow rate must be increased, with larger reactor requirements and other mass penalties. Another potential problem exists in the retention of hot hydrogen within the overall system. Even the smallest of leaks on the Moon can exhaust large quantities of gas, a function of the 10^{-12} torr lunar atmosphere, a "hard" vacuum by Earth standards. However, lunar ilmenite in the soil contains considerable hydrogen (hundreds of ppm), present as solar-wind implanted protons. This hydrogen is easily released upon heating and can counteract, to some extent, consumption of hydrogen reactant through losses caused by leaks, etc. The problem of operating a plant in a vacuum is applicable to all the processes for oxygen production on the lunar surface.

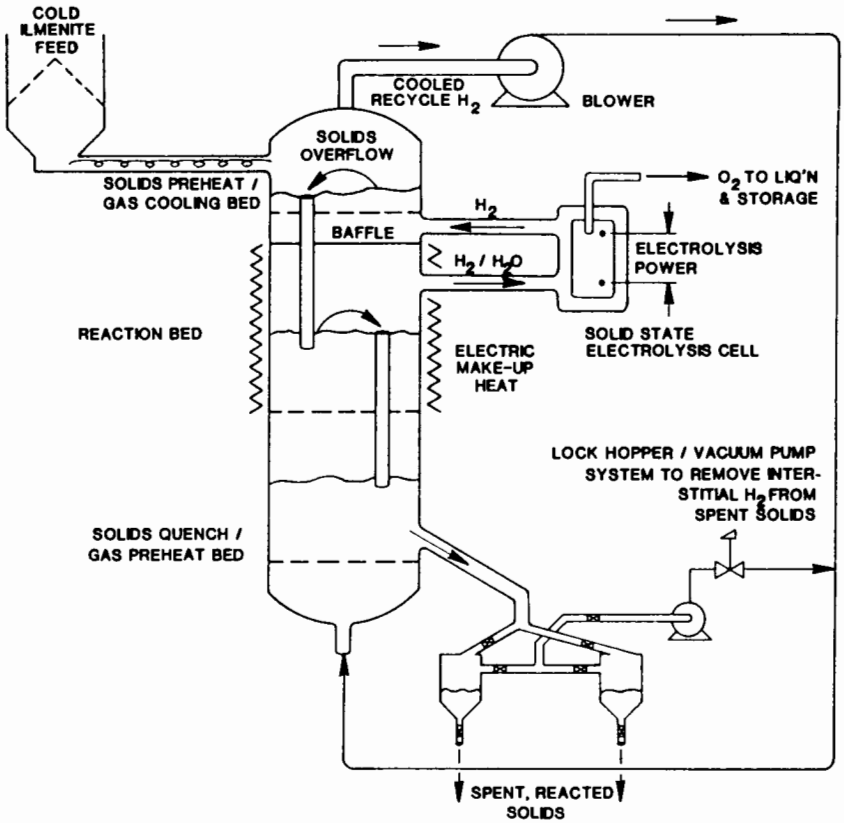


Figure 2. Continuous fluidized-bed process for ilmenite reduction by hydrogen for the production of LLOX, modified after Gibson and Knudsen (1988*a*).

Acquiring a feedstock which is ilmenite-rich will necessitate a high-Ti mare ore—a site-specific requirement. Considerable beneficiation of a mare basalt or soil will be necessary. In addition, the presence of the sulfide mineral troilite (FeS), commonly associated with ilmenite in mare rocks and soils, may be a problem in this process, as well as for most other processes discussed below. It can lead to the generation of corrosive H_2S and will necessitate a process step for removal of either the troilite from the feedstock or H_2S from the products. Of course, there are several methods by which troilite, with its sulfur component, can be removed from the feedstock, but this additional step in the beneficiation scheme is a complication which will be an economic consideration. In general, the simplicity of the beneficiation process is of major importance for the lunar production facility. Indeed, the mining and beneficiation processes may prove to be the key discriminator among competing oxygen production processes.

B. Ilmenite Reduction by Carbon Monoxide

The idea of using CO as the reducing gas to process ilmenite is based upon the same general scheme as with the use of hydrogen as the reductant. Chang (1959) stated that either H_2 or CO or some combination of both could be used as the reducing gas. Zhao and Shadman (1990) have recently investigated this process concept using CO.

Figure 3 depicts CO as the reductant involved in the reactions:

$$\mathrm{FeTiO_3 + CO \Longleftrightarrow Fe + TiO_2 + CO_2}$$

$$\mathrm{CO_2 \Longleftrightarrow CO + 1/2O_2}. \tag{3}$$

As can be seen, the system design used is similar to that for the fluidized-bed scheme proposed for hydrogen reduction, as discussed above.

Evaluation of CO Reduction of Ilmenite Process. The product from this reduction process is CO_2 which must be "cracked" to liberate oxygen. This is a very endothermic reaction which will consume significant energy. In addition, many of the same advantages and disadvantages mentioned in the hydrogen section above are equally applicable here.

There is solar-wind implanted carbon in the soil (20–30 ppm), particularly in ilmenite; with the roasting of the soil for the liberation of hydrogen and helium, this carbon also will be released and could be recovered. The ilmenite feedstock, with its solar-wind implanted particles, will also contribute carbon, as well as hydrogen, to the overall process. Thus, it may be possible to make up consumption losses of the reactant from indigenous sources, but the presence of solar-wind hydrogen (hundreds of ppm) from the ilmenite may complicate the overall process into one involving methane as well.

Zhao and Shadman (1991) have studied the kinetics of the CO reduction of ilmenite. They have demonstrated that the rate of reduction of ilmenite by CO is slower than that with H_2, as expected. However, for a given temperature, the difference is less than an order of magnitude (Y. Zhou, personal communication). These are unexpected but very encouraging results. Therefore, the kinetics of CO versus H_2 reduction may not be crucial to selection of the ultimate process.

C. Ilmenite Reduction with Methane

The reduction of ilmenite by methane has been proposed by Friedlander (1985). It is probable that this process could take on a scheme similar to that for hydrogen or carbon monoxide reduction (Figs. 2 and 3). An equation which represents this reduction reaction can be written:

$$\mathrm{FeTiO_3 + CH_4 \Longleftrightarrow Fe + TiO_2 + CO + 2H_2}$$

$$\mathrm{2CO + 6H_2 \Longleftrightarrow 2CH_4 + 2H_2O}$$

$$\mathrm{H_2O \Longleftrightarrow H_2 + 1/2O_2}. \tag{4}$$

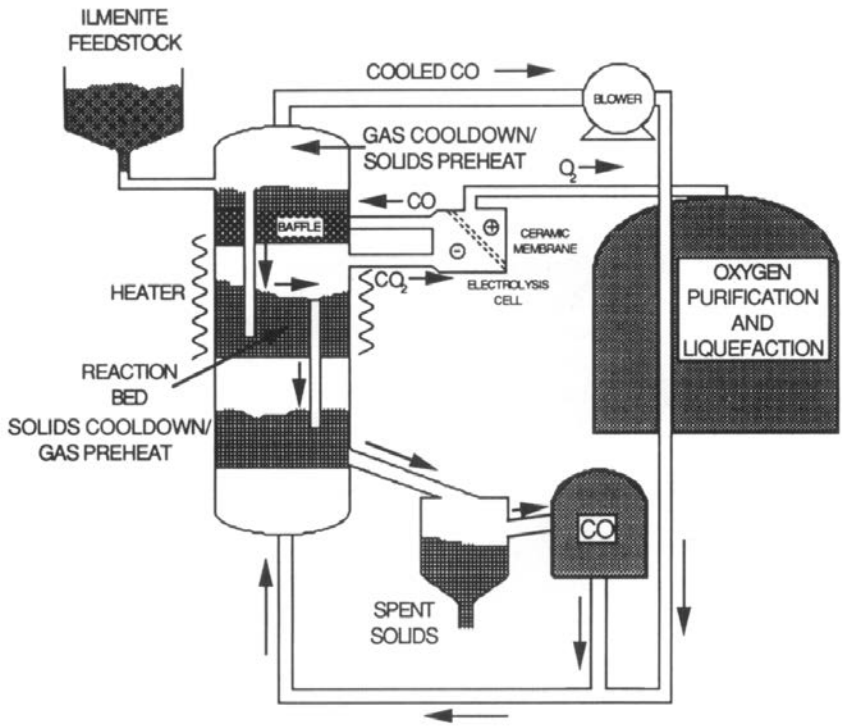


Figure 3. Carbon monoxide reduction of ilmenite for the production of LLOX, using fluidized-bed processing. The use of methane as the reductant, or the reduction of glass with hydrogen, could have the same configuration.

The CO and H_2 products are reacted with additional H_2 at 800 to 1000°C, using a nickel catalyst, in order to regenerate methane and produce H_2O. And the water is dissociated to produce O_2 and recyclable H_2. It is reported that 85 to 90% reduction of fine ilmenite (0.25–0.5 mm) was accomplished in a fluidized bed at 1000°C in only 5 to 7 min. It is suspected that this rate may be somewhat slower, based upon the kinetics of reduction of ilmenite (Hammond and Taylor 1982).

A fluidized-bed process, similar to the Carbotek scheme for reduction by hydrogen, would seem appropriate for the use of methane as the reductant. Any loss of reactant methane from the overall processing, or hydrogen/methane from the thermochemical reduction cell, possibly could be made up with indigenous carbon and hydrogen, present in the ilmenite, without the need for additional supply from Earth.

D. Glass Reduction with Hydrogen

Lunar glass, particularly from mare regions, can contain FeO contents up to 20 wt%. McKay et al. (1991) have recently demonstrated that this FeO

component of the glass can be readily reduced by hydrogen to yield elemental Fe° (native Fe) and water. Thermodynamically, the glass is considerably more unstable than the silicate minerals from which it formed; hence, the rapid kinetics of reduction. The water so produced can be easily hydrolyzed or otherwise electrolyzed to yield oxygen and recyclable hydrogen.

$$FeO_{(glass)} + H_2 \Longleftrightarrow Fe^{\circ} + H_2O$$

$$H_2O \Longleftrightarrow H_2 + 1/2O_2. \quad (5)$$

Glass is an abundant constituent of the lunar regolith. The lunar soil contains considerable glass, mostly as the result of melting produced by meteorite and micrometeorite impacts but also from volcanic activity (e.g., the orange soil at Apollo 17). The impact glass usually welds together rock and mineral fragments into aggregates called agglutinates (L. A. Taylor 1988); and these agglutinates frequently constitute over 50 vol% of a given mature soil. In addition, there are entire regions of the Moon which have blankets (1–4m deep) of volcanic glass as pyroclastic deposits. These are termed dark mantled deposits.

Although this scheme for oxygen production is a relatively recent development, it is not too soon to consider its feasibility. It may be possible to utilize a fluidized bed reaction vessel, similar to that proposed for the reduction of ilmenite with hydrogen (Fig. 2). Considerations of the possible feedstocks would depend upon the actual engineering of the process. It would be easiest to use a pyroclastic soil as it consists almost entirely of volcanic glass with a narrow range of grain sizes; an example is the orange soil from Shorty Crater. This would require no beneficiation; however, it is more likely that a mare soil will be used. The agglutinitic glass of typical high-Ti mare soil can be readily beneficiated by magnetic separation, as has been demonstrated by Taylor and Oder (1990). However, because glass sinters at much lower temperatures than the silicate minerals, it may be better to use unbeneficiated mare soil because the silicate mineral and rock fragments might hinder sintering of the glass particles, which could be a significant problem. An additional possible benefit from the use of largely unbeneficiated soil may be realized if the plant is actually located in a hi-TiO_2 mare. The feedstock would undoubtedly contain ilmenite, which may well have kinetics of reduction similar to the glass. This oxide phase would also undergo reduction and release its oxygen, thereby adding to the recovery/unit mass of feedstock.

E. Reduction with Hydrogen Sulfide

Dalton and Hohman (1972) proposed to reduce unbeneficiated lunar soil with hydrogen sulfide gas. In particular, the oxides of Fe, Ca and Mg are proposed to be reduced as shown below; this includes the recovery of sulfur, the electrolysis of the water, and cyclical production of reductant hydrogen sulfide:

$$MO + H_2S \Longleftrightarrow MS + H_2O \ (M = Fe, Ca, Mg)$$

$$MS + \text{Heat} \Longleftrightarrow M + S$$

$$H_2O \Longleftrightarrow H_2 + 1/2O_2$$

$$H_2 + S \Longleftrightarrow H_2S. \tag{6}$$

Although this process appears simplistic on paper, it is far more complicated. It is the silicate minerals in the lunar rocks and soils that contain all of the Ca and Mg and most of the Fe. And these silicates are not easily broken down by hydrogen sulfide—i.e., the kinetics are complicated and slow, even at 1000°C (L. A. Taylor, unpublished data). It would be difficult to handle soil at and above such high temperatures since the agglutinate (glass) component will readily sinter, thereby decreasing porosity and permeability. At higher temperatures, necessary for reasonable process yields, the soil will melt. In addition, because of the toxic nature of hydrogen sulfide, it would be necessary to purify the oxygen extensively, if it were to be used for life-support purposes. Of course, the problem of purification of the oxygen is applicable to most oxygen generation processes, particularly since troilite, the sulfide mineral, is ubiquitous in all mare rocks and soils, as well as present in many highland sources, albeit at much lower contents.

F. Extraction with Fluorine

Fluorine gas, F_2, is highly reactive and can be utilized to liberate metal from all oxides, thereby creating metal fluorides (Burt 1988,1992; Christiansen et al. 1988; Seboldt et al. 1991). The proposed feedstocks can be either anorthite or ilmenite in a two-stage fluidized bed reactor (Burt 1988). Using anorthite, the reaction which occurs in the first stage is:

$$CaAl_2Si_2O_8 + 2F_2 \Longleftrightarrow CaF_2 + Al_2SiO_4F_2 + SiO_2 + O_2(g). \tag{7}$$

Additional fluorine is fed into the bottom of the second stage:

$$CaF_2 + Al_2SiO_4F_2 + SiO_2 + 6F_2 \Longleftrightarrow CaAlF_5 + AlF_3 + 2SiF_4(g) + 3O_2(g). \tag{8}$$

The product gas is passed through a bed of NaF which effectively scrubs out the SiF_4 gas, but the resultant Na_2SiF_6 must be treated with Na to recover the NaF and produce Si. The $CaAlF_5$ and AlF_3 are treated with Na to yield CaF_2, NaF and Al metal. Additional steps are required to recover reactants and release additional oxygen. In all, some 8 steps are required, using several separate reactors, and each with its own operating conditions. Burt (personal communication) claims that the number of steps could probably be reduced to 4, with some sacrifice of efficiency.

The chemical reactivity of fluorine is well established. Unlike the reduction processes discussed above (e.g., H_2, CO), the extraction with fluorine liberates oxygen directly without the need for a thermochemical reduction to split off the oxygen from the reductant (e.g., H_2O, CO_2). Instead, the SiF_4

gas must be totally scrubbed out, and the oxygen must be depoisoned for human consumption. The feedstock requirements will require considerable beneficiation of the rocks/soils from the highlands (anorthite concentrate) or high-Ti mare (ilmenite).

The overall process of extraction with fluorine is complicated, involving many steps and several reactors. The process implications would seem impractical at this point. Complete recovery of fluorine would seem unlikely, and the corrosive nature of fluorine will require special inert materials, necessitating a large Earth supply involvement. Considerable technology development is required, and process mass and energy requirements are not known for comparative purposes. Lastly, the considerable danger of working with fluorine or metal fluoride gases is a major safety consideration.

G. Carbochlorination

The carbochlorination process involves using a CO-Cl_2 gas mixture or Cl_2 in the presence of solid carbon to react with anorthite and/or ilmenite (Bhogeswara 1979; Christiansen et al. 1988). A fluidized-bed reactor operating at 770°C was proposed to react:

$$CaAl_2Si_2O_8 + 8C(s) + 8Cl_2 \Longleftrightarrow CaCl_2 + 2AlCl_3(g)$$

$$+2SiCl_4(g) + 8COFeTiO_3(s) + C(s) + 3/2Cl_2(g) \Longleftrightarrow FeCl_3(g)$$

$$+TiO_2(s) + CO(g). \qquad (9)$$

Several staged condensation, hydrolysis, and electrolysis steps would be necessary to separate gaseous components—$AlCl_3$, $SiCl_4$, $FeCl_3$, Cl_2 and CO, and to recover reactants and oxygen. Chlorine gas at temperatures <1000°C normally only reacts with oxides, such as ilmenite. In order to facilitate effective reaction with the silicate minerals, it is necessary to have carbon present. The C acts as a reducing agent, forming CO, while the Cl_2 oxidizes the metal, forming a volatile chloride. In this manner, a new surface is continually exposed for reaction (Lynch 1989).

There are a large number of processing steps involved in this overall scheme; such complexities add problems with regards to system reliability, efficiency of reactant recovery, etc. Chlorine is a hazardous gas to work with. In addition, Lynch (1989) has identified some 136 C-Cl-O by-product compounds which may be produced during carbochlorination. The efficient recovery of carbon and chlorine reactants is unlikely, even using a large number of processing steps with large mass and energy requirements. On the positive side, this process would seem to be capable of using bulk lunar soil, either mare or highlands.

H. Chlorine Plasma Reduction

Lynch (1989) has recently advocated the use of a "cold" plasma reactor to create a chlorine plasma. Calculations and preliminary experiments suggest

that in such a plasma, stable metal oxides (e.g., $FeTiO_3$–ilmenite) can undergo chlorination and yield oxygen as a by-product. The term "cold" plasma is misleading in that the plasma is only relatively cooler than a normal plasma. Molecular temperatures of up to 2000°C are encountered. The term "cold" indicates a significant difference between the molecular temperature and the kinetic temperatures of the electrons in the molecule, which can be thousands of degrees hotter than the overall temperature of the molecule. The extremely reactive nature of the cold plasma is a function of these temperature differences between the electrons and the overall molecule. It is the presence of broken bonds, partially filled orbitals, and unbalanced charges that makes these molecules so highly reactive. The basic process can be written:

$$Cl_2 + MO \Longleftrightarrow MCl_2 + 1/2O_2. \tag{10}$$

After the oxygen has been released from the metal oxide, it is necessary to remove the chlorine from the metal chloride. This can be accomplished by electrolysis. Experimentation is currently in progress (Lynch 1989). The kinetics of chlorination are being examined as a function of temperature, chlorine content in the gas, total pressure, power levels, and plasma density. At this point in time, it is not possible to evaluate thoroughly the advantages and disadvantages of the chlorine plasma reduction process.

V. SILICATE/OXIDE MELT

Bulk rock and soil with their silicate minerals—for example, anorthite [$CaAl_2Si_4O_8$], olivine [$(Mg,Fe)_2Si_2O_4$], pyroxene [$Ca(FeMg)Si_2O_6$], and oxide minerals—much as, e.g., ilmenite [$FeTiO_3$], chromite [$FeCr_2O_4$]—in a molten state can be acted upon to produce oxygen. This can be achieved by electrolysis or pyrochemical techniques. Derivation of oxygen by molten silicate (magma) electrolysis has been amply demonstrated experimentally in silicate melts of a variety of compositions (see, e.g., Bockris et al. 1952*a*, *b*; Simnad et al. 1954; Oppenheim 1968,1970; Kesterke 1971; Lindstrom and Haskin 1979). Fluoride flux also has been added to the silicate melt to lower the temperatures of melting and to increase the efficiency of the electrolysis process (see, e.g., Keller 1986,1988; Keller and Taberaux 1991). Other schemes for utilizing silicate and oxide melts involve the addition of various reducing agents (e.g., NaOH, C, CO) to breakdown the molten components to produce oxygen (see, e.g., Dalton and Hohman 1972; Rosenberg et al. 1966).

A. Molten Silicate Electrolysis

In its conceptually simple form, electrolysis entails immersing two electrodes in a vat of molten silicate and imposing a current between the electrodes (Fig. 4). Oxygen is derived at the anode and metal (mostly Fe and Si) at the cathode. No moving parts or reagents are needed, and the one-step process takes place in a single reaction pot. A design for a molten silicate electrolysis

cell by McCullough and Mariz (1990) is shown in Fig. 5, where the anodes are at the upper portion of the furnace and the cathodes are at the bottom. As can be seen, this cell is designed for a continuous mode of operation, rather than by batch mode.

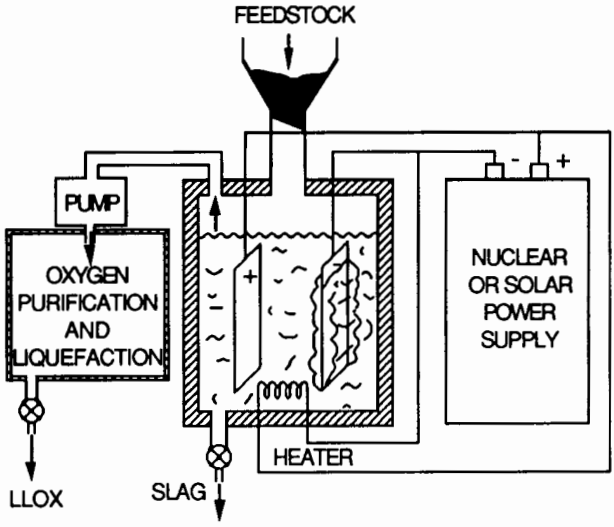


Figure 4. Schematic diagram of the electrolysis of molten silicate depicting the production of oxygen at the anode (negative electrode) and various metals at the cathode (positive).

The energy requirements of the electrolysis process depend both on the composition of the silicate melt and the design of the electrolysis cell. The fraction of the current that goes towards producing oxygen (O_{2eff}) depends primarily on the concentrations of FeO in the melt (Haskin et al. 1992). The presence of FeO substantially increases the energy required to produce oxygen because Fe^{2+} may be oxidized by anode products to Fe^{3+}; that is, FeO decreases the O_{2eff}. Efficient oxygen production appears to require that the melt be less than about 5 wt% FeO. Electrical conductivities of silicate melts have been measured in simple metal oxide-silica systems (Bockris et al. 1952*a*, *b*); temperature dependence has been determined (du Fresne and Schroeder 1983); the compositional dependence has been measured and modeled (Haskin et al. 1992). Conductivities in melts of interest for silicate melt electrolysis range over more than 2 orders of magnitude, with conductivities being lowest (<0.2 $cm^{-1}ohm^{-1}$) at low temperature in melts rich in SiO_2 and Al_2O_3 and highest (>2 $cm^{-1}ohm^{-1}$) at high temperature in melts rich in MgO, CaO, and FeO.

Effects of compositional parameters on the energy required to produce a given amount of oxygen by electrolysis of silicate melt are complex but

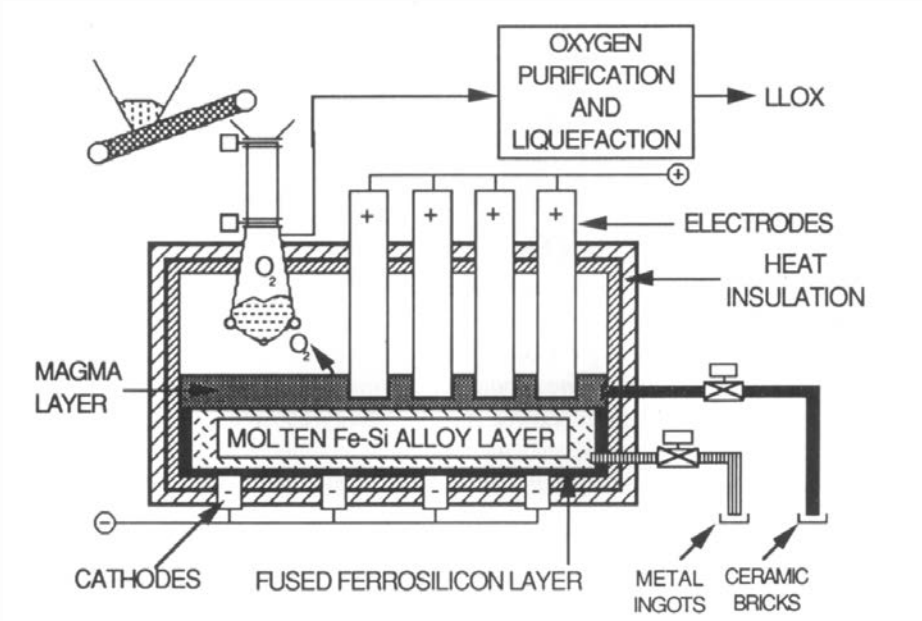


Figure 5. Magma electrolysis cell design for the production of LLOX from lunar regolith, modified after McCullough and Mariz (1990).

known to a first approximation (Colson et al. 1991). In general, the energy required increases with increasing concentrations of SiO_2, Al_2O_3, and FeO in the melt. The energy required decreases as the size of the electrolysis cell increases (increasing electrode surface area) and as the distance between electrodes is decreased. Additional compositional constraints are placed on the steady-state melt by the need to maintain container and electrode materials in equilibrium. Composition can be controlled through variations in the feedstock composition, rate of feedstock flow-through, and electrode potential.

Evaluation of Molten Silicate Electrolysis Process. The number of process steps and equipment is low. Oxygen is liberated directly with no need for further refinement (versus H_2O or CO_2 products which require reduction). There are no anticipated major mass penalties to be supplied from Earth because there is no need for reactants. Electrode life is still an unknown. The production of iron may be possible, albeit impure. Haskin et al. (1992) have calculated that the energy requirements for the molten silicate electrolysis process are less than for most alternative processes (Colson and Haskin 1990; see also their Chapter).

Because of limited industrial experience with high-temperature silicate melts, the actual design of a working electrolysis cell is less well determined than the physical-chemical properties of the melts. Experimental design in most experiments (see, e.g., Colson 1990; Kesterke 1971; du Fresne and Schroeder 1983) cannot be simply scaled up. Also, at the high temperatures

required for this process (1300–1700°C), anode stability and corrosion problems are significant and must be overcome. Colson et al. (1991) discussed the considerable advances which have been made with regards to possible container and electrode materials (Colson and Haskin 1990; Haskin et al. 1991; McCullough and Mariz 1990). However, substantial study of high-temperature refractory materials in silicate melts is needed before appropriate container and electrode specifications can be chosen.

B. Fluxed Molten Silicate Electrolysis

The employment of a flux, such as a fluoride melt, to dissolve the silicate feedstock can alleviate some of the difficulties of high-temperature corrosion experienced by the simple molten silicate process. Operating temperatures are decreased and electrolyte conductance is increased, versus molten silicate electrolysis without flux. Lower specific energy consumption can be projected. Fluxed electrolysis has the additional advantage of being similar to electrolysis processes currently used to produce aluminum on Earth (see, e.g., Macmillan 1987).

Initially, the fluxed molten silicate process was aimed at recovery of Si and Al and was only secondarily concerned with oxygen (Keller 1986,1988). This process used Al as a reactant, in addition to fluoride flux. Recently, a fluoride fluxed process without Al has been applied directly to the production of oxygen (Keller and Taberaux 1991). This has permitted a somewhat simpler scheme to be implemented. Because this latter scheme is more basic, it will be described first.

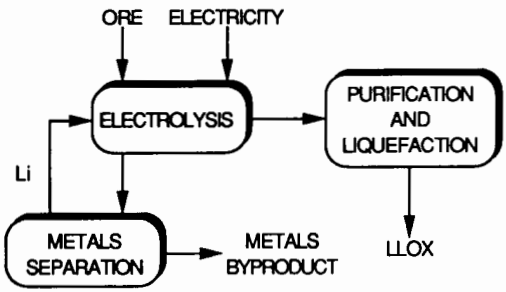


Figure 6. Flow diagram for LLOX production by electrolysis of fluxed molten silicate.

In the process recently being pursued experimentally by Keller and Taberaux (1991), lunar soil is added to a LiF-CaF_2 electrolyte at about 1000°C. The electrolysis decomposes all added component oxides. Conditions are maintained such that Ca is reduced at the cathode, along with all the other metallic components of the feed. Oxygen is liberated directly at the anode. Unfortunately, some Li also is co-reduced and has to be recovered in an auxiliary process step. The approach is summarized in Fig. 6. It is designed to utilize

a broad spectrum of feed material without any or only minimal beneficiation.

The fluoride-fluxed approach was initially investigated by Keller (1986, 1988) to obtain Si and Al, as well as oxygen. An early concept of the process is depicted in Fig. 7. Anorthite, $CaAl_2Si_2O_8$, is dissolved in a molten mixture of fluorides at about 1000° C. Aluminum is added to the melt and reduces the silica to Si:

$$3CaAl_2Si_2O_8 + 8Al \Longleftrightarrow 6Si + 3CaO + 7Al_2O_3. \qquad (11)$$

When all of the available silica has been reduced to Si, the Si is recovered from a hyper-eutectic Si-Al alloy by cooling the alloy melt to 700°C and filtering the solid Si formed. The electrolyte which now contains alumina and calcium oxide is pumped into an electrolysis cell where the oxides are reduced. Electrolysis results in the formation of oxygen at the anode and Al and Ca at the cathode. Auxiliary process steps are conceived to beneficiate the ore, to separate pure Si, to recover electrolyte components, and to purify the off-gases. This approach has been refined to the schematics of Fig. 8 (Keller 1988).

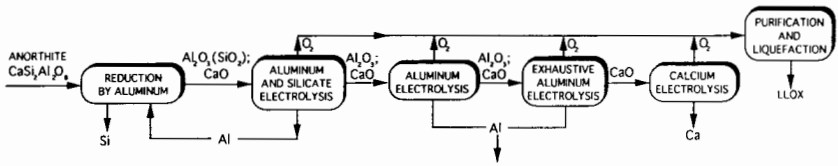


Figure 7. Flow diagram for fluxed molten silicate electrolysis to produce LLOX by the initial addition of Al. This scheme was modified from Keller (1986) and Anthony et al. (1992).

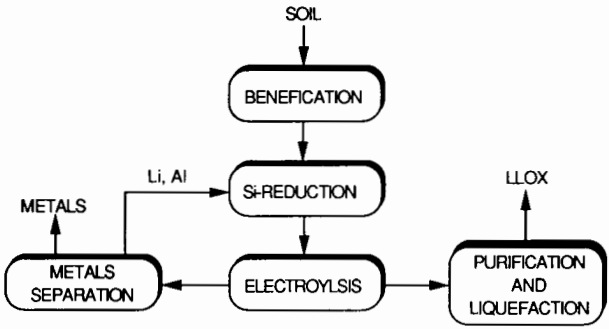


Figure 8. Flow diagram to produce LLOX, as well as Si and Al, by fluxed molten silicate electrolysis by the initial addition of Al.

Evaluation of Fluxed Molten Silicate Process. Most of the advantages mentioned under the molten silicate electrolysis process are equally applicable here. However, the fluxed-melt process, in practice, may have lower energy requirements and lesser corrosion problems, both of which are mainly due to lower reactor temperatures. Importantly, oxygen is recovered directly at the anode. And the process can be expanded to produce valuable metals. The one-step nature of this process should aid in attainment of a continuous mode of operation. The use of unbeneficiated lunar soil for feedstock and versatility of feedstock requirements are important assets of the fluxed molten silicate electrolysis process. On the other hand, the reduction of anorthite by Al to yield Si and oxygen, although demonstrated experimentally (Anthony et al. 1992), necessitates two or more major steps which adds significantly to the complexity of this modification of the flux process. And the need for an anorthite feedstock will require beneficiation of highland soil.

Many of the disadvantages for the molten silicate electrolysis process apply here as well. The problems to be solved are dominated by the necessity for a stable oxygen-evolving anode. The simplicity of the overall scheme is compromised by the need to recover the fluxing reagents (Keller 1988) so that resupply mass penalties are kept to a minimum. The makeup of fluoride flux by Earth supply may be prohibitive. There is a need for additional research in order to quantify yields and optimize operating conditions, including an efficient way to recycle electrolyte components back into the system. It would appear that the simple fluxed molten silicate process, versus the one using the addition of Al, will be the scheme to be researched more thoroughly in the future.

C. Caustic Solution and Electrolysis

In this process, molten NaOH at 400°C is used to dissolve minerals from bulk lunar soil (Dalton and Hohman 1972). This caustic molten solution is then electrolyzed to yield oxygen at the anode and sodium at the cathode. The sodium then reduces the oxides in solution, producing metals and Na_2O. This process can be made continuous if the mixed caustic and solid products are withdrawn from near the cathode of the reactor, and another unit is used to separate and recycle the caustic agent (NaOH) to the pre-electrolysis solution tank.

Except for magnesium and calcium oxides, reduction of the other oxides appears possible, thus resulting in a high oxygen yield. With considerable development of several additional steps, it may be possible to separate the mixed metal product into its constituent metals. The recovery of the caustic reagent from reactor residual solids may require centrifuges and dryers. It is anticipated that the mass penalties from caustic loss will be substantial. But, it is possible that caustic loss can be supplemented with indigenous sodium present in the mineral plagioclase. Considerable research would be required for this process, particularly the electrode stability and the long-term performance of the electrolyzer cell. Using KOH as an alternative flux for this

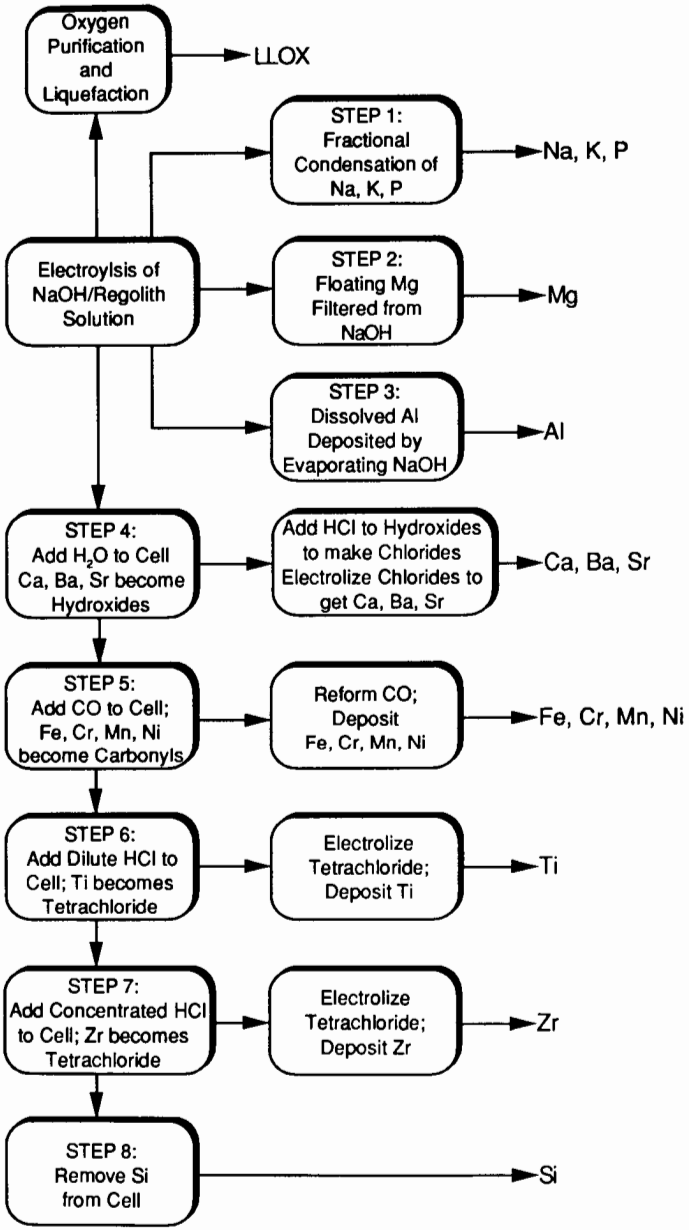


Figure 9. Flow diagram for caustic dissolution and electrolysis with the production of LLOX and metals, modified after Binder (1990).

process, as proposed by Cutler (1984), would face most of the same problems. A variation of this caustic solution process has been proposed by Binder (1990). As shown in the flow diagram in Fig. 9, this scheme involves several steps involving fractional condensation, flotation and filtration, evaporation, hydration, electrolysis, etc. In this complicated fashion, it is proposed to be capable to separate all of the metals and oxygen from the lunar regolith.

Waldron and Criswell (1982) proposed the use of NaOH in a leaching process that would entail a complicated series of non-electrolytic, pyrochemical steps for oxygen production and recovery of metals and the NaOH reactant. It would appear that this process and the other caustic solution processes, of Dalton and Hohman (1972) and Binder (1990) described above, consist of several complex steps, each with its own specifics, including temperature. The complex, multi-step nature of these schemes, coupled with the hazardous nature of the NaOH, is a distinct disadvantage which may make this overall process impractical.

D. Carbothermal Reduction

Schemes for the production of oxygen using molten reactant and involving carbon in some form are called carbothermal processes. These combine chemistry from steel-making and from coal-gas forming with electrolysis or thermolysis of water. Although these processes operate best on melts of simple oxide minerals, like ilmenite ($FeTiO_3$), it may be possible for the carbothermal reduction of melts of anorthite ($CaAl_2Si_2O_8$), but at extremely high temperatures (2000° C; Bhogeswara et al. 1979).

The reduction of molten silicates and oxide phases by carbon in various forms (i.e., C, CH_4, CO) was the subject of the pioneering work of Rosenberg (1985) and Rosenberg et al. (1966). Experiments were performed on the reduction of molten silicates by methane at about 1600°C according to the following process chemistry:

$$\begin{aligned} Mg_2SiO_{4(l)} + 2CH_4 &\Longleftrightarrow 2MgO + Si + 4H_2 + 2CO \\ CO + 3H_2 &\Longleftrightarrow CH_4 + H_2O \\ H_2O &\Longleftrightarrow H_2 + 1/2O_2. \end{aligned} \tag{12}$$

The feedstock mineral considered above is olivine, but other phases such as pyroxene [$Ca(Mg,Fe)Si_2O_6$] should work equally well. In other words, it may well be possible to utilize unbeneficiated soil. The second reaction demonstrates the hydrogenation of the carbon monoxide, using a nickel catalyst, to recover the reactant methane and water. And the water can be electrolyzed to release oxygen. Carbon or carbon monoxide, in place of methane, can also be used as the reductant. In fact, with methane, it seems that the CH_4 actually cokes to $C + 2H_2$, and the carbon is the reductant according to the equations:

$$CH_4 \Longleftrightarrow C + 2H_2$$

$$Mg_2SiO_{4(l)} + 2C \Longleftrightarrow 2MgO + Si + 2CO. \quad (13)$$

There is the possibility that some SiC may form in this process, thereby tying up carbon and placing a penalty on reactant recovery (R. Keller, personal communication). However, S. Rosenberg (personal communication) reported that none was found in residual melts from such experiments.

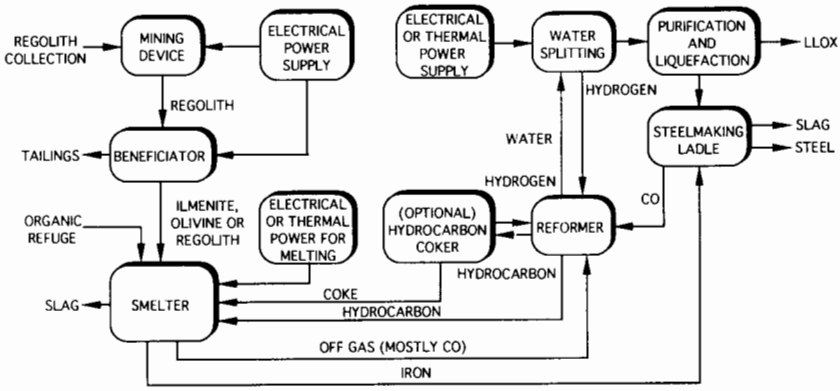


Figure 10. Flow diagram of the production of LLOX by the carbothermal reduction process, modified after Cutler and Krag (1985).

Cutler and Krag (1985) have proposed an oxygen and iron production scheme using coke (devolatilized carbon) to reduce molten ilmenite. As shown in Fig. 10, the process includes 3 major steps: ilmenite smelting, iron decarburization (steel making; middle right in Fig. 10), and hydrocarbon reforming. The product water is electrolyzed to yield oxygen and recyclable hydrogen. Ilmenite is melted (1400°C) in the smelting step and reacts endothermically with the carbonaceous materials (refuge of Fig. 10) to form iron by the following reaction:

$$FeTiO_{3(l)} + C(s) \Longleftrightarrow Fe(l) + CO(g) + TiO_2(s)$$

$$CO + 3H_2 \Longleftrightarrow CH_4 + H_2O$$

$$H_2O \Longleftrightarrow H_2 + 1/2O_2. \quad (14)$$

Approximately 10% anorthite is added as a flux to form a slag with a melting point below iron. Effectively, the ilmenite (90%) + flux (10% anorthite) + carbon-bearing solids form FeC_x + CO + TiO_2. The carbon-bearing Fe is decarburized by controlled addition of oxygen to form CO. The CO is reacted with hydrogen to form water and methane. The hydrocarbon is cracked to yield carbon and hydrogen, and the water is electrolyzed to yield oxygen and hydrogen, which is recycled. The several steps in this process have been well established in the steel industry, but operate under rather extreme conditions.

Evaluation of Carbothermal Reduction Process. The scheme as proposed by Rosenberg et al. (1966) can utilize mare regolith, largely unbeneficiated. Silica and ferrous oxide, originally present in the feedstock in olivine and pyroxene, as well as in ilmenite, are reduced. Theoretically, this should yield much more oxygen per unit of regolith than for several other processes—e.g., ilmenite reduction with hydrogen. However, this advantage occurs at the sacrifice of system complexity. The complexity of this overall process may severely compromise its feasibility. Note, however, that the complete cyclic process has been reduced to practice in a batch process. As much as one pound of granite or basalt has been processed in each batch (S. Rosenberg, personal communication).

The high temperatures ($\cong$ 1600°C) needed for the reactor and several of the steps will require considerable energy input. The high temperatures and the corrosive nature of the melt will also take a toll on reactor linings. The reductant makeup from Earth may be considerable, estimated at 5 to 20% of the carbon. However, indigenous solar-wind implanted carbon in lunar soil (average = 20 ppm) could offset some of this supply. In addition, although steel is a necessary by-product, steel making is a batch process, at least on Earth. Therefore, the automation for a continuous process will be difficult. Much of the carbothermal process can be evaluated with present knowledge. Terrestrial counterparts exist for many of these steps (e.g., smelters, steel making, hydrocarbon reforming), and considerable process experience exists.

E. Magma Partial Oxidation

Waldron (1989) has presented a five-step process which starts with an Fe-rich mare basalt rock or soil. The principle is to extract the FeO from the rock by first getting it all into a magnetic phase which can be separated from the remainder of the rock components. In order to accomplish this, the rock and/or soil is melted in the presence of 1 atmosphere of oxygen. After controlled cooling to promote crystallization, the solid is pulverized, and the mineral magnetite (Fe_3O_4) is extracted magnetically. This is dissolved in an "aqueous mineral acid" which is then electrolyzed to recover Fe and O_2. Although this scheme would seem to be straightforward, no experiments have been performed to determine the feasibility of such a process. Indeed, the beneficiation of the solidified silicate melt to release the magnetite might be difficult if the oxide is fine-grained and/or closely intergrown with silicate phases, both likely possibilities.

F. Li or Na Reduction

Semkow and Sammells (1987) proposed an indirect electrochemical reduction of lunar oxides using lithium (or sodium) to reduce oxides to metal and Li_2O. Either bulk lunar soil or mineral separates, including ilmenite, could be utilized. A temperature of 727°C was proposed for ilmenite. Lithium will reduce FeO, TiO_2 and SiO_2, but not Al_2O_3, CaO or MgO. The reaction product would consist of metals, the unreduced oxides, and Li_2O. A sublimation step

conducted at 700°C would separate the Li_2O, which would then undergo a reduction step in an electrolytic cell at about 900°C, requiring LiF and LiCl as fluxing, thereby lowering viscosity and resistivity. The liquid Li would thereby be recovered at the cathode with the evolution of oxygen at the anode.

This process reportedly can utilize any bulk regolith; thereby, the location of the plant would be versatile, i.e., non-site specific. Theoretically, the yield of oxygen by this process would be high, because lunar regolith typically contains about 45 wt% SiO_2. On the negative side, the complete recovery of Li_2O from the Li reduction reactor's solid product will be difficult, if not impossible, and require substantial sublimation at vacuums of about 0.02 torr. The electrolytic cell, to be run at 900°C, would need further research in order to evaluate cell corrosion, general degradation of materials (particularly the anode and cathode), flux loss, and effects of long-term operation. The mass penalty for Li, LiF and LiCl makeup supply from Earth may be large.

VI. PYROLYSIS

Pyrolysis is the application of heat to induce chemical change, usually a partial decomposition of metal oxides association with vaporization, as depicted in Fig. 11. The processes within this category involve high temperatures in the range of 2000 to 10,000°C. Such temperatures can be generated in a plasma torch, microwave plasma, electron beam, solar furnace, or related systems. These processes can be made virtually independent of Earth-supplied reagents. In theory, it would seem possible to recover most of the O_2 from the feedstock. For purposes of presentation, we have divided pyrolysis into low- and high-temperature processes—vapor-phase reduction and ion (plasma) reduction, respectively.

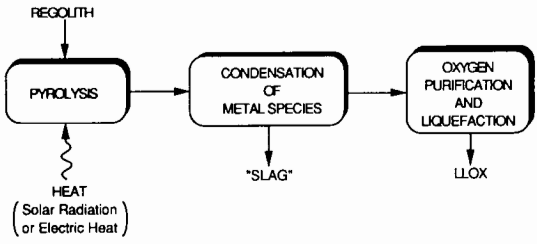


Figure 11. Flow diagram for the pyrolysis of lunar soil for the production of LLOX.

A. Vapor Phase Reduction

With vapor pyrolysis, high temperatures (2200 to 2700°C) are utilized to vaporize the feedstock (Fig. 12) and transform oxygen-bearing compounds into monoxides and oxygen (Boundy 1983; Steurer and Nerad 1983; Chapter

by Senior). After vaporization, the gas is rapidly cooled, so that everything except the oxygen is condensed back into a liquid or solid. This condensation step is critical because it must be accomplished before the metal species can recombine with the oxygen. Long-term fouling of condenser surfaces will lower process efficiency. Senior (see her Chapter) has demonstrated the feasibility of the pyrolysis portion of this process using a solar furnace and anorthite and ilmenite. However, unbeneficiated regolith can also be used as feedstock. The direct use of solar energy is a key benefit to this process.

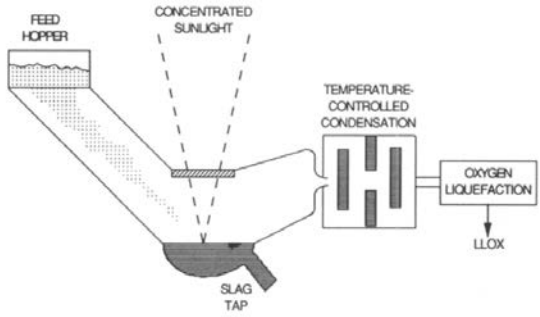


Figure 12. Schematic diagram for the production of LLOX by vapor phase pyrolysis, modified after notes of C. Senior.

This process is characterized by its total reliance on space resources, namely "hard" vacuum and solar energy, thereby eliminating the need for consumables from Earth (Steurer and Nerad 1983). Vacuum reduction and distillation of metals are well-known terrestrial processes. The condensation step in this process is in need of considerable study, although vacuum coating technology should be applicable here (Chapter by Senior).

B. Ion (Plasma) Separation

At temperatures of 7000 to 10,000°C, oxide dissociation products are ionized (Steurer and Nerad 1983). At 7700°C, over 90% of the metallic dissociation products (Fe, Ti, Al, Mg) and 25% of the Si are ionized. However, only 1% of the oxygen is ionized. The highly ionized metals are extracted from the vapor by electrostatic or electromagnetic fields, whereas neutral oxygen flows downstream for collection. In theory, this has the advantage of higher oxygen yields than the vapor reduction process. The condenser problem will be significant because considerable quantities of non-ionized metals and non-ionized Si must be removed.

Unlike the vapor phase reduction process where a solar furnace is possible, this scheme with temperatures approaching 10,000°C will be extremely energy intensive. Considerable problems with containment materials are to be

expected. Because the concepts for this process represent largely theoretical efforts not substantiated by experimental work or terrestrial analogs, it would seem that the vapor phase reduction process is the preferred pyrolysis scheme for further development at this time.

C. Plasma Reduction of Ilmenite

Allen et al. (1988) pointed out that most of the above processes for oxygen extraction from ilmenite are inefficient in that they usually recover less than 30% of the available oxygen. This is primarily because of the relatively low temperatures utilized. Allen et al. (1988) proposed a scheme involving processing of ilmenite at much higher temperatures; considerations for their calculations were at temperatures of 3000 to 6000°C. At such high temperatures, the ilmenite is completely dissociated into elemental iron, elemental titanium and oxygen. These components must be selectively condensed, which is no easy task. Although their experiments were conducted using a plasma torch, it may be possible to fashion a power source from a solar furnace. It should be emphasized that this process is still in the experimental stage, with considerable development remaining to be accomplished. However, plasma processing as applied to metallurgy has received significant study. In theory, this process has promise for oxygen and metal production from ilmenite.

VII. AQUEOUS SOLUTIONS

A. HF Acid Dissolution

Waldron and Criswell (1979,1982) and Waldron (1985) described a hydrofluoric acid-leach process (Fig. 13), wherein lunar regolith is dissolved to create metal fluorides and water. Several batch-mode, acid-leach reactors would produce steam and SiF_4 and precipitate the metal fluorides. Fluorine and HF are recovered from the metal fluorides in a complex procedure with multiple unit operations involving high-temperature hydrolysis (>1000°C), electrolysis, ion exchange, distillation, centrifuges and drying steps. This scheme is presented here as an example of just how complex one overall process can become.

Plant location for this process would not be site-specific, because any feedstock can be used. Calculations suggest that 586 tonnes of plant mass will be required to produce 1000 tonnes of oxygen per year. This is high in comparison to some of the processes reviewed above. Unfortunately, the disadvantages to this HF acid dissolution and electrolysis process are many (Christiansen et al. 1988). The acid-leach reactors are operated in batch mode, making automation difficult, if not impossible, and some reactant HF is lost, and fresh HF must be added during each new batch set-up. The large number of separate steps in the process require many small pieces of equipment. This will require greater design, development and evaluation costs than processes with fewer but larger unit operations.

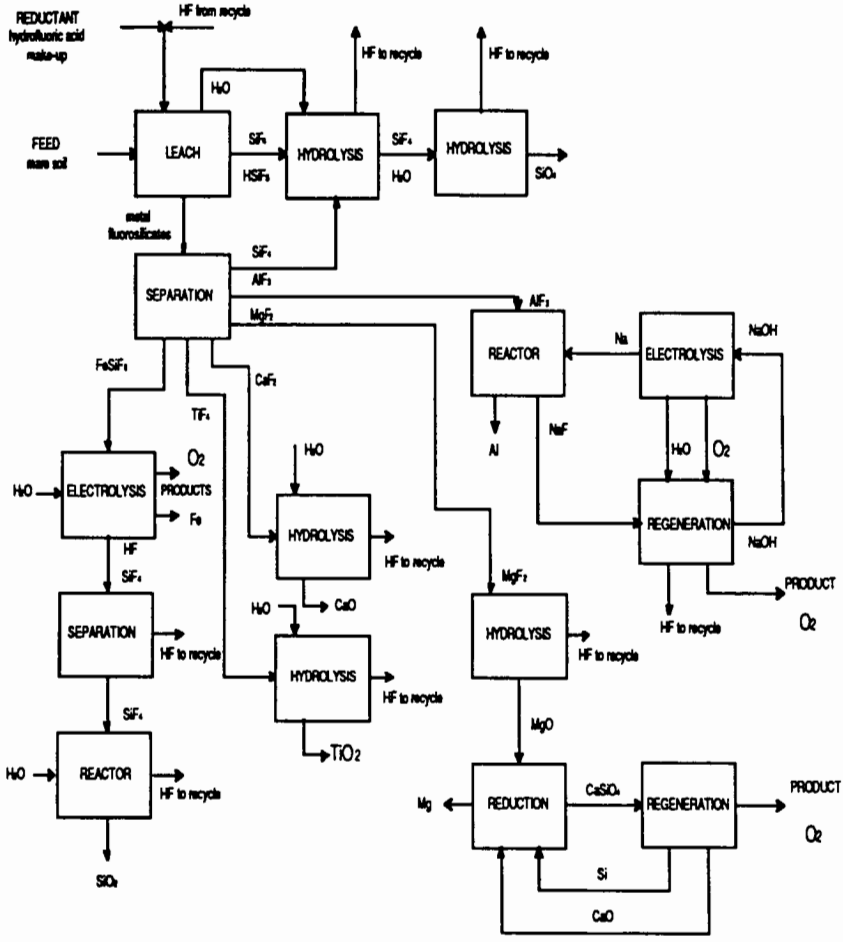


Figure 13. Flow diagram for the HF acid leaching and electrolysis of lunar soil for the production of LLOX, modified after Altenberg (1990). Note the lack of simplicity to the entire operation.

B. H_2SO_4 Acid Dissolution

Sullivan (1990) has recently proposed a sulfuric acid dissolution process to digest an ilmenite feedstock, according to the reaction:

$$FeTiO_3 + H_2SO_4 \Longleftrightarrow FeSO_4 + H_2O + TiO_2. \tag{15}$$

The resultant slurry is filtered to remove any unreacted solids (e.g., silicates), and the clean solution is passed through an electrolysis cell wherein oxygen and iron are separated by the reaction:

$$FeSO_4 + H_2O \Longleftrightarrow H_2SO_4 + Fe + 1/2O_2. \tag{16}$$

Elements of this process are based on well-known commercial practices and procedures; but these industrial processes are aimed at recovering Fe and TiO_2, not oxygen. Experiments are currently under way at NASA Johnson Space Center to test the feasibility of this somewhat complex process.

VIII. CO-PRODUCT RECOVERY OF WATER

The lack of a shielding atmosphere on the Moon permits solar-wind particles to impinge upon the lunar soil and become implanted into the various phases which comprise the soil (L. A. Taylor 1990*b*). In particular, relatively large quantities of solar-wind implanted hydrogen (50–100 ppm) and helium (10–50 ppm) are present (Carter 1985). And ilmenite acts as a sponge for these solar-wind components, retaining 5 to 10 times more than the silicate minerals. It is probable that these volatile components of the lunar soil will be collected on the Moon (Bustin and Gibson 1992). And thermal recovery of solar-wind H_2 from lunar soil liberates water from the soil as well.

A. Hydrogen/Helium/Water Production from Soil

Roasting of lunar soil to temperatures of 600°C releases approximately 80% of the hydrogen; with heating to 900°C, complete release is accomplished (Bustin and Gibson 1992). The hot hydrogen can react with oxide minerals, ilmenite and spinels, to produce water, much the same as in the ilmenite reduction with hydrogen process, discussed above. However, the kinetics for this reaction are too slow to have much of the hydrogen/ilmenite reaction occur in only seconds to minutes at 600°C or so. Instead, most (up to 90%) of the water recovered in the step-wise heating experiments on lunar soils is probably of terrestrial origin, i.e., contamination from the "wet" N_2 in which all lunar samples are stored.

Some of the water released may well be lunar in origin. Gibson (personal communication) has performed an extensive evaluation of H_2 and H_2O released during step-wise heating experiments on lunar soils. Indeed, he feels that some of the water released is not from contamination. It is possible that some of the abundant solar-wind protons in the outer skin of the ilmenite actually form hydroxyl units which diffuse out during the heating. The amount of lunar water so formed is probably only equivalent to about 10% of the mass of the hydrogen released, only some 5 to 10 ppm at most. And electrolysis could be used to dissociate this water into H_2 and O_2.

Christiansen et al. (1988) were the first to consider the roasting of lunar soil as a possible source of water/oxygen. However, the quantities of soil which would have to be handled are tremendous, compared to all other techniques for oxygen production. This is well illustrated in Fig. 14, which is probably too optimistic for this process, i.e., the curve should be even steeper. If oxygen is the primary product sought and this thermal release process is to be used, it is obvious that it would not be feasible to mine the lunar soil.

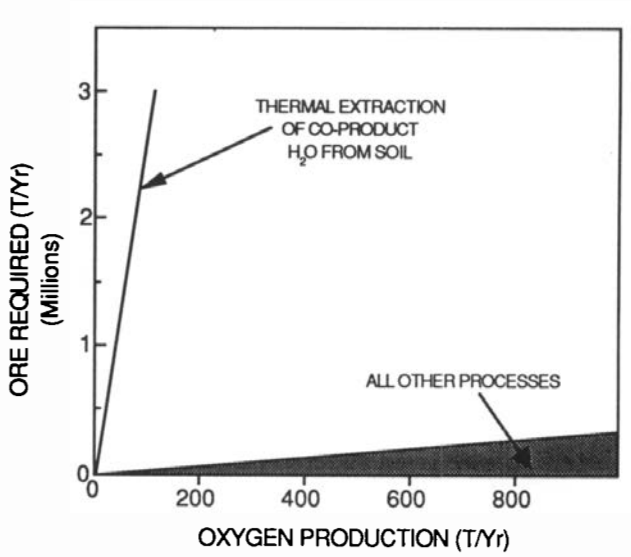


Figure 14. LLOX production versus the estimated mass of lunar regolith required, based upon estimations discussed in the text.

However, with the mining of the soil for the recovery of hydrogen, water (i.e., oxygen) could be a valuable co-product.

There are several advantages to co-product production of oxygen from the hydrogen/helium recovery process (L. A. Taylor 1990*a*,1991). Both hydrogen and oxygen (as well as helium) can be obtained by roasting to moderate temperatures of 600 to 900°C. There should be no complicated development costs involved in a stationary unit, although movement of large quantities to this site will pose major problems. However, a mobile processor will need power-source considerations, possibly involving direct use of solar energy. Probably the biggest disadvantage to this process is the huge amount of soil which will have to be handled, both mechanically and thermally. And the presence of abundant ilmenite is requisite for the liberation of the water. This would require a high-Ti mare soil, thus making this process site-specific.

IX. EVALUATION AND RANKING

Twenty potential processes for the production of lunar oxygen have been reviewed in the preceding pages. Table I lists the processes and the important references. Table II summarizes the feedstock required for each process; and Table III is a qualitative comparison of each of the potential processes, in terms of technology readiness, the number of major steps, the process conditions, and feedstock requirements. It must be thoroughly appreciated that there is considerable subjectivity in these rankings, and even that the four different factors themselves are not really of identical importance. Nevertheless, a

ranking of processes was accomplished. Based upon these rankings in Table III, plus the evaluations presented throughout this chapter, 12 processes are considered to have lesser potential for developing into the selected scheme for oxygen production on the Moon. The 8 candidates considered to be the more feasible processes are listed in Table IV. These are given further considerations, mainly from their engineering aspects.

TABLE II
Feedstocks for Oxygen Production on the Moon

Process	Feedstocks/Locations
Solid/Gas Interaction	
Ilmenite reduction with hydrogen	Ilmenite—Hi-Ti mare
Ilmenite reduction with C/CO	Ilmenite—Hi-Ti mare
Ilmenite reduction with methane	Ilmenite—Hi-Ti mare
Glass reduction with hydrogen	Mare or pyroclastics
Reduction with hydrogen sulfide	Mare or pyroclastics
Extraction with fluorine	Mare or pyroclastics
Carbochlorination	Mare or pyroplastics
Chlorine plasma reduction	Mare or pyroclastics
Silicate/Oxide Melt	
Molten silicate electrolysis	Mare or highlands
Fluxed molten silicate electrolysis	Mare or highlands
Caustic dissolution and electrolysis	Mare or highlands
Carbothermal reduction	Mare or highlands
Magma partial oxidation	Mare
Li or Na reduction of ilmenite	Ilmenite—Hi-Ti mare
Pyrolysis	
Vapor phase reduction	Mare or highlands
Ion (plasma) separation	Mare or highlands
Plasma reduction of ilmenite	Ilmenite—Hi-Ti mare
Aqueous Solutions	
HF acid dissolution	Mare or highlands
H_2SO_4 acid dissolution	Ilmenite—Hi-Ti mare
Co-Product Recovery	
Hydrogen/helium water production	Mare

A. Resupply Mass for Reagent Makeup

In any chemical/physical process involving recycling of reagents, there are always unavoidable losses. In terrestrial applications, many processes that

TABLE III

Qualitative Comparison of Lunar Oxygen Processes

Process	Tech-nology[a]	No. of Steps[b]	Process Conditions[c]	Feed-Stock[d]	Total	Rank
Solid/Gas Interaction						
Ilmenite red. with H_2	8	9	7	3	27	4
Ilmenite red. with C/CO	7	8	7	3	25	7
Ilmenite red. with CH_4	7	8	7	3	25	8
Glass red. with H_2	7	9	7	6	29	2
Reduction with H_2S	2	6	6	8	22	12
Extraction with F_2	5	1	2	10	18	16
Carbochlorination	3	3	3	10	19	15
Cl_2 plasma reduction	4	5	5	10	24	9
Silicate/Oxide Melt						
Molten sil. electrol.	6	8	5	10	29	3
Fluxed sil. electrol.	6	6	5	10	27	5
Caustic dis. electrol.	5	4	3	10	22	13
Carbothermal reduction	6	3	3	10	22	14
Magma partial oxid.	2	2	4	5	13	19
Li or Na red. of ilm.	2	3	5	2	12	20

Vaporization						
Vapor phase reduction	6	8	6	10	30	1
Ion (plasma) sep.	4	8	4	10	26	6
Plasma red. of ilmenite	7	8	6	3	24	10
Aqueous Solutions						
HF acid dissolution	5	1	2	10	18	17
H_2SO_4 acid dissolutn.	5	3	3	5	16	18
Co-Product Recovery						
H_2/He water production	7	9	8	1	25	11

[a] Technology: 1 = major technologic development required; 10 = no major unknowns.
[b] No. of steps: 1 = many (>5); 10 = one step.
[c] Process conditions (temperature, energy, plant mass, corrosion): 1 = severe; 10 = low.
[d] Feedstock requirements: 1 = huge quantities; 2 = mare, beneficiated (ilm); 5 = mare, unbeneficiated; 10 = any feedstock, unbeneficiated.

TABLE IV
Candidate Processes for Oxygen Production on the Moon

Processes	References
Solid/Gas Interaction	
Ilmenite reduction with hydrogen	Gibson and Knudsen (1988*a*)
Ilmenite reduction with C/CO	Chang (1959); Zhou and Shadman (1991)
Ilmenite reduction with methane	Friedlander (1985)
Glass reduction with hydrogen	McKay et al. (1991)
Silicate/Oxide Melt	
Molten silicate electrolysis	Haskin (1985); Colson and Haskin (1990)
Fluxed molten silicate electrolysis	Keller (1986); Keller and Taberaux (1991)
Pyrolysis	
Vapor phase reduction	Steurer and Nerad (1983); Chapter by Senior
Ion (plasma) separation	Steurer and Nerad (1983)

appear attractive on paper are found to be uneconomical because of the cost of the lost reagent. In lunar applications, minimizing makeup requirements will be even more important because the transportation cost dwarfs the material cost.

For example, in the ilmenite reduction with hydrogen process, if all of the hydrogen were lost and could not be recycled, then one-eighth tonne of hydrogen would have to be transported from Earth for each tonne of oxygen produced on the Moon: a product-to-reagent (payback) ratio of only 8 is obviously not economical. Exactly how great the payback ratio needs to be in order for lunar oxygen production to be economical is not known. But it probably must be at least 50, otherwise the capital investment would never be recovered. As shown in Table V, in order to have a payback ratio of 50, the recovery of hydrogen for ilmenite reduction must be 84%, which is probably achievable, particularly when one considers the hydrogen indigenous to the ilmenite feedstock itself. On the other hand, CO and CH_4 reduction require 98.8% and 98.2% recovery, respectively, in order to have a payback ratio of 50.

If a continuous process for the molten silicate electrolysis process can be demonstrated to be practical, then the only consumables may be the anodes and cathodes. In this case, the payback ratio would probably exceed 50. In addition, although recent experimentation on the use of NaF and LiF for the fluxed molten silicate electrolysis process has produced some promising results, no evaluation of reagent consumption has been made to date. The

TABLE V

Reagent Makeup for Potential Oxygen Production Processes

Processes	Product	Reagent	Required % Recovery of Reagent Product/Reagent = Payback				
			= 10	20	50	100	200
Ilmenite: High-Ti Mare							
Reduction with hydrogen	O_2	H_2	20	60	84	92	96
Reduction with CO	O_2	CO	94	97	98.8	99.4	99.7
Reduction with methane	O_2	CH_4	91	96	98.2	99.1	99.6
Mare or Highlands							
Glass reduction with H_2	O_2	H_2	20	60	84	92	96
Molten sil. electrolysis	O_2	?[a]	?				
Fluxed molten sil. electrol.	O_2	LiF, NaF[a]	?				
Vapor phase reduction	O_2	?[a]	?				
Ion (plasma) reduction	O_2	?[a]	?				

[a] The resupply masses for reagent makeup for these processes have not been determined to date. Therefore, the percent recovery of the reagents cannot be estimated.

vapor phase reduction process has yet to be completely defined. Indeed, if this process relies solely on space resources, the payback ratio theoretically would be infinity. In reality, it will probably be well above 50.

B. Quantity of Lunar Ore Required

Christiansen et al. (1988) concluded that a production rate of 1000 tonne/year of lunar oxygen would make a significant impact on space transportation. The quantity of ore required to produce this much oxygen with each of the 8 candidate processes is presented in Table VI. Note that the process "throughput" is very important, because this quantity strongly affects the percent recovery of reagents that is actually achievable, i.e., the greater the throughput, the lower the actual percent recovery. Stated another way, if two processes have similar required reagent recoveries in order to have a payback of 50, then the process with the smaller throughput is more likely actually to achieve the needed percent recovery.

Processes using beneficiated regolith (e.g., processes involving ilmenite) require values of about 200,000 tonne/year. At the high end, hydrogen-helium-water co-product recovery requires $>15\times10^6$ tonne/year; molten silicate electrolysis only requires 5000 tonne/year of unbeneficiated regolith, largely because oxygen from much of the entire feedstock is evolved and recovered. The large quantity of ore associated with the hydrogen-helium-water co-product recovery makes this process impractical for oxygen recovery alone.

C. Plant Mass Required

The estimated plant mass required to produce 1000 tonne/year of lunar oxygen is also shown in Table VI. For the processes involved with gaseous reduction of either ilmenite or glass, estimated plant sizes fall in the 200 tonne range, largely a function of the large throughputs. On the other hand, the other processes listed in Table VI have relatively small throughputs which will not require such massive plants.

D. Energy Requirement

The estimated energy required to mine, transport, and process lunar ore to produce 1000 tonne/year of oxygen is also shown in Table VI. Where known, the energy for the different processes fall in the 2 to 4 MWyr range.

E. Final Candidate Processes

Much of the data used to compile Tables V and VI is weak, some approaching guesswork, some recent and reliable, others older and subjective. Many of the estimates given are only guesses. The main source of data was Christiansen et al. (1988). These tables are intended only for casual evaluation and comparison and should not be adhered to. It is not felt that any further winnowing of these 8 processes is warranted at this time.

TABLE VI

Mine and Plant Characteristics: 1000 Tonnes of Oxygen per Year

Processes	Ore (T/yr)		Plant Mass	Energy
	Raw	Process Troughput	(T)	(MWyr)
Ilmenite: High-Ti Mare				
Reduction with hydrogen	210,000[a]	21,000[b]	200	3
Reduction with CO	210,000[a]	21,000[b]	225	3.5
Reduction with methane	210,000[a]	21,000[b]	225	3.5
Mare or Highlands				
Glass reduction with H_2	160,000[c]	80,000[d]	200	4
Molten sil. electrolysis	5,000[e]	5,000[f]	70	3
Fluxed molten sil. electrol.	5,000[e]	5,000[f]	80	3.5
Vapor phase reduction	5,000[e]	5,000[f]	40	2
Ion (plasma) separation	5,000[e]	5,000[f]	40	2.5

[a] Assumes feedstock with 50 wt% ilmenite from an ore with 5% available ilmenite for a beneficiation factor of 10.
[b] Assumes approximately 90% conversion of ilmenite.
[c] Assumes soil with 25% glass beneficiated to 50% glass for a beneficiation factor of 2.
[d] Assumes 15% FeO in glass (= 3.35% O_2 in glass) and 75% conversion of FeO to $Fe + \frac{1}{2}O_2$.
[e] No beneficiation necessary.
[f] Assumes about 43 wt% O_2 in soil with 50% recovery.

X. SUMMARY

Although over 20 processes have been put forth for the liberation of oxygen from lunar materials, many are largely untested and/or extremely complicated and difficult to implement. Simplicity and resupply mass are the keywords for the process(es) which will ultimately be selected for the initial production of oxygen on the Moon.

The processes that have received the most attention to date involve reduction of ilmenite by hydrogen, molten silicate electrolysis, fluxed molten silicate electrolysis, and vapor phase pyrolysis. However, there are several other processes which should be studied further, notably ilmenite reduction with carbon monoxide or methane, glass reduction with hydrogen, and ion (plasma) pyrolysis.

Acknowledgments. This chapter has been improved considerably by the constructive reviews of R. Colson, R. Keller, S. Rosenberg, C. Senior, and several anonymous reviewers. We gratefully acknowledge their efforts on our

behalf. This chapter was researched and prepared with the partial support of the NASA Space Resources Utilization Office at Johnson Space Center and the Lunar Geotechnical Institute in Lakeland, Florida.

REFERENCES

Allen, P. H., Prisbrey, K. A., and Detering, B. 1988. Plasma processing of lunar ilmenite to produce oxygen. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 411–419.

Altenberg, B. 1990. Processing Lunar In-Situ Resources. Tech. R & D Project, Final Rept., Job No. 90634-002, Bechtel Group, Inc.

Bhogeswara, R., Choudaray, U., Erstfeld, T., Williams, R., and Chang, Y. 1979. Extraction processes for the production of aluminum, titanium, iron, magnesium and oxygen from nonterrestrial sources. In *Space Resources and Space Settlements*, eds. J. Billingham and W. Gilbreath, NASA SP-428, pp. 257–274.

Binder, A. B. 1990. LLOX—Metal production via NaOH electrolysis. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 339–346.

Bockris, J. O'M., Kitchener, J. A., and Davies, A. E. 1952*a*. Electric transportation liquid silicates. *Trans. Faraday Soc.* 48:536–548.

Bockris, J. O'M., Kitchener, J. A., Ignatowicz, W., and Tomlinson, J. W. 1952*b*. Electric conductance in liquid silicates. *Trans. Faraday Soc.* 48:75–91.

Boundy, R. A. 1983. Executive summary. In *Research on the Use of Space Resources*, ed. W. F. Carroll, NASA JPL Publ. 83-36, pp. 1–3.

Burt, D. M. 1988. Lunar production of oxygen and metals using fluorine: Concepts involving fluorite, lithium, and acid-base theory. *Lunar Planet. Sci.* XIX:123–124 (abstract).

Burt, D. M. 1992. Lunar mining of oxygen using fluorine. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 423–428.

Bustin, R., and Gibson, E. K. 1992. Availability of hydrogen for lunar base activities. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 437–445.

Carter, J. L. 1985. Lunar regolith fines: A source of hydrogen. In *Lunar Bases Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 571–581.

Chang, M. C. S. 1959. Process for Treating Materials Containing Titanium and Iron. U. S. Patent No. 2,912,320, United States Patent Office, Washington, D. C.

Christiansen, E. L., Euker, H., Maples, K., Simonds, C. H., Zimprich, S., Dowman, M. W., and Stovall, M. 1988. Conceptual Design of a Lunar Oxygen Pilot Plant. EEI Rept. No. 88-182 (Houston: Eagle Engineering Inc.).

Colson, R. O. 1990. Characterization of metal products of silicate melt electrolysis. *Lunar Planet. Sci.* XXI:214–215 (abstract).

Colson, R. O., and Haskin L. A. 1990. Lunar oxygen and metal for use in near-earth space: Magma electrolysis. In *Engineering, Construction, and Operations*

in Space: Proc. Space 90, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 187–196.

Colson, R. O., and Haskin, L. A. 1991. Magma electrolysis—An update. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 12.

Colson, R. O., Haskin, L. A., and Keller, R. 1991. Electrolysis of lunar soil to produce oxygen and metals. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 7.

Cutler, A. H. 1984. An alkali hydroxide based scheme for lunar oxygen production. Lunar Bases and Space Activities of the 21st Century, Oct. 29–31, Washington, D. C., Abstract book, p. 14.

Cutler, A. H., and Krag, P. 1985. A carbothermal scheme for lunar oxygen production. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 559–569.

Dalton, C., and Hohmann, E., eds. 1972. Conceptual Design of a Lunar Colony. NASA/ASEE Systems Design Inst., NASA Grant NGT 44-005-114.

du Fresne, E., and Schroeder, J. E. 1983. Magma electrolysis. In *Research on the Use of Space Resources*, ed. W. F. Carroll, NASA JPL Publ. 83-36.

Friedlander, H. N. 1985. An analysis of alternate hydrogen sources for lunar manufacture. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 611–618.

Gibson, E. K., Chang, S., Lennon, K., Moore, G. W., and Pearce, G. W. 1975. Carbon, sulfur, hydrogen, and metallic iron abundances in Apollo 15 and 17 basalts. *Proc. Lunar Sci. Conf.* 6:1287–1301.

Gibson, M. A., and Knudsen, C. W. 1988*a*. Lunar oxygen production from ilmenite. In *Papers Presented to the Symp. on Lunar Bases and Space Activities in the 21st Century*, April 5–7, Houston, Tex., LPI Contrib. 652, p. 94.

Gibson, M. A., and Knudsen, C. W. 1988*b*. Lunar oxygen production from ilmenite. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 400–410.

Gibson, M. A., Knudsen, C. W., and Roeger, A. 1990. Development of the Carbotek process (TM) for lunar oxygen production. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 357–367.

Hammond, P. A., and Taylor, L. A. 1982. The ilmenite/titano-magnetite assemblage: Kinetics of re-equilibration. *Earth Planet. Sci. Lett.* 61:143-150.

Haskin, L. A. 1985. Toward a Spartan scenario for use of lunar materials. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 435–443.

Haskin, L. A., Colson R. O., Lindstrom, D. J., Lewis, R. H., and Semkow, K. W. 1992. Electrolytic smelting of lunar rock for oxygen, iron and silicon. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 411–422.

Heiken, G., Vaniman, D., and French, B., eds. 1991. *Lunar Sourcebook* (Cambridge: Cambridge Univ. Press).

Keller, R. 1986. Dry Extraction of Silicon and Aluminum from Lunar Ores. Final Report. SBIR contract NAS 9-17575, EMEC Consultants.

Keller, R. 1988. Lunar production of aluminum, silicon and oxygen. In *Metallurgical Processes for the Year 2000 and Beyond*, eds. H. Y. Sohn and E. S. Geskin (Warrendale, Pa.: Minerals, Metals & Materials Soc.), pp. 551–562.

Keller, R., and Taberaux, A. T. 1991. Electrolysis of lunar resources in molten salt. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC

Jan. 7–10, Tucson, Ariz., Abstract book, p. 10.

Kesterke D. G. 1971. Electrowinning of oxygen from silicate rocks. U. S. Bureau Mines Rept. Invest. 7587.

Lindstrom D. J., and Haskin L. A. 1979. Electro-chemical preparation of useful material from ordinary silicate rocks. In *Space Manufacturing Facilities*, eds. J. Gray and C. Krop (New York: AIAA), pp. 129–134.

Lynch, D. C. 1989. Chlorination processing of local planetary ores for oxygen and metallurgically important metals. In *Space Engineering Research Center for Utilization of Local Planetary Resources Annual Progress Report 1988–1989* (Tuscon: The University of Arizona).

Macmillan, D. W., ed. 1987. *Proc. of the International Symposium on Quality and Process Control in the Reduction and Casting of Aluminum and Other Light Metals* (New York: Pergamon Press).

McCullough, E., and Mariz, C. 1990. Lunar oxygen production via magma electrolysis. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 347–356.

McKay, D. S., Morris, R. V., and Jurewecz, A. J. 1991. Reduction of simulated lunar glass by carbon and hydrogen and its implications for lunar base oxygen production. *Lunar Planet. Sci.* XXII:881–882 (abstract).

Oppenheim, M. J. 1968. On the electrolysis of molten basalt. *Mineral. Mag.* 36:1104–1122.

Oppenheim, M. J. 1970. On the electrolysis of basalt, II: Experiments in an inert atmosphere. *Mineral. Mag.* 37:568–577.

Rosenberg, S. D. 1985. A lunar-based propulsion system. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 169–176.

Rosenberg, S. D., Guter, G. A., and Miller, F. E. 1966. The on-site manufacture of propellant oxygen from lunar resources. In *Aerospace Chemical Engineering*, eds. L. Isenberg and S. E. Stephanou (New York: American Inst. of Chemical Engineering), pp. 228–234.

Seboldt, W., Lingner, S., Hoernes, S., and Grimmeisen, W. 1991. Oxygen extraction from lunar soil by fluorination. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 11.

Semkow, K. W., and Sammells, A. F. 1987. The indirect electrochemical refining of lunar ores. *J. Electrochem. Soc.* 134:2088–2089.

Sherwood, B., and Woodcock, G. R. 1991. Costs and benefits of lunar oxygen: Engineering, operations and economics. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 12.

Simnad, M. T., Derge, G., and George, I. 1954. Ionic nature of liquid iron-silicate slags. *J. Metals* 6:1386–1390.

Simon, M. C. 1985. A parametric analysis of lunar oxygen production. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 531–541.

Steurer, W. H., and Nerad, B. A. 1983. Vapor phase reduction. In *Research on the Use of Space Resources*, ed. W. F. Carroll, NASA JPL Publ. 83-36.

Sullivan, T. A. 1990. Process engineering concerns in the lunar environment. In *Proc. AIAA Space Prog. and Technologies Conf.* (New York: AIAA), AIAA Paper 90-3753.

Taylor, L. A. 1988. Generation of native Fe in lunar soil. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 67–77.

Taylor, L. A. 1990*a*. Hydrogen, helium, and other solar-wind components in lunar soil: Abundances and predictions. In *Engineering, Construction, and Operations in Space: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 68–77.

Taylor, L. A. 1990*b*. Rocks and minerals in the regolith of the Moon: Resources for a lunar base. In *Advanced Materials: Applications of Mining and Metallurgical Processing Principles*, ed. V. I. Lakshmanan (New York: Soc. of Mining, Metallurgy and Exploration), pp. 29–47.

Taylor, L. A. 1991. Helium abundances on the Moon: Assumptions and estimates. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 40.

Taylor, L. A. 1992. Resources for a lunar base: Rocks, minerals, and soil of the Moon. In *Lunar Bases and Space Activities of the 21st Century II*, ed. W. W. Mendell, NASA CP-3166 (Houston: Lunar and Planetary Inst.), pp. 361–371.

Taylor, L. A., and Carrier, W. D., III. 1991. Oxygen production processes on the Moon: An overview. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 12.

Taylor, L. A., and Oder, R. R. 1990. Magnetic beneficiation and hi-Ti mare soils: rock, mineral, and glassy components. In *Engineering, Construction, and Operations in Space: Proc. Space 90*, ed. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 143–152.

Taylor, L. A., Williams, R. J., McCallister, R. H. 1972. Stability relations of ilmenite and ulvöspinel in the Fe-Ti-O system and application of these data to lunar mineral assemblages. *Earth Planet. Sci. Lett.* 16:282–288.

Taylor, L. A., Cooper, B., McKay, D. S., and Colson, R. O. 1990. Oxygen production on the Moon: Processes for different feedstocks. In *Advanced Materials: Applications of Mineral and Metallurgical Processing Principles*, ed. V. I. Lakshmanan (Littleton, Co.: Soc. for Mining, Metallurgy, and Exploration).

Taylor, S. R. 1975. *Lunar Science: A Post-Apollo View* (New York: Pergamon Press).

Taylor, S. R. 1982. *Planetary Science: A Lunar Perspective* (Houston: Lunar and Planetary Inst.).

Vaniman, D. T., and Heiken, G. 1990. Getting lunar ilmenite: from soils or rocks? In *Engineering, Construction, and Operations in Space: Proc. Space 90*, ed. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 107–116.

Waldron, R. D. 1985. Total separation and refinement of lunar soil by the HF acid leach process. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSS Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 132–149.

Waldron, R. D. 1989. Magma partial oxidation: A new method for oxygen recovery from lunar soil. In *Space Manufacturing 7, Proc. of the Ninth Princeton/AIAA/SSI Conf.: Space Resources to Improve Life on Earth*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 69–77.

Waldron, R. D., and Criswell, D. R. 1979. Over-view of methods for extraterrestrial materials processing. In *Space Manufacturing 3, Proc. of the Fourth Princeton/AIAA Conf.: Space Manufacturing Facilities*, eds. J. Grey and C. Krop (New York: AIAA).

Waldron, R. D., and Criswell, D. R. 1982. Materials processing in space. In *Space Industrialization*, vol. I, ed. B. O'Leary (New York: AIAA), pp. 97–130.

Williams, R. J. 1985. Oxygen extraction from lunar materials: An experimental test of an ilmenite reduction process. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 551–558.

Williams, R. J., and Mullins, O. 1983. Enhanced production of water from ilmenite:

XIV:34–35 (abstract).

Williams, R. J., McKay, D. S., Giles, D., and Bunch, T. E. 1979. Mining and beneficiation of lunar ores. In *Space Resources and Space Settlements*, NASA SP-428, pp. 275–288.

Zhao, Y., and Shadman, F. 1990. Kinetics and mechanism of ilmenite reduction with carbon monoxide. *AIChE J.* 36:443.

Zhao, Y., and Shadman, F. 1991. Reaction engineering for materials processing in space: Reduction of ilmenite by hydrogen and carbon monoxide. Resources of Near-Earth Space: Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 12.

PRODUCING OXYGEN BY SILICATE MELT ELECTROLYSIS

RUSSELL O. COLSON and LARRY A. HASKIN
Washington University

Because of the Earth's substantial gravity well, the Moon, with its lower gravity, is a potentially valuable source of materials for use in space. Oxygen, fuel for both humans and rockets, is the most abundant element on the Moon and a potentially valuable lunar resource if it can be easily extracted from the rocks in which it is chemically bound. The unique conditions on the Moon, such as vacuum, absence of many reagents common on the Earth, and presence of very nontraditional "ores" suggest that a unique and nontraditional process for extracting materials from these ores may prove the most practical. This process should be simple, problem-free and energy efficient. Electrolysis of molten silicate has the advantages of simplicity of concept, absence of need to supply reagents from Earth, and low-power and mass requirements for the processing plant, thus meeting the criteria above. Electrolysis experiments using 1 to 2 g quantities of silicate melt have been done at low cell voltages and indicate that the process can be very energy efficient. Materials for container and electrodes have been identified that are stable under electrolysis conditions for at least 2 hr. Larger experiments of longer duration are needed to further test the durability of container and electrode materials. Silicate melt electrolysis is versatile with respect to feedstock composition, meaning that almost any lunar material can be used as feedstock. In addition, the process is not sensitive to expected variations in feedstock composition, meaning that compositional variations over a limited area, such as might be mined for feedstock, do not require compensating adjustments to the process. Developing a process to extract oxygen from lunar materials is an important step in learning to live in space. As such, its value to humanity may be much greater than indicated by a calculation of the cost versus return.

I. INTRODUCTION

One consequence of Earth's relatively high gravity is a relatively high energy cost for launching material from Earth into space, even into near-Earth space. Because of this, materials available from extraterrestrial sources should become economically attractive as the need for propellants and constructional materials in near-Earth space grows. We believe that the Moon, because of its proximity and our knowledge of its surface, is the most sensible source of such material for the near future. Its gravity is only about one-sixth that of Earth, low enough to improve greatly the payload/liftoff mass ratio for conventional rockets but still high enough to facilitate separation and manufacturing processes. On this basis, it has been argued that lunar production of

material for use in space can be economically viable (see, e.g., Criswell 1983; Koelle 1988). However, others have concluded that costs of developing a lunar industry could overwhelm returns over any reasonable time frame (see, e.g., Simon 1985; Sherwood and Woodcock 1991).

If we take it as given that the destiny of humankind should include space as more than an outpost of scientific or military activity, that it is ultimately a place for people to live, then the long-term economic benefit of learning to live there is not something readily evaluated by weighing the costs of launching material from Earth versus producing material in space. It can be argued that the most valuable asset of space is space. It is potentially a place to live. The most valuable investment in space may be simply learning how to live there. We propose that learning how to use the resources found in space is an essential step in this task.

One material that we expect to need in increasing quantities in space is oxygen. It will be used for people to breathe, but much larger amounts will be used as the oxidizer in rocket propellant. Oxygen can be supplied from the Moon, where it is the most abundant element (Haskin and Warren 1991). Although free oxygen does not exist on the Moon, it can be extracted from lunar rock and regolith where it exists in chemical combination with metallic elements, primarily in silicate phases. One way of extracting the oxygen from the lunar regolith is by electrolysis.

Oxygen is present in abundance in nearly all lunar material (see, e.g., Haskin and Colson 1990) and can theoretically be extracted by electrolysis from any known lunar rock. Derivation of oxygen by direct electrolysis of silicate melt has been demonstrated experimentally in silicate melts of a variety of compositions (see, e.g., Bockris et al. 1951,1952*a*; Simnad et al. 1954; Oppenheim 1968,1970; Kesterke 1971; Lindstrom and Haskin 1979; Colson and Haskin 1990; Haskin et al. 1992). Electrolysis of silicate melt with additions of "fluxing" agents has also been amply demonstrated (see, e.g., Keller 1988; Grjotheim et al. 1990; Binder 1990). This chapter deals only with direct silicate electrolysis and for discussion of electrolysis using various "fluxing" agents the reader is referred to the references given above.

In its conceptually simple form, electrolysis entails immersing two electrodes in a vat of molten silicate and imposing a direct current between the electrodes. Oxygen is derived at the anode and metal (mostly Fe and Si, see, e.g., Haskin et al. 1992) is deposited at the cathode. In order to evaluate silicate electrolysis as a means for extracting oxygen on the Moon, and to compare it to alternative processes, it is necessary not only to know that oxygen can be derived by electrolysis but also how much energy the process requires, how large a facility is needed to produce a given amount of oxygen, and whether container and electrode materials can be designed that survive the corrosive, high-temperature conditions of a silicate melt. This knowledge requires a fundamental understanding of the physical chemistry of silicate melts and of the possible container materials. Much of the research to attain a first-order understanding of this has now been done and is reviewed below.

II. ENERGY CONSIDERATIONS

On Earth, electrolysis has found many important applications (of which reduction of ores is only a single class) in the chemical and metallurgical industries. For example, it is used to extract aluminum from bauxite and to split water into hydrogen and oxygen. On the Moon, electrolysis will almost certainly have a major role in any oxygen production process, either directly through electrolytic smelting of lunar ores or indirectly through electrolysis of water produced by other ore-reduction processes. The viability of electrolysis as an oxygen-producing process must be measured in terms of the costs of raw materials, energy, labor and capital. The relative influence of each of these on oxygen production on the Moon will be very different from our experience on Earth and, in the short term, the total cost is likely to be most strongly influenced by the very high costs of labor, capital and imports from Earth. Initial low rates of oxygen production will also tend to de-emphasize the importance of energy efficiency in the short term. However, in so far as we view early lunar oxygen production as a part of "gaining a foothold in space" and learning to live there (Mueller 1983), it is wise to consider the longer-term importance of energy efficiency in choosing the production process.

Traditionally, the efficiency of electrolysis has been measured in terms of percent yield, percent conversion per pass, selectivity, current efficiency and energy consumption. We discuss the meanings of these terms below.

Percent yield refers to the amount of desired product derived from a given amount of starting material. It is equal to (atoms product)/(atoms-starting material consumed) times 100%. About 60 atom% of typical lunar materials are oxygen, representing an upper limit to the percent yield that can be attained for oxygen production. The percent yield is a measure of how much feedstock is required to obtain a given amount of product. It is also useful to consider *yield efficiency*, or the percent of the desired element present in the feedstock that is obtained as product. This value will be less than 100% if not all the oxygen is extracted from the feedstock.

It is possible that the amount of product derived by any one pass of the material through the electrochemical cell is low even though the percent yield is improved by many passes (although repeated passes of material is not common in industrial electrolysis on Earth). Because each cycle of product separation and processing adds to the cost of the process, it is useful to define *percent conversion per pass* as the amount of desired product derived from a given amount of starting material for a single pass of the material through the electrolysis cell.

Selectivity is useful when a mixture of products is obtained. It refers to the percent of the products comprised of the desired product. If gases other than oxygen are produced by the process, this value will decrease from 100%.

The term *current efficiency* refers to the fraction of the current that results in formation of the desired product. If parasitic reactions occur at the anode, if the desired product is partially consumed by back-reactions with other

products or the melt itself, or if some of the current flows electronically rather than ionically, then not all of the electrical current will contribute to the desired production of oxygen. This nonproductive use of electrical energy can substantially increase the amount of energy required to produce a given amount of product.

Energy consumption can be expressed by the following equations (Rieger 1987):

$$\text{Energy (Joules}/n \text{ moles e}^-) = nFE \tag{1a}$$

$$E = E_c - E_a - |n_c| - |n_a| - i(R_{\text{cell}}). \tag{1b}$$

Here n is the number of electrons involved in the charge transfer, F is Faraday's constant (9.65×10^4 C/mole), E is the cell potential (which is proportional to energy), $E_c - E_a$ is the potential required to drive the oxidation-reduction reaction, $-|\eta_c| - |\eta_a|$ is the overpotential required to overcome any kinetic problems, i is the current, and R_{cell} is the resistance of the electrolysis cell and equals $L/(\kappa A)$, where L is the distance between electrodes, A is the surface area of the electrodes, and κ is the melt conductivity.

Both i and $-|n_c| - |n_a|$ increase as the rate of production is increased (for any given cell configuration). Therefore, energy requirements increase as production rate increases. Also, the energy required to produce at a given rate increases as the electrolysis cell is made smaller (smaller values of A). Although some of this excess energy can be used to melt incoming feedstock, much will have to be removed as waste. In general, a higher production rate and smaller cell size result in lower costs of labor and capital per unit product but higher energy costs. Therefore, there is a trade-off between energy costs on the one hand and labor and capital costs on the other. The optimum oxygen production cell must balance production rate, cell size, and energy consumption.

We point out that estimation of *energy consumption per unit product* requires that the current efficiency also be considered. For example, the current, designated by the term i above, can be related to current efficiency by the expression

$$i = (\text{current needed to yield desired production rate at}$$

$$100\% \text{ efficiency})/(\text{current efficiency}) \tag{2a}$$

and energy is expressed by

$$\text{Energy (J/mole product)} = nFE/(\text{current efficiency}). \tag{2b}$$

Achievable production rates are also influenced by the rate at which material is transported to the electrode surface where oxidation or reduction takes place. If mass transport is slow (diffusional), then net electrolysis at a given cell voltage can be negligible even if it is energetically favorable. In

such a case, additional energy (i.e., larger cell voltage) will be required to drive the reaction at a reasonable rate. The most obvious way to increase the rate of mass transport is to stir the solution. This could be accomplished by either active stirring by some stirring apparatus in the solution or passive stirring by convective mixing or stirring induced by the formation of an electrode product such as bubbles of gas. In either case, the viscosity of the solution is an important parameter in determining the mass transport rate. Viscosity, along with density, is also important in governing how readily products can be separated from slag.

From the discussion above, we can infer that a prediction of the energy needed to produce oxygen from lunar materials requires a good understanding of the physical chemistry of the silicate melts from which the oxygen is to be electrolytically extracted. This includes knowledge of the reactions involved at both the anode and the cathode, the energy and kinetics of these reactions, the other reactions that compete with the product-forming reactions, and the viscosities and conductivities of the melts. Much of this knowledge is available in the scientific literature. For example, viscosity of silicate melts as a function of composition was studied by Bockris et al. (1955) and Bottinga and Weill (1972), energy of reactions in silicate melts are reported in several studies (see, e.g., Semkow et al. 1982; Semkow and Haskin 1985; Schreiber 1987; and Colson et al. 1990), and Haskin et al. (1992) report melt conductivities and studies of electrode reactions specific to melt electrolysis. We briefly review here these results (Haskin et al. 1992).

In the electrolysis of basaltic silicate melt, there are two primary cathode reactions that produce metal. They are the following:

$$Fe^{2+} + 2e^- \leftrightarrow Fe^o \tag{3}$$

and, in simplified form (because the actual reaction requires depolymerization of silicate),

$$Si(IV) + 4e^- \leftrightarrow Si^o. \tag{4}$$

Haskin et al. (1992) infer that the reduction of Fe^{2+} is a simple, two-electron transfer with reaction rates sufficiently fast not to interfere with efficient electrolysis. They do not discuss the kinetics of the reduction of Si(IV). Reductions of additional elements existing in silicate melts at lower concentrations (such as Cr^{3+}, Cr^{2+}, Ti^{4+}, Ti^{3+} and Mn^{2+}) are also expected to occur in the same potential range as the reductions of Fe^{2+} and Si(IV). At sufficiently negative cathode potentials, less readily reduced cations in the melt, including Al^{3+}, Mg^{2+} and Ca^{2+}, will begin to react. Because they are less readily reduced, these elements are not expected to be key players in silicate melt electrolysis to extract oxygen. There is an additional competing reaction at the cathode if the oxygen pressure above the cell is sufficiently high:

$$O_2 + 4e^- \leftrightarrow 2O^{2-}. \tag{5}$$

However, this reaction is not significant for oxygen fugacities less than about 10^{-3} because of the low concentration of O_2 in the melt at low oxygen fugacities (less than 2.3×10^{-5}moles/ℓ at $f_{O_2} < 10^{-3}$ atm; Semkow and Haskin 1985).

At the anode, the principal desired oxygen-producing reaction is the reverse of Eq. (5). However, this reaction is not significant because the concentration of O^{2-} in silicate melts is low and formation of new O^{2-} from silicate polymer chains is a kinetically slow process relative to desired electrolysis rates. Therefore, the actual reaction during electrolysis involves oxidation of silicate polymer chains to form oxygen in the neutral state (Haskin et al. 1992) as shown in Eq. (6). This process increases the extent of polymerization of the silicate,

$$4(\mathrm{SiO}-) \leftrightarrow 2(\mathrm{Si}-\mathrm{O}-\mathrm{Si}) + \mathrm{O}_2 + 4\mathrm{e}^-. \tag{6}$$

This reaction is sufficiently fast to provide for electrolysis at a reasonable rate.

Serious competing reactions at the anode include the oxidation of reduced species that have higher oxidation states, such as $Fe^{2+} \rightarrow Fe^{3+}$, $Cr^{2+} \rightarrow Cr^{3+}$, or $Ti^{3+} \rightarrow Ti^{4+}$. The most important competing reaction, involving iron, is the following:

$$\mathrm{Fe}^{2+} \leftrightarrow \mathrm{Fe}^{3+} + \mathrm{e}^-. \tag{7}$$

Other competing reactions at the anode can include oxidation of metal products (primarily Si^0 and Fe^0) that are transported to the anode from the cathode either by suspension of small metal beads or by dissolution of the metal in the silicate melt. We have measured the solubilities of transition metal species in silicate melts and found them to be on the order of tenths of a percent (Colson 1990, and unpublished). This is too low to decrease current efficiency substantially. The transport of metal in suspension in the silicate melt is a complex function of cell design and dynamics and is not readily predicted from our small-scale experiments.

Haskin et al. (1992) conclude that reaction (7) is the primary competing reaction at the anode (because Cr^{2+} and Ti^{4+} are generally present at lower concentrations than Fe^{2+}) and that little or no electronic conduction in the melt occurs. Therefore, they are able to relate the current efficiency to the concentration of Fe^{2+} in the melt. By their data, the current efficiency equals $(0.049/X_{\mathrm{FeO}})/(1 + [0.049/X_{\mathrm{FeO}}])$. Here, X_{FeO} is the mole fraction FeO (or Fe^{2+}) in the melt. This has the consequence that electrolysis to produce oxygen as a main product is most efficiently carried out in melts with relatively low iron concentrations (<2%).

Electrical conductivities of silicate melts have been measured in simple metal oxide-silica systems (Bockris et al. 1952*a*, *b*); temperature dependence has been determined (du Fresne and Schroeder 1983); and the compositional dependence has been measured and modeled (Haskin et al. 1992). Du Fresne and Schroeder measured temperature dependence of conductivities for silicate melts of two different compositions: a basaltic melt and an anorthositic melt. Temperature dependence of the conductivities (in $\mathrm{ohm}^{-1}\mathrm{cm}^{-1}$) determined in

their study can be expressed as $\ln\kappa = 8.9 - 140500/RT$ (for basaltic melt) and $\ln\kappa = 8.8 - 160500/RT$ (for anorthositic melt) (R = gas constant in J/Kelvin-mole). Conductivities decrease about an order of magnitude with a 300° decrease in temperature.

Haskin et al. (1992) model empirically the dependence of melt conductivities on composition at a single temperature (1425°C) using an expression of the form

$$\ln\kappa = A + \Sigma B_i X_i \tag{8}$$

where A and B_i are constants determined by linear regression and are related to the diffusivity and charge of ions in the melt, and values of X_i are the mole fractions of the various ionic species. The expression they derive is

$$\ln\kappa = 5.736 - 12.6 \cdot SiO_2 - 10 \cdot AlO_{1.5} - 3.7 \cdot TiO_2 +$$

$$1.89 \cdot FeO + 0.07 \cdot MgO - 1.25 \cdot CaO. \tag{9}$$

It is seen that SiO_2 and $AlO_{1.5}$ have the greatest effect on conductivity, and higher concentrations of these oxides sharply decrease melt conductivity.

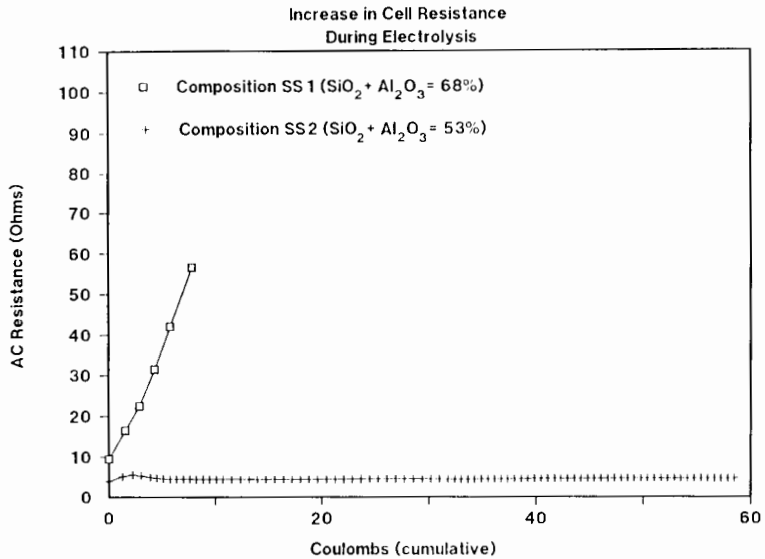


Figure 1. Dependence of frothing (indicated by increase in resistance with current) on melt composition. Both experiments involved electrolysis using Pt electrodes about 0.5 cm apart with current densities of about 0.1 A cm^{-2}. Higher viscosities in the composition with higher concentrations of SiO_2 + Al_2O_3 result in a large increase in cell resistance due to bubbles forming at the anode faster than they can escape from the melt.

SiO_2 and $AlO_{1.5}$ also increase viscosity (Bottinga and Weill 1972). As discussed above, higher viscosity results in lower transport rates to the electrodes and therefore lower rates of production. Also, for the case of oxygen production, higher viscosities will inhibit the escape from the melt of the oxygen produced at the anode. This can sharply increase the cell resistance, thereby increasing the amount of energy required for electrolysis. We have observed that frothing of the melt occurs in electrolysis experiments in which the combined concentrations of SiO_2 and $AlO_{1.5}$ are higher than about 55 wt%. With lower concentrations of these polymerizing cations, viscosities are lower and the frothing does not occur. This is shown in Fig. 1. Therefore, because of both conductivity and viscosity effects, efficiency of oxygen production will be optimized by maintaining low concentrations of SiO_2 and $AlO_{1.5}$ in the electrolysis cell.

To summarize the efficiency of silicate melt electrolysis in the terms defined above, and using results of Haskin et al. (1992): percent yield $\sim$ percent conversion per pass (only one pass assumed) $\sim$30% (corresponds to about 20 wt%); percent selectivity is nearly 100% for oxygen production at the anode; and current efficiency is at least as high as 0.85 in low-FeO melts (presuming that a low-Fe steady-state melt can be maintained even if the feedstock has higher FeO concentrations). $E_c - E_a \sim -1.4$ V; $-|\eta_c| - |\eta_a| \sim$ 10% of $E_c - E_a$ (estimated by Haskin et al. 1992); and $i(R_{\text{cell}})$, the energy lost to resistive heating, is calculated on a theoretical basis as about 0.7 V (for a steady state composition derived by subtracting about half of the SiO_2 and all of the FeO from a basaltic melt). Compared to the energy required by alternative processes proposed for extracting oxygen from lunar materials, these values appear very good (Colson and Haskin 1990).

We have measured the energy lost to resistive heating caused by both melt resistivity and resistance due to bubbles forming at the anode during electrolysis in experiments using 1 to 2 g of silicate melt with less than 55 wt% SiO_2+ $AlO_{1.5}$. Electrodes were 0.5 cm apart and current density was 0.1 A cm^{-2}. The value of $i(R_{\text{cell}})$ was $<$0.14 V, which is less than 10% of the potential needed to remove oxygen from SiO_2 in silicate melt. If we presume that the current efficiency is 0.85 or larger (Haskin et al. 1992), then the energy requirement of electrolysis in this experiment is calculated to be $<$1.3 times the theoretical minimum. This energy requirement is substantially lower than has been suggested for most other processes, which require 2 to 4 times the theoretical energy (Colson and Haskin 1990) and it is lower than the energy requirement calculated on theoretical grounds for batch electrolysis according to Haskin et al. (1992). However, Haskin et al. (1992) presume a current density of about 1 A cm^{-2}, whereas these experiments had a current density of about 0.1 A cm^{-2}. In other experiments using this same composition but much higher current densities (0.8 to 1 A cm^{-2}), resistance varied erratically, as though frothing periodically increased and dissipated, and energy requirements were higher (at 1 A cm^{-2}, $i(R_{\text{cell}})$ rose to as high as 5 V).

III. TRANSPORT AND MAINTENANCE CONSIDERATIONS

As discussed above, short-term costs of oxygen production will likely be dominated by capital and labor costs. Therefore, it is important to minimize the initial plant mass that must be brought from Earth (roughly proportional to the capital investment) and the amount of supplies that need to be brought on a continuing basis. It is also very important that the process used to produce oxygen on the Moon be relatively problem-free to minimize the amount of astronaut time required for maintenance and to minimize down-time. It is important that feedstock for the process be easy to get. These considerations require a more practical analysis of the process than has been done above.

Because of its conceptual simplicity, silicate melt electrolysis is potentially one of the most problem-free of the processes proposed. It is a single process taking place in a single reaction pot with a minimum of moving parts and no imported reagents. It is because of this conceptual simplicity that silicate melt electrolysis is a promising candidate for producing oxygen on the Moon, where it is important to keep the process simple and problem-free (see, e.g., Haskin 1985). However, high-temperature silicate melts are very corrosive, and few materials, if any, are completely inert in them. It has not yet been demonstrated by any experiment on a scale that is directly comparable to what is needed for lunar electrolysis that any of the materials proposed for container or electrodes can survive without substantial maintenance. Below, we discuss the materials that have been proposed in the literature for use as anodes or container, report the behavior of selected electrodes and container in our own electrolysis experiments, and finally, discuss the problem of scaling up.

Several materials have been proposed for use as anodes in silicate melt electrolysis. As pointed out by McCullough and Mariz (1990) and Taylor et al. (1991), the anodes must meet several criteria. These include being either inert in or thermodynamically stable with respect to the silicate melt, having reasonable strength and creep characteristics at the electrolysis conditions, and being sufficiently conductive to minimize power loss due to resistive heating in the anode. McCullough and Mariz (1990) introduce three classes of electrodes: homogeneous (composition is constant throughout electrode); nonhomogeneous (diffusional gradient or layer-structure yields a core and exterior of the electrode of different compositions), and composite (combination of ceramics and metals is used to decrease problems with creep). All of their proposed anodes involve the use of some Fe or Cr spinel with various impurities to provide improved conductivity and various layering or compositional gradients to provide improved strength or stability. Although they discuss the need to control oxygen fugacity to stabilize the spinels, they do not discuss the problem that Fe or Cr activity in the melt must also be controlled. These activities would need to be maintained at levels that are high relative to what is generally found in lunar materials and at levels that are sufficiently high to decrease current efficiency substantially (as discussed above) and reduce

percent yield (because in order to maintain a sufficiently high concentration of iron so that the Fe-Cr spinel will be stable, only a small amount of the iron can be removed by electrolysis). However, their proposed anodes have the possible advantage that they can be constructed from lunar materials.

Use of platinum in anodes was proposed by both Haskin et al. (1992) and McCullough and Mariz (1990). Platinum seems to be a reasonable candidate for an anode material because of its apparently inert behavior in silicate melts. It has been used for decades in experimental work involving high-temperature silicate melts (see, e.g., Donaldson et al. 1975). However, when Pt is used as an anode, it is not perfectly inert (Minenko et al. 1961; Semkow and Haskin 1986). Semkow and Haskin (1986) and Haskin et al. (1992) conclude that a Pt-oxide film quickly forms on the platinum anode during electrolysis creating a diffusional barrier that inhibits additional Pt-oxide formation, or, alternatively, a dynamic equilibrium is established in which the rate of decomposition of Pt-oxide matches the formation of new Pt-oxide.

Figure 2. Pt anode surface after 16 min of electrolysis at about 1.4 A cm^{-2} and approximately 10 V between electrodes. The bright Pt blobs seen in the silicate glass up to 25 μm from the anode are metal, not oxide, suggesting that if the anode is being oxidized during electrolysis, it is quickly reduced by the silicate melt. Backscattered electron image, bar is 10 μm in length (experiment number A12-1ch).

Our use of platinum anodes in electrolysis experiments of durations of up to 45 min confirms that the platinum anodes do not quickly oxidize and fail. However, close examination of some of the anodes after the experiments reveals that particles of the anode have separated and migrated as much as 25 μm (Fig. 2). We believe that this indicates that the Pt-oxides that form at the anode can begin to migrate into the silicate melt and that they are quickly reduced by the melt itself, leading to establishment of a dynamic equilibrium in which formation of new Pt-oxide is balanced by Pt-oxide reduced by the

silicate melt. Whether this dynamic equilibrium is maintained sufficiently close to the anode to preserve the integrity of the Pt anode over extended periods of electrolysis is not known.

A second question we have addressed experimentally is how readily platinum metal dissolves in silicate melts. Our recent measurements of the solubility of Pt metal in silicate melts indicate that it is soluble at roughly the tens of ppm level. Concentrations of Pt 50 to 100 μm from the anode have been found to be as high as 100 to 200 ppm after electrolysis experiments (presumably as both Pt metal and Pt-oxides dissolved in the melt). Although these concentrations are sufficiently low to be considered inert in a short-term experiment, Pt may not be inert relative to extended electrolysis. For example, if we presume that the kinetics of Pt dissolution in silicate melt are sufficiently fast to maintain equilibrium, production of 300 metric tonnes of oxygen per year corresponds to loss of more than 300 kg Pt per year (presuming solubility of Pt is 200 ppm). This is about 0.1% of current world production.

Colson and Haskin (1990) identify four general classes of possible containers (also applicable to electrodes). These are: selection of an inert material that is not significantly soluble in the silicate melt; selection of a material that is in thermodynamic equilibrium with the melt; selection of a material that can be maintained in steady-state equilibrium such as an "iron skull" in which the interior is molten but the exterior is cooled to form a solid skin (see, e.g., Greiner et al. 1933, p. 17); and selection of a material that is destroyed by the silicate melt, but only slowly.

In addition to holding the silicate melt, the container must also hold the iron-silicon metal cathodic product of electrolysis. This prevents the use of platinum as a container material because Pt forms an alloy with silicon with a melting temperature of 830°C.

Ceramic container materials have been proposed by both Colson and Haskin (1990) and McCullough and Mariz (1990). McCullough and Mariz suggest the use of refractory fire brick of unspecified composition. Although it is unlikely this brick would be in equilibrium with silicate melt of lunar composition, aluminum-rich refractory material would probably dissolve only slowly in the silicate melt and could be readily replaced by materials available on the Moon. Keller and Larimer (1991) used Al_2O_3 crucibles successfully in electrolysis experiments lasting 3 hr. Colson and Haskin (1990) suggest the use of Mg-Al spinel as a container material because electrolysis tends to change typical lunar compositions in such a way as to make this spinel stable. This refractory material was used in our electrolysis experiments and was in contact with the silicate melt for up to 2 hr without any indication of corrosion by the melt (Fig. 3).

IV. SCALING UP

There are several aspects to scaling up, including larger experiments more closely emulating a working oxygen factory, experiments of longer duration,

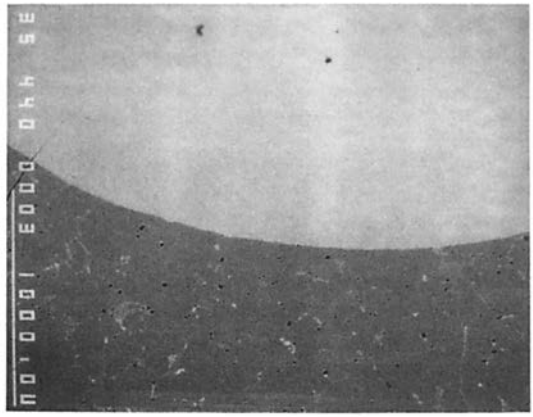


Figure 3. Spinel crucible (dark) in contact with silicate melt (light) for 2 hr shows little indication of corrosion by the silicate melt. The melt was of a composition expected to be in equilibrium with the spinel based on phase equilibria (approximately SiO_2 = 61.8 wt%, Al_2O_3 = 20%, MgO = 26.5%, CaO = 26.5 (Levin et al. 1969) plus 5.6% of a synthetic composition similar to Apollo 12 soil 12001 (Taylor 1982)]. Glass stringers permeating the spinel follow fractures formed when the crucible and melt were repeatedly cooled and reheated over the course of four separate electrolysis experiments. Backscattered electron image, bar is 1000 μm in length (experiment number ELECSS6).

and experiments more closely emulating lunar conditions (such as lower gravity). Some of the observations that are particularly sensitive to size include dynamics of melt mixing and transport rate to the electrodes, heat budget within the cell, and the dynamics of any magnetic-field-induced instabilities in the melt. Other observations are particularly sensitive to experiment durations. These include experiments dealing with the durability of materials, such as those on the durability of platinum and spinel proposed for use as anode and container material. Measurements that depend on gravity include determination of the ability of the oxygen bubbles produced at the anode to escape (although this can be estimated from the melt viscosities, the effect of the cell configuration on how easily bubbles escape is not easily determined theoretically), and the tendency of melt droplets formed as oxygen bubbles escape from the melt to stick to the cell roof.

Some of the experimental measurements done so far are not particularly sensitive to problems encountered in scaling up. These include studies of the physical chemistry of silicate melts. Reaction energies, reaction kinetics, effects of competing reactions, viscosities, diffusivities, conductivities, and the like are sufficiently insensitive to gravity or surface tension that these observations can be considered applicable to scaled-up electrolysis. (We note that McCullough and Mariz [1990] listed behavior of parasitic reactions and formation of diffusion-inhibiting polymers in the melt as measurements that require scaled-up experiments in lunar gravity. We disagree with their

assessment.)

Caution must be used, however, in applying these observations made in small experiments to calculations relevant to large-scale electrolysis. Although the physical chemistry of the melt provides all the information we need in theory to calculate the dynamics and efficiency of electrolysis, in practice these calculations are often too complex to make a reasonable prediction possible. Although the current efficiency depends almost entirely on the concentration of Fe^{2+} at the anode (significant effects of temperature and overvoltage were not observed in our experiments), the concentration of Fe^{2+} at the anode can be a complex function of cell dynamics not easily calculated from the known quantities of viscosity, diffusion, etc. For example, it is possible that stratification of the cell could eliminate this inefficiency problem by providing low Fe^{2+} melt at the anode and higher Fe^{2+} melt elsewhere. The consequence of such approaches are difficult or impossible to assess based on the results of our small experiments (because of the complex relationships among stirring, diffusion, and design of the actual cell). Also, Cutler (1985) suggests that very high current densities could improve current efficiency, albeit at the expense of the power needed by the process. Such high current densities cannot be achieved in our experiments because the sample froths out of its container.

Therefore, larger-scale experiments concentrating on engineering and design of the cell are required before actual efficiencies can be more than roughly estimated. To this extent, the small-scale experiments we report here serve to constrain the type of larger-scale experiments that are worth attempting.

V. CONSIDERATIONS OF FEEDSTOCK COMPOSITION

From the discussion above it can be inferred that efficiency of silicate melt electrolysis is improved by decreasing the concentrations of SiO_2, $AlO_{1.5}$, and FeO in the melt in the electrolysis cell. By constraining SiO_2+Al_2O_3 to be less than 55 wt% and FeO less than 2 wt%, the energy required by silicate melt electrolysis of 1 to 2 g samples can be kept below 130% of the theoretical energy needed to extract oxygen from SiO_2 in the melt (as discussed above). If a container material is chosen that must be maintained in thermodynamic equilibrium with the melt (such as spinel discussed above), phase equilibria place additional constraints on the composition of the material in the cell. It is interesting to determine whether, in the worst-case scenario in which all of these constraints must be met before the electrolysis process is competitive with other processes, these constraints can be met using known lunar materials.

The composition of the melt in the cell is not simply the composition of the input feedstock. Rather, it is a steady-state composition resulting from the steady removal of the products (mainly Si, Fe and O_2) from the feedstock as it is added to the cell. The variables we can use to control the steady-state composition include the feedstock composition (which can be varied

within certain limits), the composition of the metal product (which can be varied by adjusting the amount of feedstock electrolyzed via cell potential and feedstock flow-through rate), and the temperature (which places constraints on the composition of a melt in equilibrium with a container or electrode material).

In the following example, we constrain the steady-state concentration of FeO to be less than 2 wt% (to maximize current efficiency; Haskin et al. 1992), SiO_2+Al_2O_3 to be less than 55% (to minimize frothing and maximize melt conductivity; Haskin et al. 1992), and the steady state melt to be in equilibrium with spinel ($MgAl_2O_4$). Depending on the potential imposed between the electrodes, Si/Fe in the metal product can vary between 0 and Si(feedstock)/Fe(feedstock). In this exercise, we assume the feedstock achieves equilibrium at the imposed potential. Therefore, the amount of Fe, Ti and Si metal produced are defined by $K_{Fe} = f(E)$ = FeO(melt)/Fe(metal), $K_{Si} = f(E)$ = SiO_2(melt)/Si(metal), and $K_{Ti} \sim K_{Si}$. Here, E is cell potential, and, for this exercise, the activities are approximated by wt% SiO_2 in melt and mole fraction Si or Fe in metal. Although these definitions for activities are only crude approximations, errors in these approximations affect corresponding cell potential only minimally and do not affect at all the results we discuss below. The amount of spinel removed from the melt is determined from phase equilibria (see, e.g., Levin et al. 1969, p. 185). After a sufficient length of time, the steady state composition is defined by the following relationship:

$$\text{wt fraction } A \text{ in cell} = (a - bc)/(1 - c) \tag{10}$$

where a = weight fraction of A in feedstock, b = weight fraction of A in product, and c = weight fraction of feedstock that yields product.

As an example of how the steady-state melt composition can be calculated from these constraints, we give the following. We choose an example feedstock of composition SiO_2 = 46.77 wt%, TiO_2 = 1.46%, Al_2O_3 = 16.75%, FeO = 12.4%, MgO = 10.35%, CaO = 10.95%, Na_2O = 0.45%, K_2O = 0.21%, MnO = 0.17%, and Cr_2O_3 = 0.3% (Apollo 15, station 6 soil 15012; Morris et al. 1983). As SiO_2, TiO_2 and FeO are removed by electrolysis according to selected values for K_{Si} (= 0.48) and K_{Fe} (= 0), the composition of the residual melt moves into the spinel stability field whereupon spinel ($MgAl_2O_4$) precipitates from the melt. The values for steady state composition of the melt and product composition were derived by iterative calculations. First SiO_2, TiO_2 and FeO were removed from the melt according to the chosen values for K_{Si}, K_{Ti} and K_{Fe}. Then spinel was removed to bring the residual melt onto the spinel stability surface {which we approximated for temperature = 1450°C from Levin et al. [1969, p. 185] by the relationship MgO = ([Al_2O_3−15])/5·(13−2·x)+ (1−(Al_2O_3−15)/5)·(21−5·x) where x = (39−4·(wt%Al_2O_3−15)/5-wt%SiO_2)/7, valid for the values of Al_2O_3 and SiO_2 relevant in this calculation and ignoring thermodynamically nonideal contributions from TiO_2, Na_2O, K_2O, MnO and Cr_2O_3}. These calculations

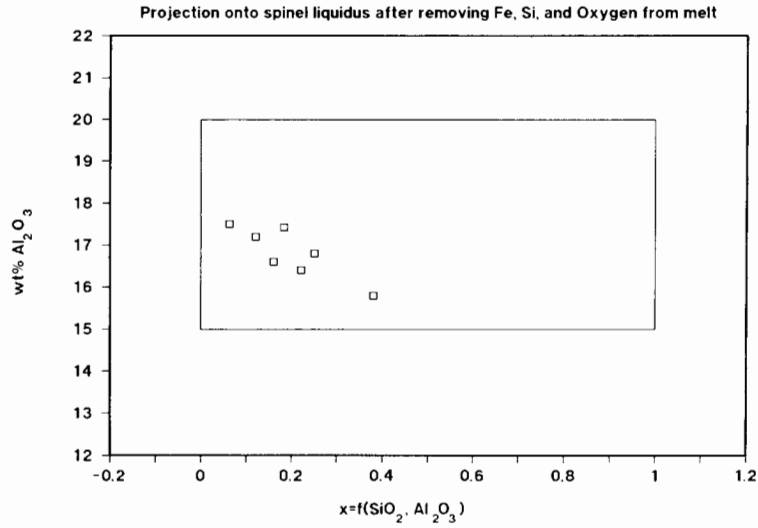


Figure 4. Theoretical calculation of the steady-state melt composition of several Apollo 15, station 6 soil compositions sampled from an area of about 700 m^2, illustrating that the electrolysis process is not sensitive to expected local variations in feedstock composition. The box indicates the region where the criteria for efficient electrolysis are met ($SiO_2 + Al_2O_3 < 55\%$ and $FeO < 2\%$) and spinel ($MgAl_2O_4$) is a stable phase. Axes of the figure are chosen such that the curved surface of the spinel liquidus boundary in the system SiO_2-Al_2O_3-CaO-MgO can be roughly plotted on this planar figure [$x = (39-4\cdot(\text{wt\% } Al_2O_3 - 15)/5 - \text{wt\%}SiO_2)/7$ and is intended to model the curvature of the spinel stability surface at 1450°C]. These projections are done as discussed in the text by iteratively removing FeO, TiO_2, and SiO_2 from the melt according to the equilibrium relations $K_{Fe} = 0$ and $K_{Si} = 0.48$ (according to Haskin et al. [1992], this corresponds to a potential of between 1.3 and 1.5 V), then projecting from spinel toward the 1450°C spinel liquidus surface (to model the equilibrium crystallization of spinel triggered by the removal of Fe and Si from the melt). Data are from Korotev (1987) and Morris et al. (1983).

were repeated iteratively until all equilibrium criteria were met simultaneously. For this example feedstock, products are the following:

25.3 wt% metal (59.5 wt%Si, 2.4%Ti, 38.2%Fe)
20.5% oxygen
13.3% spinel (71.8% Al_2O_3, 28.2% MgO)
40.9% residual (steady-state) melt (35.6% SiO_2, 1.1% TiO_2,
17.6% Al_2O_3, 16.1% MgO, 26.8% CaO,
1.1% Na_2O, 0.5% K_2O, 0.4% MnO, 0.7% Cr_2O_3).

Two questions are of interest. First: how much variation can there be in feedstock composition for a given set of electrolysis conditions (temperature, cell potential) without causing the steady-state compositions to deviate from the ranges defined above (that is, how sensitive is the process to feedstock

compositional variations)? Second: how wide a variation in feedstock can be accommodated by adjusting temperature and cell potential to bring the steady-state composition within the compositional range defined above (that is, to what extent can we use temperature and cell potential to compensate for more drastic changes in feedstock composition)?

Figure 4 illustrates that the process is not sensitive to expected local variations in feedstock composition. Over a region of about 700 m^2 at station 6, Apollo 15, variations in the soil composition are small enough that the steady states for all compositions are within the ranges defined above. The axes of Fig. 4 are chosen such that the stability field of spinel at 1450°C can be approximately plotted on the figure. The region of the spinel stability field that also meets the other criteria outlined above is shown by a box on the figure.

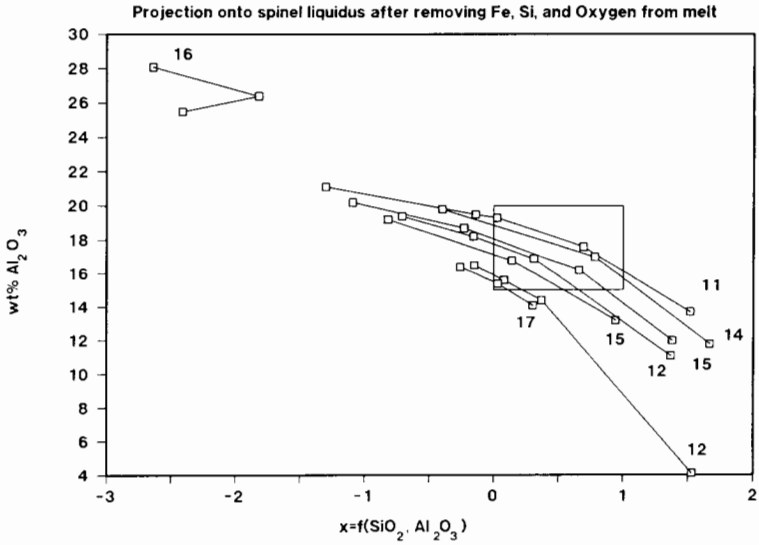


Figure 5. Illustration demonstrating that the electrolysis process is versatile with respect to various lunar soil compositions. Projection is done as in Fig. 4, but for each soil composition, the projection is calculated for a range of values of K_{Si}, generally somewhere between 0.4 and 0.8. This corresponds to fairly small variations in cell potential of less than 0.1 Volt. The numbers on the figure indicate from which Apollo mission the soil sample is taken (data are from Taylor 1982). Most of the soils can be brought into the box that delineates the compositional region where electrolysis is efficient and spinel is stable, but to do this requires a different value of K_{Si} for each composition.

When data for more soils, also from Apollo 15 and sampled over several square kilometers, are included, the results for some of them do not fall within the box. However, by adjusting the value of K (which as defined above is a function of cell potential and therefore can be controlled), the

steady state arising from these other soil compositions can be brought into the box on the figure. In fact, steady states deriving from many of the soils sampled by the Apollo missions can meet the demanding criteria defined above (Fig. 5), demonstrating the versatility of the electrolysis process with respect to feedstock composition. With small adjustments of cell potential, the steady-state composition for all the example soils except that from Apollo 16 can meet the criteria required for high conductivity, low viscosity, high production efficiency, and spinel on the liquidus.

We point out that whether or not spinel is ultimately chosen as a container material determines whether the rather stringent criteria of maintaining the steady state composition in equilibrium with spinel is actually required. If some other container is ultimately chosen, this criterion may be relaxed, or it may be desirable to keep some other phase present on the liquidus. This discussion is given as an example illustrating that the process is versatile with respect to feedstock composition and is not particularly sensitive to expected variations in feedstock composition over a small region.

Acknowledgments. We gratefully thank NASA for support for this research through UA/NASA Space Engineering Center for utilization of local planetary resources. We thank E. McCullough and R. Keller for helpful reviews of this manuscript.

REFERENCES

Binder, A. B. 1990. LLOX—Metal production via NaOH electrolysis. In *Engineering, Construction and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 339–346.

Bockris, J. O'M., Kitchener, J. A., and Davies, A. E. 1951. Ionic transport in liquid silicates. *J. Chem. Phys.* 19:255.

Bockris, J. O'M., Kitchener, J. A., and Davies, A. E. 1952*a*. Electric transport in liquid silicates. *Trans. Faraday Soc.* 48:536–548.

Bockris, J. O'M., Kitchener, J. A., Ignatowicz, W., and Tomlinson, J. W. 1952*b*. Electric conductance in liquid silicates. *Trans. Faraday Soc.* 48:75–91.

Bockris, J. O'M., MacKenzie, G. D., and Kitchener, J. A. 1955. Viscous flow in silica and binary liquid silicates; *Trans. Faraday Soc.* 51:1734–1748.

Bottinga Y., and Weill, D. F. 1972. The viscosity of magmatic silicate liquids: A model for calculation. *Amer. J. Sci.* 272:438–375.

Colson, R. O. 1990. Solubility of Ni° in silicate melts and implications for metal/melt and crystal/melt partitioning. *Ann. Meeting Geol. Soc. Amer., Abstract Booklet* 22:A164.

Colson, R. O., and Haskin L. A. 1990. Lunar oxygen and metal for use in near-Earth space: Magma electrolysis. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 187–196.

Colson, R. O., Haskin, L. A., and Crane, D. 1990. Electrochemistry of cations in diopsidic melt: Determining diffusion rates and redox potentials from voltammetric curves. *Geochim. Cosmochim. Acta* 54:3353–3367.

Criswell, D. R. 1983. A transportation and supply system between low Earth orbit and the Moon which utilizes lunar derived propellants. *Lunar Planet. Sci.* XIV:11–12 (abstract).

Cutler, A. 1985. Slag-Metal Equilibrium in Lunar Smelting and Arc Electrowinning. Soc. of Manufacturing Engineers Technical Paper EM85-978 (Dearborn, Mich.: SME).

Donaldson, C. H., Williams, R. J., and Lofgren, G. 1975. A sample holding technique for the study of crystal growth in silicate melts. *Amer. Mineral.* 60:324–326.

du Fresne, E., and Schroeder, J. E. 1983. Magma electrolysis. In *Research on the Use of Space Resources*, ed. W. F. Carroll, JPL Publ. 83-36, pp. 3-1–3-22.

Greiner, E. S., Marsh, J. S., and Stoughton, B. 1933. *The Alloys of Iron and Silicon* (New York: McGraw-Hill).

Grjotheim, K., Matiasovsky, K., Danek, V., and Stubergh, J. R. 1990. Electrochemical deposition of metals; alloys and oxygen from natural minerals—I. Physiochemical properties of molten cryolite-potassium feldspar mixtures. *Canadian Metallurgical Quart.* 29:39–42.

Haskin, L. A. 1985. Toward a Spartan scenario for use of lunar materials. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 435–443.

Haskin, L. A., and Colson, R. O. 1990. Lunar resources, toward living off the lunar land. In *Space Mining and Manufacturing*, ed. T. Triffet, pp. I12–I19.

Haskin, L. A., and Warren, P. 1991. Chemistry. In *Lunar Sourcebook*, eds. G. Heiken, D. Vaniman and B. French (Cambridge: Cambridge Univ. Press), pp. 357–474.

Haskin, L. A., Colson, R. O., Lindstrom, D. J., Lewis, R. H., and Semkow, K. W. 1992. Electrolytic smelting of lunar rock for oxygen, iron and silicon. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century, Vol. II*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 411–422.

Keller, R. 1988. Lunar production of aluminum, silicon and oxygen. In *Metallurgical Processes for the Year 2000 and Beyond*, eds. H. Y. Sohn and E. S. Geskin (Warrendale, Pa.: Minerals, Metals & Materials Soc.), pp. 551–562.

Keller, R., and Larimer, K. T. 1991. Experimental study of the electrolysis of silicate melts. In *UA/NASA SERC Annual Progress Report 1990–91 (Tucson: UA/NASA SERC), pp. IA-63–IA-69.*

Kesterke, D. G. 1971. Electrowinning of Oxygen from Silicate Rocks. Report of Investigations 7587 (Washington, D. C.: U. S. Bureau of Mines).

Koelle, H. H. 1988. The influence of lunar propellant production on the cost-effectiveness of cis-lunar transportation systems. In *Papers Presented to the Symp. on Lunar Bases and Space Activities in the 21st Century*, Houston, Tex., April 5–7, LPI Contrib. 652, p. 140 (abstract).

Korotev, R. L. 1987. Mixing levels, the Apennine front soil component, and compositional trends in the Apollo 15 soils. *Proc. Lunar Sci. Conf.* 17, *J. Geophys. Res. Suppl.* 92: E411–E431.

Levin, E. M., Robbins, C. R., and McMurdie, H. F., eds. 1969. *Phase Diagrams for Ceramists, Vol 2* (Columbus, Oh.: American Ceramic Society).

Lindstrom, D. J., and Haskin, L. A. 1979. Electrochemical preparation of useful material from ordinary silicate rocks. In *Space Manufacturing Facilities*, eds. J. Gray and C. Krop (New York: AIAA), pp. 129–134.

McCullough, E., and Mariz, C. 1990. Lunar oxygen production via magma electrolysis. In *Engineering, Construction and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil

Engineers), pp. 347–356.
Minenko, V. I., Petrov, S. M., and Ivanova, N. S. 1961. Behaviour of a platinum electrode in silicate melts. *Russian J. Phys. Chem.* 35:753–755.
Mueller, G. 1983. A giant foothold for mankind. In *Lunar Planet. Sci.* XIV:31 (abstract).
Morris, R. V., Score, R., Dardano, C., and Heiken, G. 1983. *Handbook of Lunar Soils*, NASA/JSC Publ. 67 (Houston: Johnson Space Center).
Oppenheim, M. J. 1968. On the electrolysis of molten basalt. *Mineral. Mag.* 36:1104–1122.
Oppenheim, M. J. 1970. On the electrolysis of basalt, II: Experiments in an inert atmosphere. *Mineral. Mag.* 37:568–577.
Rieger, P. H. 1987. *Electrochemistry* (Englewood Cliffs, N. J.: Prentice-Hall).
Schreiber, H. D. 1987. An electrochemical series of redox couples in silicate melts: A review and applications to geochemistry. *J. Geophys. Res.* 92:9225–9232.
Semkow, K. W., and Haskin, L. A. 1985. Concentrations and behavior of oxygen and oxide ion in melts of composition CaO-MgO-xSiO_2. *Geochim. Cosmochim. Acta* 49:1897–1907.
Semkow, K. W., and Haskin, L. A. 1986. Observations of Oxide Film Formation on Platinum Electrodes During Silicate Melt Electrolysis. McDonnell Center for the Space Sciences Report TEL-108 (St. Louis, Mo.: Washington Univ.).
Semkow, D. W., Rizzo, R. A., Haskin, L. A., and Lindstrom, D. J. 1982. An electrochemical study of Ni^{2+}, Co^{2+}, and Zn^{2+} ions in melts of composition $CaMgSi_2O_6$. *Geochim. Cosmochim. Acta* 46:1879–1889.
Sherwood, B., and Woodcock, G. R. 1991. Costs and benefits of lunar oxygen: Engineering, operations and economics. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 19.
Simnad, M. T., Derge, G., and George, I. 1954. Ionic nature of liquid iron-silicate slags. *J. Metals* 6:1386–1390.
Simon, M. C. 1985. A parametric analysis of lunar oxygen production. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 531–541.
Taylor, L. A., Cooper, B., McKay, D. S., and Colson, R. O. 1991. Oxygen production on the Moon: Processes for different feedstocks. In *Metallurgy Processing Fundamentals: Lunar Mining and Processing* (New York: Soc. of Mining, Metallurgy and Exploration), preprint 91-83.
Taylor, S. R. 1982. *Planetary Science: A Lunar Perspective* (Houston: Lunar and Planetary Inst.).

LUNAR OXYGEN EXTRACTION USING FLUORINE

WOLFGANG SEBOLDT and STEPHAN LINGNER
German Aerospace Research Establishment

STEPHAN HOERNES and WOLFGANG GRIMMEISEN
University of Bonn

REINHARD LEKIES and RALF HERKELMANN
Solvay Fluor und Derivate GmbH, Hannover

and

DONALD M. BURT
Arizona State University

Lunar soil and rocks are substantial resources for oxygen which could probably be exploited during future lunar activities. For the liberation of this oxygen, the use of fluorine gas as a strong reagent is discussed. Fluorine is capable of decomposing a variety of silicates. Results of fluorination experiments support the idea that nearly complete separation of elemental oxygen from lunar soil components at relatively moderate temperatures can be expected. Two processing concepts are outlined. They differ in several aspects, especially with regard to beneficiation needs and recycling strategies. Recycling of fluorine should be achieved to a high degree to make the process economically promising. Reduction of reaction by-products for fluorine recycling is proposed to be carried out by either sodium or atomic hydrogen, producing also useful metals and other materials. For both concepts fluorine recovery (final electrolysis of resulting NaF or HF) seems to be the most energy-demanding step. Handling of fluorine and its compounds should pose no major problems, inasmuch as several container materials are known which effectively resist corrosion.

I. INTRODUCTION

Processing and utilization of local planetary material is considered to be a key technology, possibly enabling enhanced future space activitites in the planetary system and, particularly, on the Moon. This technology might lead, for example, to cost-reduced scenarios for manned lunar bases.

Oxygen is an interesting element, both as fuel component for space transportation and as a gas for life support. Given the deficiency of lunar volatiles,

extraction of oxygen from the lunar lithosphere is considered, which contains an average of about 45% O_2 by weight. Useful by-products might include: structural materials such as iron, magnesium, titanium and aluminum; silicon for solar cells; possibly lime (CaO) for cement or ceramics. The nature of lunar raw material is known from returned soil and rock samples of Apollo- and Luna-landing missions (S. R. Taylor 1975,1982). It is a mixture of glassy melt particles, mineral fragments and their comminuted source rocks and consists almost entirely of complex metal oxides. Crystalline silicates are made up of tetrahedral $(SiO_4)^{4-}$ units with Si^{4+} in the center and interlattice metal ions for charge balance. The Si-position is partly substituted by Al in some minerals (e.g., feldspars). Silicate structures are exceptionally stable and are inert against many chemical agents.

Various processing concepts have been developed in the past (see other chapters in this book) which are based upon treatment of lunar soil, rocks or their components by physical, chemical or electrochemical means. These processes differ also in the demand for exploration and beneficiation efforts as well as in their output of several (usable) by-products (metals, ceramics, etc.).

II. THE FLUORINATION PROCESS

One possibility of liberating elemental oxygen from lunar silicates is the reaction with a stronger oxidizing agent than O_2 itself. Fluorine (chemical symbol: F, molecular formula: F_2) is the lightest halogen and the most reactive element of the periodic system. It is one of the strongest oxidizing agents, thus even capable of decomposing silicates.

The oxygen-fluorine exchange is shown by the following idealized reaction scheme (M = metal):

$$\underset{\text{(metal oxides)}}{M_{xy}O_y + yF_2} \longrightarrow \underset{\text{(metal fluorides)}}{xyMF_{2/x} + y/2O_2} \tag{1}$$

For aluminum oxide (x=2/3, y=3), e.g., this equation becomes:

$$Al_2O_3 + 3F_2 \longrightarrow 2AlF_3 + 3/2O_2. \tag{2}$$

From stoichiometry each mole O_2 is substituted by two moles F_2 in a complete reaction, changing metal oxides to corresponding simple and mixed-element fluorides, releasing oxygen as gas. The reaction itself is exothermic and runs spontaneously; reaction rates are, however, strongly temperature dependent.

The amount of fluorine needed depends upon the molar oxygen contents of the minerals as shown in the simplified equations below[a]:

[a] Mixed-element fluorides (e.g., $CaAlF_5$) might also occur as well as ferric fluorides (FeF_3).

Feldspar (anorthite):

$$CaAl_2Si_2O_8 + 8F_2 \longrightarrow CaF_2 + 2AlF_3 + 2SiF_4 + 4O_2 \quad (3)$$

Pyroxene (augite):

$$(Ca, Mg, Fe)SiO_3 + 3F_2 \longrightarrow (CaF_2, MgF_2, FeF_2) + SiF_4 + 1.5O_2 \quad (4)$$

Olivine:

$$(Fe, Mg)_2SiO_4 + 4F_2 \longrightarrow 2(FeF_2, MgF_2) + SiF_4 + 2O_2. \quad (5)$$

Parts of the reaction by-products give a solid mixture of metal fluorides, whereas SiF_4 will be gaseous at room temperature. The distribution of fluorides naturally depends upon the composition of the initial silicate material.

Hydrofluoric acid has been applied in laboratories to leach natural rock samples for their chemical analysis by conventional wet methods. Oxygen extraction with fluorine and halogen-fluorides (BrF_5, ClF_3) is a proven technique in analytical chemistry, e.g., for measuring the O_2-content in metal-oxides and to release oxygen for investigations of isotopic compositions in rocks and minerals (Baertschi and Silverman 1951; Taylor and Epstein 1962). Geochemists, who demand total oxygen extraction for isotope analyses, have found no better reagent despite considerable experimentation. The fluorination is typically carried out at temperatures between 700 K and 950 K in nickel reaction vessels and runs over a period of hours or days.

Fluorine is inexpensive and abundant on Earth (Kilgore et al. 1985), where it is typically extracted from the natural mineral fluorite, CaF_2 (Ellis and May 1986). The fluorite is reacted with sulfuric acid, H_2SO_4, to yield hydrofluoric acid, HF, plus calcium sulfate, $CaSO_4$; the hydrofluoric acid (in a molten salt bath with potassium fluoride, KF) is then electrolyzed to yield hydrogen gas at the cathode and fluorine gas at the anode. The fluorine gas evolved is stored in iron or nickel containers until used.

A major user is the aluminum industry, where fluorides act as fluxes for Al_2O_3 electrolysis. The reason is that the fluoride ions are about the same size as the oxide ions (fluorides thus dissolve oxides), that fluoride compounds typically have much lower melting temperatures and that the melts have lower viscosities and better electrical conductivities than oxide melts. Most of the metal fluorides are nonvolatile, but silicon tetrafluoride, SiF_4, is a gas at normal conditions.

III. SUGGESTED USES FOR FLUORINE IN SPACE

The unusual properties of fluorine and its compounds have led to a number of suggested uses related to space exploration. The simplest is the use of fluoride molten salts as a storage medium for solar heat (see, e.g., Whittenberger and Misra 1987; Boyle et al. 1988; Misra 1988). A more complex use might be

in a molten salt electrolytic cell for oxygen extraction (see, e.g., Semkow and Sammells 1987; Tsai et al. 1990). Thirdly it could be used as a flux to facilitate silicate electrolysis (Kesterke 1970 and 1971; this parallels the use of fluorite as a flux in the steel industry) or as a molten salt solvent for oxides (Jarrett et al. 1980; Keller 1988; this parallels the use of synthetic fluoride molten salts in the aluminum industry, mentioned above). Finally, of course, fluorine either as hydrofluoric acid, HF (see, e.g., Waldron 1985), or fluorine gas, F_2 (Dalton and Degelman 1972; Downs 1972; O'Donnell 1972; Burt 1988,1992), has been proposed for oxygen liberation from lunar rocks. Geochemists have successfully used fluorine gas for oxygen extraction throughout the last three decades, without major safety problems.

IV. RESULTS OF FLUORINATION EXPERIMENTS

As already mentioned, experience has been gained in the fluorination method with oxygen extraction from terrestrial rock samples and also from original lunar material and corresponding terrestrial simulants (Taylor and Epstein 1970; Epstein and Taylor 1971,1975; O'Donnell 1972).

From the early work of Taylor and Epstein (1962) it is known that different minerals show different resistance to fluorine attack. Framework- and chain-silicates (e.g., feldspars and pyroxenes) are easily decomposed at temperatures around 670 K while some orthosilicates like olivines have significant lower reaction rates at the same conditions. This behavior is thought to be due to different crystal structure and silica content. Silicon is removed from the sites of reaction as gaseous SiF_4, which creates continuously fresh surfaces for the attack of hot fluorine. Fluorine resistant minerals are believed to form solid reaction products at exposed surfaces, protecting the interior from further decomposition. These passivation effects clearly increase with growing grain size.

The first experimental attempt to extract oxygen from simulated lunar surface material with fluorine was carried out by O'Donnell (1972). The diffusion of fluorine into coarse-grained minerals (about 250 μm) was directly observed under a microscope. Areas of changed optical properties were observed to form on prominent surface structures and along cracks. These areas were interpreted as fluoride rims. The experiments gave very low O_2 yields (only up to 4%) due to the low temperatures ($\leq$523 K) and the reduced fluorine pressures chosen.

Haimson and Knauth (1983) showed a clear linear correlation of oxygen yields with available fluorine in the fluorination procedure. Nearly total oxygen release (in 12 hr at 723 K) from different terrestrial SiO_2-modifications was obtained with a molar fluorine to oxygen ratio of more than 2.

Recently, fluorination experiments with silicates (Seboldt et al. 1991) were carried out to study the dependence of oxygen yields with reaction temperature and feedstock composition in order to evaluate favorable conditions for a future lunar oxygen production plant. Four lunar soil simulants were used

consisting half of anorthositic or Ti-basaltic glasses (Corning Glass Works 1989). Variable amounts of bytownite (An 79-feldspar), augite (clinopyroxene) and Mg-rich olivine (Fo 90) (Flucks and Deutsch 1990) were admixed in various proportions. The chosen grain size distribution was between 63 μm and 90 μm, representing typical particle sizes of lunar soil (Carrier et al. 1973). The resulting simulants contained similar amounts of oxygen as well as silicon and differed mainly in Al-, Mg- and Fe-contents (see Fig. 1), thus reflecting characteristics of typical lunar highland and mare sites.

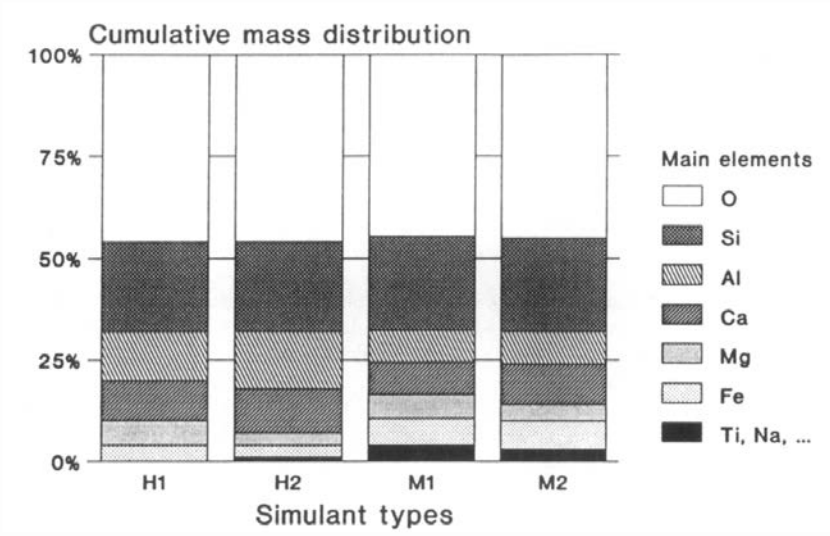


Figure 1. Chemical composition of lunar soil simulants H1, H2, M1, M2, representing different lunar highland (H) and mare (M) regions.

The fluorination line at University of Bonn originally dedicated to routine oxygen isotope analyses of terrestrial material was used for the experiments. It is similar to the first fluorination line described by Taylor and Epstein (1962). One major difference is the introduction of a fluorine purification part (Asprey 1976), improving isotopic analyses results. Multiple fluorination runs of the four prepared samples at five temperature levels between 623 K and 923 K revealed strong temperature dependence of the oxygen yields (Fig. 2), ranging between 7% and 100%. Other reaction parameters were kept constant. Saturation effects are clearly visible from Fig. 2, occurring around 700 K and leading to an oxygen release of 70 to 80% at this temperature. Lunar soil simulants H2 and M2 allow near total oxygen release at temperatures above 800 K. This behavior can be interpreted as a consequence of their olivine-free batch composition.

In summary, one can conclude from the experiments that about 80% of oxygen is extractable from lunar soil of any site at relatively moderate temperatures of 720 K. Using olivine-free highland soils, even total oxygen release

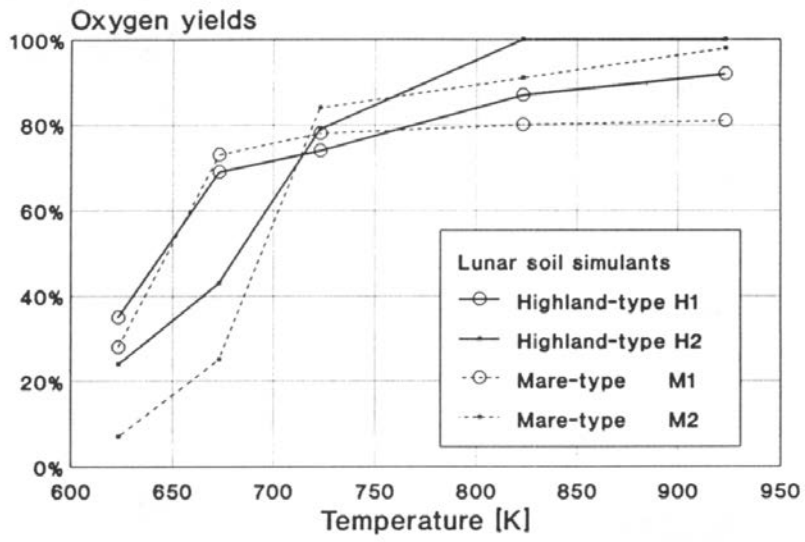


Figure 2. Temperature dependence of oxygen yields from fluorinated olivine-bearing (H1, M1) and olivine-free (H2, M2) soil simulants; F2-pressure 2.5 bar, reaction time 16 hr, sample mass 20 mg, reaction volume 20 cm^3.

may be attained at somewhat higher temperatures (820 K). These conclusions drawn from laboratory experiments, however, hold strictly true only for small sample masses (several ten milligrams). Extrapolation to industrial processing conditions with tons of raw material is difficult and further experiments with larger quantities of soil simulants are needed. Furtheron, technical solutions must be found to handle these large quantities in specific processing facilities placed in extraterrestrial environments.

V. FLUORINATION CONCEPTS FOR LUNAR OXYGEN PRODUCTION

This section describes two processing schemes proposed for extraterrestrial oxygen production using fluorine. Regarding the intended oxygen extraction both concepts are based upon the same reaction principle, the substitution of oxygen by fluorine. They differ mainly in the need for beneficiation of the raw material and the method of fluorine recovery. The major application seems to be a lunar base scenario, but the processes could also be adapted to other extraterrestrial bodies, containing significant amounts of silicates on the surface.

The (metal) elements present in lunar rocks differ in their tendency of reaction with oxygen, fluorine and the tendency of fluorine-oxygen exchange. They form three distinct sequences as shown in Fig. 3, using the chemical potential μ (corresponding to the energy liberated during reaction) at 800 K

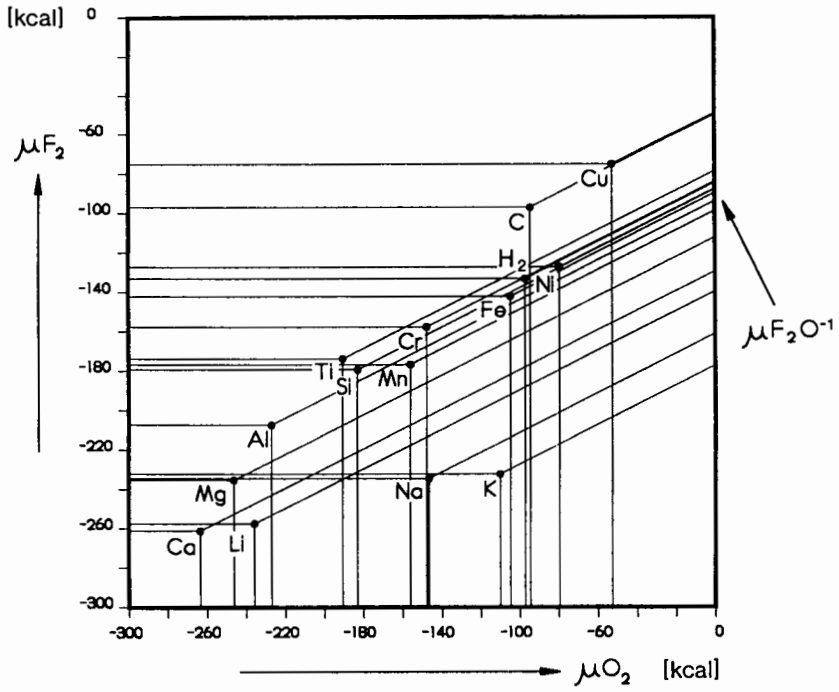


Figure 3. Chemical potential (μ-μ) diagram (in kcal) showing the relative affinities of some lunar and other elements for oxygen (vertical lines intersecting horizontal axis), fluorine (horizontal lines intersecting vertical axis) and for fluorine-oxygen exchange (slanting lines extending to upper right) at 800 K and 1 atm. Data from Pankratz et al. (1984).

and 1 atm. Reaction tendencies increase with decreasing μ values. Vertical lines intersecting the horizontal axis represent the sequence for reactions of the type

$$2xM + O_2 \longrightarrow 2/yM_{xy}O_y \tag{6}$$

and horizontal lines for those of the type

$$xM + F_2 \longrightarrow xMF_{2/x}. \tag{7}$$

Elements with lower μ values generally are capable of reducing oxides/fluorides of elements with higher μ values. Slanting lines of slope 1/2 (due to the molar oxygen-fluorine exchange factor of 0.5) generated at the intersections of the vertical and horizontal lines represent the sequence for oxygen-fluorine exchange reactions (described in Sec. II) with increasing tendency from upper left to lower right in the diagram. For these exchange reactions μF_2O^{-1} values result from $\mu F_2O^{-1} = \mu F_2 - 1/2\mu O_2$. For example, the oxygen-fluorine exchange reaction for calcium ($\mu F_2O^{-1} \approx -130$ kcal) is clearly stronger than for iron (≈ -90 kcal). Positive μF_2O^{-1} values exclude oxygen-fluorine exchange. Because only simple compounds are considered in Fig. 3, conclusions

for reactions involving complex compounds (silicates, mixed-element fluorides) should be drawn, however, with caution.

A. Stepwise Fluorination and Sodium Reduction

At equilibrium, the stages of stepwise fluorination of feedstock of a given composition can be predicted from thermodynamic data or, where these are lacking, from the stable fluorine-rich mineral assemblages seen in nature. An appropriate lunar feedstock would appear to be the mineral anorthite, ideally $CaAl_2Si_2O_8$, the most abundant ingredient of the anorthosite that forms the lunar highlands. Restricting the feedstock to a single mineral or to a simple mixture of minerals simplifies the recycling of the fluorine (inasmuch as all fluorine must be brought from Earth at great expense, all should be recycled, at least in principle).

For anorthite, the probable fluorination steps are as follows (Burt 1992):

$$CaAl_2Si_2O_8 + F_2 \longrightarrow CaF_2 + Al_2SiO_5 + SiO_2 + 1/2O_2 \tag{8}$$

$$Al_2SiO_5 + F_2 \longrightarrow Al_2SiO_4F_2 + 1/2O_2 \tag{9}$$

$$4CaF_2 + Al_2SiO_4F_2 + 2F_2 \longrightarrow 2Ca_2AlF_7 + SiO_2 + O_2 \tag{10}$$

$$2Ca_2AlF_7 + Al_2SiO_4F_2 + 2F_2 \longrightarrow 4CaAlF_5 + SiO_2 + O_2 \tag{11}$$

$$Al_2SiO_4F_2 + 2F_2 \longrightarrow 2AlF_3 + SiO_2 + O_2 \tag{12}$$

$$SiO_2 + 2F_2 \longrightarrow SiF_4 + O_2. \tag{13}$$

Each of the above reactions releases one mole of O_2 for each two moles of F_2 put in, and each except reaction (9) and reaction (13) SiO_2 (as intermediate product). Many of the phases produced at intermediate steps correspond to minerals (some to several mineral modifications, depending on temperature); e.g., CaF_2 is fluorite, Al_2SiO_5 is andalusite (or sillimanite), SiO_2 is quartz (or tridymite or cristobalite), and $Al_2SiO_4F_2$ is topaz. How many of these well-ordered phases would actually be formed in a heated nickel reaction vessel remains to be determined. If they do form, the progress of stepwise fluorination in going down a fixed or fluidized bed reactor (with F_2 gas fed in the bottom and anorthite charge fed in the top) could be followed by monitoring the charge for the "index fluoride" formed at each stage. In order, these would be (1) fluorite, (2) topaz, (3) Ca_2AlF_7, (4) $CaAlF_5$, (5) AlF_3 and (6) SiF_4 (in product gas). In any case, the overall or complete fluorination reaction would probably be:

$$\underset{\text{(anorthite)}}{CaAl_2Si_2O_8} + 8F_2 \longrightarrow \underset{\text{(mixed fluorides)}}{CaAlF_5 + AlF_3} + \underset{\text{(gas mixture)}}{2SiF_4 + 4O_2}. \tag{14}$$

Complete fluorination of any other lunar silicate would similarly produce a series of crystalline or molten mixed fluorides plus a gas mixture consisting mainly of O_2 and SiF_4 (plus unreacted F_2). A major task is to remove the

unwanted fluorine-bearing species from the gas mixture and then to recycle them plus the condensed fluorides back to a starting material from which fluorine gas can again be obtained.

In the sodium reduction process (Burt 1989,1992) fluorine would be brought from Earth as the salt NaF, then electrolyzed on the Moon to Na metal and F_2 gas. The fluorine gas alters the mineral feedstock to mixed fluorides plus oxygen gas, then the Na metal reduces the mixed fluorides to metals plus NaF, thus closing the loop chemically. Electrical energy feeds the process (via electrolysis of the NaF), and O_2, metals and oxides are the products, produced chemically from the feedstock. Transport or storage of fluorine as gas is avoided.

In detail the process is somewhat complex, even for a single-mineral feedstock such as anorthite, inasmuch as the fluorination product SiF_4 is a gas that must be removed from the O_2 gas product (together with excess F_2), and the fluorination product CaF_2 cannot be reduced by Na metal, but must be ion exchanged to CaO in an intermediate step.

The proposed processing reactions (Burt 1992) are listed below (per mole of anorthite to be treated; *l*: liquid, *g*: gaseous):

1. Electrolysis of NaF:

$$16\text{NaF}(l) \longrightarrow 16\text{Na} + 8\text{F}_2(g) \tag{15}$$

2. Fluorination of anorthite (oxygen production):
   a. Partial fluorination (F_2-scrubbing at very top of column):

$$\text{CaAl}_2\text{Si}_2\text{O}_8 \text{ (excess)} + \text{F}_2 \text{ (in F}_2 - \text{O}_2 \text{ gas mix)} \longrightarrow$$

$$\text{CaF}_2 + \text{Al}_2\text{SiO}_5 + \text{SiO}_2 + 1/2\text{O}_2(g) \tag{16}$$

   b. Complete fluorination (lower in column):

$$\text{CaF}_2 + \text{Al}_2\text{SiO}_5 + \text{SiO}_2 + 7\text{F}_2(g) \longrightarrow$$

$$\text{CaAlF}_5 + \text{AlF}_3 + 2\text{SiF}_4(g) + 7/2\text{O}_2(g) \tag{17}$$

3. Scrubbing of SiF_4:

$$2\text{SiF}_4 \text{ (in O}_2 - \text{F}_2 - \text{SiF}_4 \text{ gas mix)} + 4\text{NaF} \longrightarrow 2\text{Na}_2\text{SiF}_6 \tag{18}$$

4. Silicon production:

$$2\text{Na}_2\text{SiF}_6 + 8\text{Na} \longrightarrow 2\text{Si} + 12\text{NaF} \tag{19}$$

5. Aluminum production:

$$\text{CaAlF}_5 + \text{AlF}_3 + 6\text{Na} \longrightarrow \text{CaF}_2 + 2\text{Al} + 6\text{NaF} \tag{20}$$

6. Production of sodium oxide (for step 7):

$$2Na + 1/2O_2(g) \longrightarrow Na_2O \tag{21}$$

7. Lime production:

$$CaF_2 + Na_2O \longrightarrow CaO + 2NaF \tag{22}$$

8. Final fluorine scrubbing (cycle CaF_2 to step above):

$$F_2 \text{ (trace in } O_2 - \text{gas)} + CaO \longrightarrow CaF_2 + 1/2O_2(g). \tag{23}$$

Overall:

$$CaAl_2Si_2O_8 \longrightarrow CaO + 2Al + 2Si + 7/2O_2(g). \tag{24}$$

There is only a single initial NaF-electrolysis step, and most of the remaining 7 chemical steps produce NaF for step 1. In practice, the final fluorine scrubbing in step 8 could be performed with Na_2O from step 6, or could be omitted if sufficient scrubbing were performed in step 2a. In practice also, step 3 might be replaced by some sort of temperature-dependent condensation reaction to separate SiF_4, in which case step 4 would become:

$$2SiF_4(g) + 8Na \longrightarrow 2Si + 8NaF. \tag{25}$$

Finally, steps 6 and 7 could probably be combined into a single bulk reaction:

$$CaF_2 + 2Na + 1/2O_2(g) \longrightarrow CaO + 2NaF. \tag{26}$$

Under ideal circumstances, then, the eight reactions and reaction vessels implied by the numbered steps listed above could be replaced by only 6. Further simplifications might be possible if fewer metallic products were desired or if a different beneficiated feedstock were used.

In the scheme presented above, for each mole of anorthite to be processed, the plant is required to contain 20 moles of NaF (16 for electrolysis and 4 for scrubbing SiF_4). From formula weights, each ton of anorthite per cycle would then require 3.02 tons of NaF. All of the NaF would be recycled, so the real control on production capacity would be the length of each cycle (i.e., the amount of electricity available for electrolysis and the rapidity of the chemical reactions). The experiments needed to yield such data have not yet been performed.

B. Fluorination and Hydrogen Reduction

In order to avoid major beneficiation of lunar soil a fluorination process with fluorine recycling via hydrogen reduction has been proposed (Seboldt et al. 1991), consisting of a fluorination step and two successive fluorine recovery steps (see Fig. 4). Nearly half of the soil mass mined might eventually be released as oxygen by this method. Other abundant elements like silicon,

aluminum, calcium, magnesium and iron would be accessible, too. Base siting aspects could almost be neglected. If possible, olivine-bearing soil should be avoided due to the depleted reactivity (Seboldt et al. 1991). Most surface areas of lunar highlands and some mare regions are almost totally free of olivine. Eventually, the separation of larger rock fragments by sieving as a simple beneficiation effort could be practiced to give improved fluorination results.

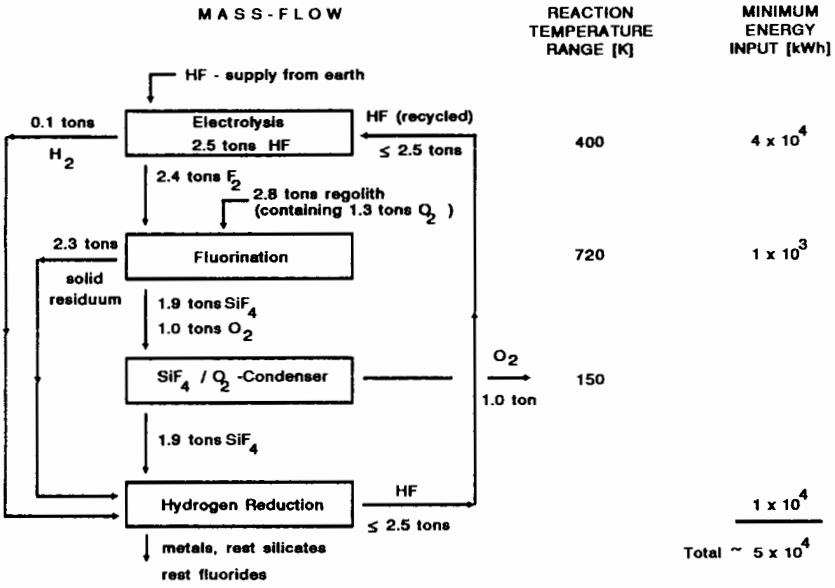


Figure 4. Mass and energy budget per metric ton of extracted lunar O_2, assuming 80% oxygen yields. If necessary a further F_2-recycling process via molten salt electrolysis may be added in connection with the hydrogen reduction step.

As mentioned in Sec. IV, technical realization concepts for large-scale fluorination facilities have to be developed in the future; conceivable are batch or continuous modes of processing. The fluorine-oxygen exchange will run according to the following idealized equation (s: solid, g: gaseous):

$$[SiO_2 \cdot Al_2O_3 \cdot CaO \cdot MgO \cdot FeO](s) + F_2(g) \overset{\approx 700\ K}{\longrightarrow}$$

$$O_2(g) + SiF_4(g) + AlF_3(s) + CaF_2(s) + MgF_2(s) + FeF_2(s). \qquad (27)$$

Parts of the solid reaction products might consist of mixed-element fluorides as well. Recall that ilmenite-bearing lunar soil of several maria (e.g., Mare Tranquillitatis) contains additionally a significant TiO_2-component, which would be reacted to its corresponding fluoride TiF_4, volatile at the

TABLE I
Reaction By-Products of Lunar Soil[a]

Compound	Concentration [wt. %]
SiF_4[b]	50.4
AlF_3	13.9
CaF_2	9.1
MgF_2	9.4
FeF_2	12.7
TiF_4[b]/NaF	4.5

[a] Idealized; calculated for total fluorination of Apollo 12-soil.
[b] Gaseous at fluorination temperature.

temperature of reaction. An example for the expected idealized distribution of fluoride reaction products is shown in Table I.

Oxygen and silicon tetrafluoride (and eventually TiF_4) evolve as volatile products, preventing coating and passivation effects on the grains. O_2 should be separable from the gas mixture by cooling and fractional distillation. SiF_4 (and TiF_4) will be trapped below 178 K (SiF_4-sublimation point) whereas oxygen will leave the condenser as gas above 90 K. Traces of excess fluorine in the cleaned volatile fraction may be removed by passing the gas through a CaO scrubber if pure oxygen (e.g., for life support) is needed. A second possibility of trapping trace fluorine is based on the selective absorption behavior of K_3NiF_6, which will react with fluorine around 520 K (Asprey 1976). Regeneration of the active nickel salt (and fluorine recovery) is possible by simple heating ($\approx$770 K). The remaining solid fluorination residue consists of mixed metal fluorides and may contain small amounts of metal fluorosilicates (topaz, actinolite) and/or silicates in case of incomplete reaction.

In order to recycle most of the fluorine, atomic hydrogen as a powerful reducing agent (Jones et al. 1973) is proposed combined with electrolysis of the reaction product HF (Seboldt et al. 1991). Before, molecular hydrogen must be dissociated into its atoms, e.g., by thermal, electric or electromagnetic means:

$$H_2 + \text{energy (104 kcal/mol)} \longrightarrow 2H\cdot \tag{28}$$

Effective production and use of atomic hydrogen is realized on Earth by discharging methods or arcs (Wood's and Langmuir's principles) with decomposition yields near 95% (Hollemann and Wiberg 1976). Considering the dissociation energy of H_2 the chemical potentials μO_2 and μF_2 for atomic hydrogen should be significantly lower than those for H_2 (compare Fig. 3). Therefore the resulting gas of energy-rich, highly reactive hydrogen radicals should be capable of reducing metal fluorides to their corresponding elements. Hydrogen fluoride is produced as volatile product and metals remain as solid residue:

$$MF_x + xH\cdot(g) \longrightarrow M + xHF(g). \tag{29}$$

From thermodynamic data, most of the occurring fluorides should react

with atomic hydrogen at moderate temperatures (≥300 K). MgF_2 will need much higher reaction temperatures up to 3000 K (*Handbook of Chemistry and Physics* 1989/90). CaF_2 is still inert against atomic hydrogen attack at these high temperatures. The reduction of most fluorides by atomic hydrogen is therefore favored from a thermodynamic point of view, giving metals as by-products and HF for further processing. The decomposition of the most abundant silicon tetrafluoride by this method is described in the literature (Shirai et al. 1989; Siemens AG 1981; General Electric Co. 1953). The reduction of other fluorides with the proposed method has to be investigated and demonstrated in future experiments.

The produced hydrogen fluoride from the hydrogen reduction step may be decomposed by electrolysis:

$$2HF + \text{energy (130 kcal/mol } F_2) \longrightarrow H_2 + F_2. \tag{30}$$

Both recovered elements, hydrogen and fluorine, could be recycled (Fig. 4). This process is carried out commercially on Earth for production of elemental fluorine by molten salt electrolysis of anhydrous potassium fluoride (KF·xHF) with varying concentrations of the hydrogen fluoride component (Kirk-Othmer 1980), the electrical conductivity is provided by KF. Basically, three temperature regimes are considered for fluorine-generating cells, from which the medium temperature type (working between 330 K and 380 K) is the commercial state of the art (for details see Kirk-Othmer 1980). This molten salt electrolysis is a very energy-consuming process, requiring electrical energy between 45 and 75 kJ per kg fluorine (Jaccaud et al. 1988).

Presuming complete recovery of fluorine from fluorides except from MgF_2 and CaF_2, about 85% of bulk fluorine may be recovered, as calculated for Apollo soil (see Fig. 5). In order to improve F_2-recycling further, the additional fluorine recovery from residual solid metal fluorides via molten salt electrolysis may be necessary, eventually allowing near total recycling of fluorine. For example, the molten salt electrolysis of CaF_2 soluted in an alkali metal-tetrafluoroborate melt is described for the synthesis of elemental fluorine (DuPont 1989) and could be considered for total lunar fluorine recovery from mixed solid fluorides including CaF_2:

$$MF_x + \text{energy} \longrightarrow M + 0.5xF_2. \tag{31}$$

An interesting further possibility could be to combine concepts for oxygen extraction by direct magma electrolysis with those for fluorine recovery in some sort of hybrid extraction procedure as fluorides are a good flux for silicate and oxide melts (A. Hepp, personal communication).

Concerning the mass flow of a future fluorination and oxygen production pilot plant, the following estimates can be made, assuming that oxygen extraction is run with 80% yields. Approximately 2.8 metric tons of lunar soil have to be mined and processed in the fluorination reactor together with

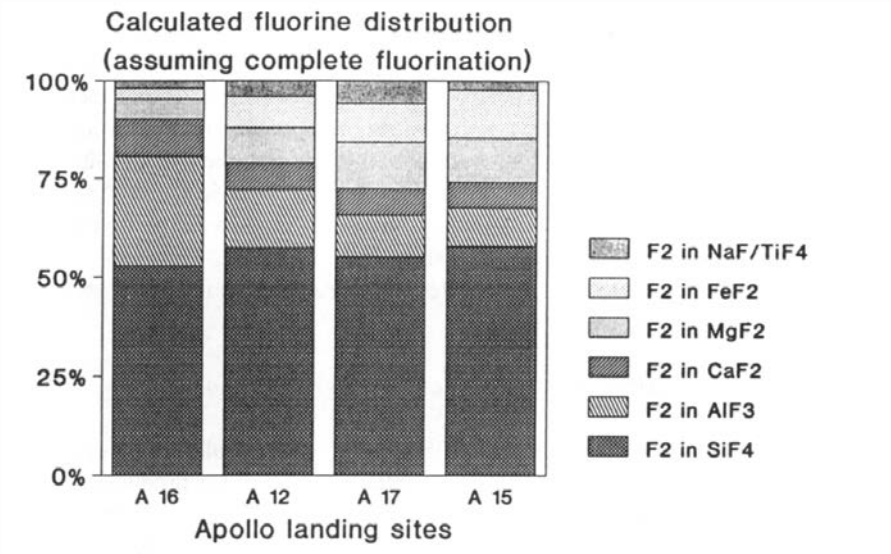


Figure 5. Calculated percentage of fluorine contained in different fluorides (idealized), assuming complete fluorination of soil from four Apollo landing sites (sample data from S. R. Taylor [1975] and Jaumann [1988]).

2.4 tons of fluorine per metric ton of oxygen produced (see Fig. 4). The necessary fluorine is separated in the electrolysis cell from 2.5 tons of HF. Therefore, instead of F_2 initially HF could be brought from Earth. Fluorine losses would have to be compensated by supply from Earth. For economic reasons, however, recovery of most processed fluorine is absolutely necessary.

In-situ mining of (lunar) fluorine could be considered but seems not very promising due to its low content in the bulk lunar soil and rocks (traces around 200 ppm according to Reed et al. [1970]). Somewhat higher concentrations of F_2 might be expected in the agglutinates fraction of the lunar soil, in lunar surface deposits of pyroclastic volcanic glasses (Apollo 17-region) and in the slightly differentiated highland crust section (Apollo 14- and Apollo 15-region) of the western hemisphere (see Chapter on lunar ores by Haskin and Colson). But in any case, local concentrations of fluorine-bearing phases like those on Earth are not known and not expected to be found in the primitive lunar crust.

According to Fig. 4, the most energy demanding step seems to be the fluorine recycling, especially the electric supply inside the electrolysis cell. The total energy input is estimated to be some 10^4 kWh per ton oxygen produced, which corresponds at least to mean power requirements of several hundred kW for an annual lunar O_2-production of 100 metric tons (Seboldt et al. 1991). For batch-mode processing power peaks could eventually be minimized by realizing a semi-continuous production process with several

identical reaction chambers in successive operation and by reuse of processing heat.

VI. HANDLING OF FLUORINE AND ITS COMPOUNDS

Today fluorine (F_2), hydrogen fluoride (HF) and fluorine compounds are established as major industrial chemicals. There are several materials known which solve the corrosion problems posed by contact with fluorine or its compounds (*Dechema Corrosion Handbook* 1989). The resistance of some metallic materials against fluorine attack depends upon the formation of protective layers of metal fluorides. These can be created by careful exposure to F_2, passivating the metal and thus preventing the subsurface from further corrosion (Taylor and Epstein 1962). Metal surfaces designed to be in contact with fluorine have to be cleaned thoroughly from traces of oil or grease beforehand (Solvay Product Information 1988).

Table II shows the corrosive effects on certain metals and metal alloys exposed to fluorine gas at varying temperatures and at atmospheric pressure. The data given are valid for materials with clean, smooth and passivated surfaces.

TABLE II

Corrosion Rates [mm/month] for Various Metals[a]

Temp. [K]	473	573	673	773	873	973
Nickel	—	—	0.017	0.125	0.71	0.83
Monel	—	—	0.012	0.049	1.5	3.7
Inconel	—	—	0.93	1.5	4.2	12.5
Copper	—	—	3.2	3.9	24	71
Alumin.	—	—	0	0.32	0.43	—
Magnes.	0	0	—	—	—	—
Steel:						
low-C	0	0.22	0.59	—	—	—
stainl.	0	0	0.76	13.7	—	—

[a] Data from Myers and DeLong 1984; —: not determined.

The findings clearly reveal the temperature dependent performance of metals and metal alloys listed in Table II. At high temperatures (above 670 K), aluminum, nickel and monel are significantly more resistant to corrosion than other metals. Therefore they should be used for vessels and valves under high-temperature (and elevated pressure) conditions, as typically occurring in fluorination processes. The other metals may be used at lower temperatures for pipes, tubes and fittings. Copper and low-carbon steel are particularly suitable for this purpose. Metals like brass, bronze, hard solders or silicates (quartz, glasses) may be useful too below 370 K. The only nonmetallic material resisting fluorine attacks at high temperatures ($\approx$1270 K) is highly sintered

clay. PTFE (poly-tetrafluorethylene) has also given excellent performance for pump and valve packings (Kirk-Othmer 1980).

VII. SUMMARY AND CONCLUSIONS

Fluorination of lunar raw materials is considered to be an interesting candidate process for future industrial oxygen production on the Moon. The great advantage of fluorine as a reagent is its high reactivity, which could lead to nearly total oxygen release from lunar silicates at relatively moderate temperatures. Owing to its inaccessibility on the Moon, recycling of reacted fluorine from by-product metal fluorides, however, must be achieved to a high degree to realize cost benefits. This seems to be the most critical requirement. In any case, certain amounts of fluorine or fluorine compounds for initiation of the process and for balance of losses have to be supplied from Earth.

The described processing concepts differ in several aspects, especially with regard to beneficiation effort and recycling strategies. Reduction of fluoride reaction by-products is considered by use of sodium or atomic hydrogen. The sodium reduction method implies transport and storage of fluorine as an inert salt, NaF, which has to be molten and electrolysed in the first processing step. Generation of fluorine by molten salt electrolysis of NaF is as yet an unproven technology. The main problem is to develop resistant anode-surfaces for fluorine evolving electrodes because of the caustic behavior of fluorine in a hot environment. Corrosion could eventually be reduced also by carrying out the NaF electrolysis at relatively low temperatures in a minimum-melting eutectic salt mixture of NaF and other fluorides. The sodium reduction concept is proposed to be used on beneficiated lunar soil, requiring the screening of the mineral anorthite from the lunar highlands as feedstock. This restricts the variety of formed fluorides, opening the possibility of complete fluorine recycling by stepwise chemical reaction with sodium, provided that reaction rates will be sufficiently high.

In the hydrogen reduction concept initial delivery and recycling of fluorine is based on the light-weight hydrogen fluoride, which implies relatively low masses for transportation and processing. Fluorination is proposed to be performed on untreated lunar fines, without beneficiation. Fluorine recycling from reactants (mixed metal fluorides) is proposed to be carried out in a reactive gas of atomic hydrogen, which should be capable of reducing most of the bulk mass of fluorides. Generation and recycling of fluorine by electrolysis of HF uses entirely conventional, proven technology (fluorine industry), which should be further improved with respect to mass- and energy-demand. Use of atomic hydrogen for reduction purposes, however, is only known from small scale processing of certain metal fluorides (Meinert 1979). Further experiments will be needed to demonstrate the feasibility of dissociating fluorides by using atomic hydrogen and/or by molten salt electrolysis.

Both concepts have some interesting aspects in common. As part of either process, purified silicon for solar cells, native aluminum and other

useful materials could be produced.

First estimates of the bulk energy required for production of oxygen have shown that by far the most energy demanding step seems to be the fluorine recovery which is dominated by electric supply of the electrolysis cell(s). Handling of fluorine and its compounds seems to pose no severe problems because certain materials can withstand the corrosion effects.

In order to verify the described theoretical concepts and to select favorable methods and conditions for fluorine recycling further work is needed. Especially, fluorination experiments with larger sample masses (grams to kilograms) should be performed to place the extrapolations of recent experimental results from laboratory scale to industrial processing dimensions on a firmer ground.

Acknowledgment. D. Stoeffler and J. Grothues (Institute for Planetology, University of Muenster/ FRG) kindly gave support with basic Moon-like material for preparation of lunar soil simulants, which where used for fluorination experiments at University of Bonn/FRG.

REFERENCES

Asprey, L. B. 1976. The preparation of very pure fluorine gas. *J. Fluorine Chem.* 7:359–361.

Baertschi, P., and Silverman, S. R. 1951. The determination of relative abundances of the oxygen isotopes in silicate rocks. *Geochim. Cosmochim. Acta* 1:317–328.

Boyle, R. V., Coombs, M. G., and Kudija, C. T. 1988. Solar dynamic power option for the space station. *Proc. of the 23rd Intersociety Energy Conversion Engineering Conf.*, vol. 3, ed. D. Y. Goswami (New York: ASME), pp. 319–328.

Burt, D. M. 1988. Lunar production of oxygen and metals using fluorine: Concepts involving fluorite, lithium, and acid-base theory. *Lunar Planet. Sci.* XIX:150-151 (abstract).

Burt, D. M. 1989. Lunar mining. *Amer. Sci.* 77:574–579.

Burt, D. M. 1992 Lunar mining of oxygen using fluorine. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 423–428.

Carrier, W. D., III, Mitchell, J. K., and Mahmood, A. 1973. The nature of lunar soil. *J. Soil Mech. Found. Div. (ASCE)* 99:813–832.

Corning Glass Works 1989. Chemical analyses of two synthetic lunar glasses. Unpublished data (Corning, N. Y.: Corning Glass Works).

Dalton, C., and Degelman, L. O. 1972. Design of a lunar colony. *NASA/ASEE Systems Design Institute* (Houston: NASA Grant NGT 44-005-114).

Dechema Corrosion Handbook Volume 5. 1989. Fluorine, hydrogen fluoride and hydrofluoric acid. Deutsche Gesellschaft für Apparatewesen, chem. Technik und Biotechnologie (Weinheim: Verlag Chemie), pp. 103–188.

Downs, W. R. 1972. Oxygen From Lunar or Similar Soil. U. S. Patent No. 3.773.913.

DuPont. 1989. Process for Preparing Fluorine by Electrolysis of CaF_2. U. S. Patent, US 4802970.

Ellis, J. F., and May, G. F. 1986. Modern fluorine generation. *J. Fluorine Chem.* 33:133–147.

Epstein, S., and Taylor, H. P., Jr. 1971. $^{18}O/^{16}O$, $^{30}Si/^{28}Si$, D/H and $^{13}C/^{12}C$ ratios in lunar samples. *Proc. Lunar Sci. Conf.* 2:1421–1441.

Epstein, S., and Taylor, H. P., Jr. 1975. Investigation of the carbon, hydrogen, oxygen, silicon isotope and concentration relationships on the grain surfaces of a variety of lunar soils and in some Apollo 15 and 16 core samples. *Proc. Lunar Sci. Conf.* 6:1771–1798.

Flucks, M., and Deutsch, A. 1990. Results from x-ray fluorescence analyses of five mineral separates. Unpublished data, University of Muenster/FRG.

General Electric Co. 1953. Hydrogen Reduction Method and Apparatus. U. S. Patent US 2768061.

Haimson, M., and Knauth, L. P. 1983. Stepwise fluorination—a useful approach for the isotopic analysis of hydrous minerals. *Geochim. Cosmochim. Acta* 47:1589–1595.

Handbook of Chemistry and Physics. 1989/90. 70th ed. (Cleveland: CRC), pp. D1–D93.

Hollemann, A. F., and Wiberg, E. 1976. *Lehrbuch der anorganischen Chemie,* 81th-90th ed. (New York: de Gruyter), pp. 62–64.

Jaccaud, M., Faron, R., Devilliers, D., and Romano, R. 1988. Fluorine. In *Ullman's Encyclopedia of Industrial Chemistry*, vol. A11, 5th ed. (Weinheim: Verlag Chemie), pp. 293–306.

Jarret, N., Das, S. K., and Haupin, W. E. 1980. Extraction of oxygen and metals and fluorine from lunar ores. In *Extraterrestrial Materials Processing and Construction,* ed. D. R. Criswell (Houston: Final Report to NASA/JSC, NSR 09-051-001), pp. 255–267.

Jaumann, R. 1988. Spektrophotometrische Analyse der chemisch-mineralogischen Zusammensetzung lunarer Oberflaechenmaterialien. Ph. D. Thesis, Univ. of Munich/FRG. German Aerospace Research Establishment Research Report.

Jones, W. E., Macknight, S. D., and Teng, L. 1973. Die Kinetik von Reaktionen des atomaren Wasserstoffs in der Gasphase. *Chem. Rev.* 73:465–486.

Keller, R. 1988. Lunar production of aluminum, silicon and oxygen. In *Metallurgical Processes for the Year 2000 and Beyond,* eds. H. Y. Sohn and E. S. Geskin (Warrendale, Pa.: Mineral, Metals and Materials Soc.), pp. 551–562.

Kesterke, D. G. 1970. Electrowinning oxygen from silicate rocks. In *Proc. of the Seventh Annual Working Group on Extraterrestrial Resources,* June 17–18, 1969, Denver, Co., NASA SP-229, pp. 139–144.

Kesterke, D. G., 1971. Electrowinning Oxygen from Silicate Rocks. U. S. Bureau of Mines Rept. of Investigations 7578.

Kilgore, C. C., Kraemer, S. R., and Bekkala, J. A. 1985. Fluorspar availability—Market economy countries and China. *U. S. Bureau of Mines Info. Circ.* 9060.

Kirk-Othmer. 1980. Fluorine. In *Encyclopedia of Chemical Technology,* vol. 10, 3rd ed. (New York: Wiley), pp. 630–654.

Meinert, H. 1979. *Fluorchemie—Zur Chemie und Anwendung*, vol. 126 (Berlin: Akademie-Verlag), p. 62.

Misra, A. K. 1988. Densities of some molten fluoride salt mixtures suitable for heat storage in space power applications. *J. Electrochemical Soc.* 135:2780–2781.

Myers, W. R., and DeLong, W. B. 1984. Fluorine corrosion: High-temperature attack on metals by fluorine and hydrogen fluoride. *Chemical Engineering Progress* 44(5):359.

O'Donnell, P. M. 1972. *Reactivity of Simulated Lunar Material with Fluorine,* NASA

TM-X-2533.

Pankratz, L. B., Stuve, J. M., and Gokcen, N. A. 1984. Thermodynamic data for mineral technology. *U. S. Bureau of Mines Bull.* 677.

Reed, G. R., Jr., Jovanovic, S., and Fuchs, L. H. 1970. Trace elements and accessory minerals in lunar samples. *Science* 167:501–503.

Seboldt, W., Lingner, S., Hoernes, S., and Grimmeisen, W. 1991. Oxygen extraction from lunar soil by fluorination. Resources of Near-Earth Space: Proc. Second Annual Symp. of the UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 11.

Semkow, K., and Sammells, A. F. 1987. The indirect electrochemical refining of lunar ores. *J. Electrochem. Soc.* 134A:2088–2089.

Shirai, H., Tanabe, H., Uematsu, T., Hanna, J., and Shimizu, I. 1989. Growth of amorphous, microcrystalline and epitaxial silicon at the same substrate temperature under control of atomic hydrogen. *J. Non-Crystalline Solids* 114(2):810–812.

Siemens A. G. 1981. Speichermaterial für Wasserstoff unter Verwendung von amorphem Silizium. German Patent DE 3110290.

Solvay Product Information. 1988. Fluor. 15-17.

Taylor, H. P., Jr., and Epstein, S. 1962. Relationship between $^{18}O/^{16}O$ ratios in coexisting minerals of igneous and metamorphic rocks. Part 1: Principles and experimental results. *Geol. Soc. America Bull.* 73:461–480.

Taylor, H. P., Jr. and Epstein, S. 1970. $^{18}O/^{16}O$ ratios of Apollo 11 lunar rocks and minerals. *Proc. Apollo 11 Lunar Sci. Conf.* 2:1613–1626.

Taylor, S. R. 1975. *Lunar Science: A Post Apollo View* (New York: Pergamon Press).

Taylor, S. R. 1982. *Planetary Science: A Lunar Perspective* (Houston: Lunar and Planetary Inst.).

Tsai, K. J., Kuchynka, D. J., and Sammells, A. F. 1990. The electrochemical generation of useful species from lunar materials. *J. Power Sources* 29:321–322.

Waldron, R. D. 1985. Total separation and refinement of lunar soils by the HF acid leach process. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSS Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 132–149.

Whittenberger, J. D., and Misra, A. K. 1987. Identification of salt-alloy combinations for thermal energy storage applications in advanced solar power systems. *J. Materials Eng.* 9:293–302.

PRODUCTION OF OXYGEN FROM LUNAR ILMENITE

Y. ZHAO and F. SHADMAN
University of Arizona

The kinetics and mechanism of reduction of ilmenite by carbon monoxide as well as hydrogen at 800 to 1100°C were investigated. The temporal profiles of conversion have a sigmoidal shape and indicate the presence of three different stages (induction, acceleration and deceleration) during CO reduction at all temperatures and H2 reduction at the temperatures below 876°C. The apparent activation energies based on the initial rates are 29.6 kcal/mole for CO reduction and 22.3 kcal/mole for H_2 reduction, respectively. The reaction is first order with respect to carbon monoxide and hydrogen under the experimental conditions studied. Both SEM and EDX analyses show that the diffusion of Fe product away from the reaction front and through the TiO_2 phase, followed by the nucleation and growth of a separate Fe phase are important steps in both reduction processes. The main difference between these two reactions is that TiO_2 can be reduced to lower oxides of titanium by hydrogen at temperatures higher than 876°C, and the reduction rate of ilmenite by H_2 is much faster than that of ilmenite by CO. A novel process flow sheet for carbothermal reduction process is also presented.

I. INTRODUCTION

The most useful material to produce on the Moon is oxygen for propellant (Cole and Segal 1964). There is a significant amount of oxygen on the Moon although very little is readily available as water or gaseous oxygen. The only practical source of oxygen on the lunar surface is igneous materials which contain typically 40 to 50% oxygen as oxides. The major minerals present in these rocks are ilmenite (the most abundant opaque mineral in lunar rocks), anorthite, and olivine. The igneous rocks have been pre-crushed by meteoritic bombardments to form regolith, the lunar equivalent of soil. This reduces the need for crushing and simplifies the mining and separation problem greatly.

Extraction of oxygen from iron oxide, as in the lunar ilmenite is of particular interest because it is energetically more favorable than extraction from silicon, aluminum, titanium, calcium or magnesium oxides. Iron oxide reduction is also attractive because of its potential for producing iron as a co-product. In addition, Agosto (1985) has concluded that ilmenite can be obtained from lunar soil at high purity using electrostatic separation techniques. A number of processes have been suggested for oxygen production from ilmenite in lunar regolith. Most of these processes require imported reagents to be recycled in the process. Therefore, low loss per pass is a critical factor and requirement for process feasibility. Among the proposed

processes, the hydrogen and carbothermal reduction of ilmenite appear very promising. Hydrogen reduction has a relatively simple process configuration; the individual steps are relatively well studied. However, the major problem is large heating and cooling loads required to condense the water and then heat the H_2 back up to its reaction temperature. Handling and storage of large amounts of hydrogen is also a problem.

Most of the available literature describes investigations on naturally occurring ores using carbon, CO or H_2 as reducing agents (Wouterlood 1979; Poggi et al. 1973; Bardi et al. 1987). From a fundamental point of view, the results of such studies are difficult to interpret because of the complex nature of the ores and the presence of many components. While CO might be an intermediate compound in any carbothermal reduction (Gupta et al. 1987; El-Guindy and Davenport 1970), the fundamental kinetics and mechanism of ilmenite reduction with CO are not clearly known. Poggi and Charette (1973) studied the reduction of synthetic ilmenite by carbon monoxide. However, their results are difficult to interpret because under the conditions used in their studies, carbon was also formed from CO disproportionation and participated in the ilmenite reduction. El-Guindy and Davenport (1970) investigated the reduction of synthetic ilmenite with graphite and discovered that the reduction starts at ~860°C at the contact points between the reactants. Up to 1020°C solid state reduction appears to be the main reaction mechanism, while above this temperature a rate increase has been observed and has been attributed to a change of mechanism to gaseous reduction of ilmenite by regenerated CO.

Briggs and Sacco (1988) studied the reduction of ilmenite by hydrogen at 600°C and 800°C. The ilmenite used in their study contained ~8% ferric iron. They found that some preoxidation of ilmenite by oxygen prior to reduction can decrease the complete reduction time of samples. During preoxidation, the ilmenite is converted to pseudobrookite (Fe_2TiO_5) and rutile. The single crystals of ilmenite are converted, therefore, into a polycrystalline array of pseudobrookite and a fine dispersion of rutile (Barnes and Pickles 1988). Carbotek (1988) has developed a fluidized-bed reactor for the reduction of ilmenite by hydrogen at the temperature between 900°C and 1000°C. Terrestrial ilmenite was used in this study. They demonstrated the feasibility of producing oxygen from terrestrial ilmenite by first reducing the terrestrial ilmenite and then electrolyzing water to produce oxygen and hydrogen which is returned to the reactor. They also reported that the reaction is first order with respect to hydrogen under their experimental conditions. Bardi et al. (1987) investigated the kinetics of hydrogen reduction of Norwegian ilmenite ore powders. The electron microprobe analysis of the reduced Norwegian ilmenite grains showed the existence of a segregated iron phase present as spheroidal nodules and a TiO_2 phase present as vein-like arrangements. Their optical microscopic study on sections of synthetically prepared $FeTiO_3$ showed separate reacted and unreacted zones in the sample particles.

Donnelly (1970) reduced ilmenite beach sands using a mixture of hydrogen and carbon monoxide and found that diluting the hydrogen reducing gas

with CO decreases the rate of reduction of the iron oxide. The lowering of the rate of reduction is attributed to the water gas shift reaction. In view of the presence of both CO and H_2 in ilmenite reduction process based on the use of carbonaceous waste as a carbon source, the kinetics and the mechanism of synthetic ilmenite reduction by H_2/CO is currently under study in our laboratory.

In this chapter, the principal aim is to describe the reduction of ilmenite by carbon monoxide and hydrogen under conditions where the original ilmenite and the final products are well characterized. The emphasis is on ilmenite with no ferric impurities present. This is important in applying the results to the reduction of lunar ilmenite. Section II describes the experimental setup. The kinetics and mechanism of ilmenite reduction by carbon monoxide and hydrogen are considered in Secs. III and IV, respectively. Section V deals with the comparison of CO and H_2 reduction of ilmenite and a new process flowsheet for a novel carbothermal reduction process of ilmenite. Finally, various conclusions are given in Sec. VI.

II. EXPERIMENTAL APPROACH

A schematic of the experimental apparatus is shown in Fig. 1. The main components of this system are an electronic microbalance (Cahn Instruments, Inc., Model 1000), a quartz flow through reactor with inlet and outlet, and a movable furnace with a PID controller. The composition of gaseous reactants and products was determined using an infrared analyzer and a gas chromatograph. Ilmenite was used in the form of thin flakes pressed from powder. Samples were suspended from the microbalance, which monitored weight changes during the course of an experiment. A thermocouple was used to monitor the temperature of the reactor around the flake. All experiments were performed under isothermal conditions. For the CO reduction, the reducing gas entering the reactor contained CO, CO_2 and N_2. The ratio of CO to CO_2 was always maintained at 99 to prevent carbon deposition due to the CO disproportionation reaction (Jones 1975; Shomate 1946). The gas flow was varied from 105 to 260 std.cc/min. For H_2 reduction, the reducing gas contained H_2 and N_2. The gas flow ranged from 260 to 660 std.cc/min.

Samples of starting material were prepared by cold pressing approximately 0.270 g of $FeTiO_3$ powder (with particles size $<45\,\mu$m) in a die at 14,500 psi for 5 minutes to form disks. The disks were then cut into flakes approximately 10 mm by 8 mm. The thickness of the disks was 0.60 mm except for experiments conducted to determine the effect of intergranular diffusion resistance (transport through void space among grain particles).

Each experiment was started by first purging the reactor system at room temperature to reduce the concentration of oxygen to levels below 25 ppm. The reducing gas was then introduced into the reactor. To initiate the reduction, the furnace was raised rapidly. Within three minutes the temperature in

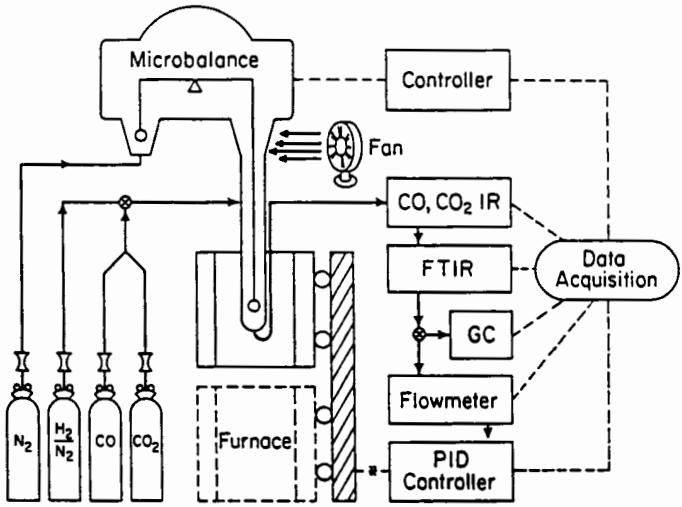


Figure 1. Schematic diagram of the reactor system. GC=gas chromatograph, IR=non-dispersive infrared analyzer, FTIR=Fourier transform infrared spectrometer.

the reactor was within 1% of the set point temperature. The experiments were terminated at a desired conversion by rapidly lowering the furnace.

Several techniques were used for chemical analysis and characterization of the starting and reduced samples. Mossbauer spectroscopy was employed to determine the oxidation state of iron in our synthetic ilmenite. X-ray diffraction (XRD) was used to identify the different crystalline phases in the starting material and products. High-resolution scanning electron microscopy (SEM) and energy-dispersive X-ray (EDX) analyses were employed to examine the polished cross section of both partially and completely reduced samples and to determine the elements present in each phase. For SEM and EDX analyses, the samples were mounted in an epoxy resin and polished to expose the cross section of the grains.

III. REDUCTION OF ILEMITE WITH CARBON MONOXIDE

The experimental results of CO reduction of ilmenite are presented in this section. The impurity content of our synthetic ilmenite used in this study is given in Table I. The XRD pattern of our synthetic ilmenite, as shown in Fig. 2a, suggests that the sample contains only the $FeTiO_3$ crystalline phase. A typical two-line Mossbauer spectrum of the sample, shown in Fig. 3b, confirms that the iron in the synthetic ilmenite is in the form of Fe^{++}. This

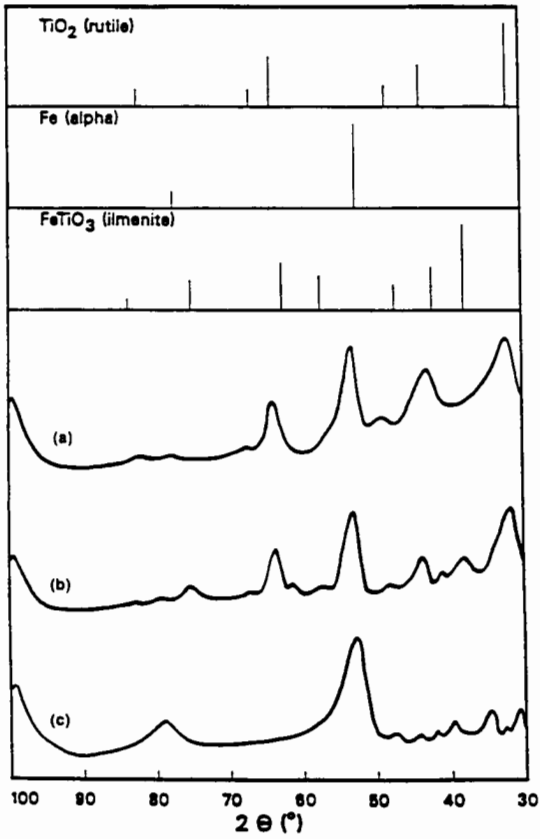


Figure 2. (a) X-ray diffraction spectrum of synthetic ilmenite. (b) X-ray diffraction spectrum of synthetic ilmenite after complete reduction. $T = 900, 1000, 1100°C$. (c) X-ray diffraction spectrum of synthetic ilmenite after partial reduction. $T = 1000°C$.

represents the oxidation state of iron in lunar ilmenite. The six-line Mossbauer spectrum of hematite or Fe_2O_3 is given in Fig. 3a.

For each experiment, the total sample weight loss was obtained based on the continuous microbalance measurement and the total amount of CO_2 was given by the on-line infrared gas analyzer, which is capable of continuously monitoring the concentration of CO_2 in the gas phase. It is found that these two measurements agree well with following stoichiometry:

$$FeTiO_3(s) + CO(g) = CO_2(g) + Fe(s) + TiO_2(s). \quad (1)$$

The first set of experiments were conducted to determine the effect of interphase transport resistance around the flakes. The experiments were conducted at the highest temperature (1100°C) with 13% CO. As shown in Fig. 4, the interphase resistance is not important if the flow rate is at least 260

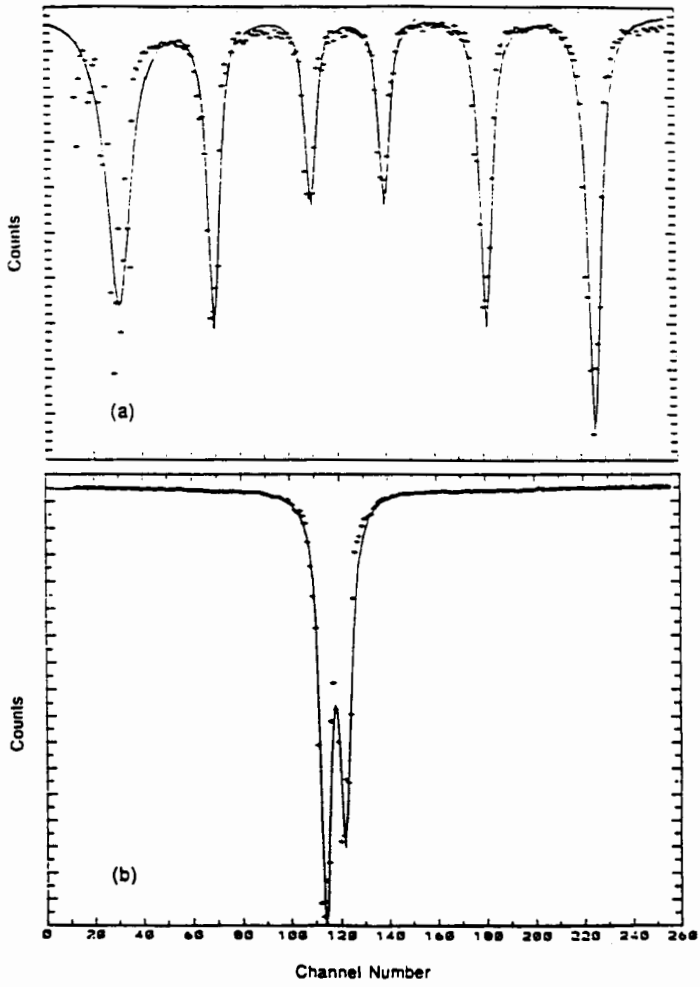


Figure 3. (a) Mossbauer velocity spectrum for hematite. (b) Mossbauer velocity spectrum for synthetic ilmenite.

std.cc/min. At temperatures below 1100°C, the interphase resistance will be even less significant.

The second set of experiments were to determine the effects of intergranular diffusion of CO on the kinetics. These experiments were conducted at the highest temperature (1100°C) with 23% CO and at a gas flow rate of 260 std.cc/min. As shown in Fig. 5, the intergranular diffusion of CO does not affect the reduction kinetics if the flake thickness is less than 0.60 mm.

To find the reaction order, a series of experiments were conducted at 13% and 23% CO concentration at 900, 1000 and 1100°C. The results, shown in Fig. 6, show that the reaction order with respect to carbon monoxide is unity

TABLE I
Maximum Impurity Content in the
As-Received Synthetic Ilmenite

Impurity	Maximum Concentration (Wt. %)
Al	0.001
Ca	0.01
Cr	0.001
Cu	0.01
Mg	0.001
Si	0.1

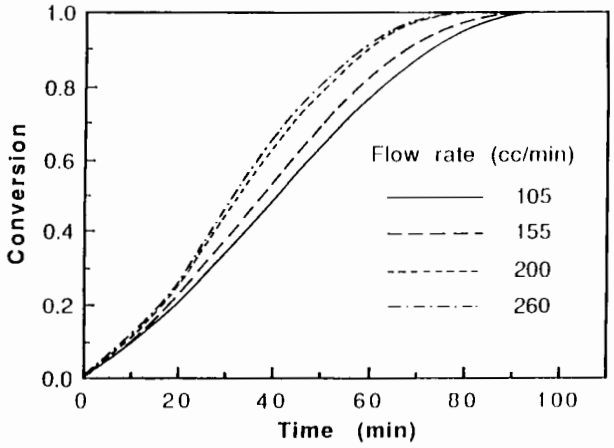


Figure 4. The effect of gas flow rate on the reduction rate of ilmenite. [CO]=13%, $T = 1100°$.

in this range of temperature and CO concentration. Therefore, the rate of reaction can be written as

$$r = k(T, X)C_{CO} \tag{2}$$

where rate is expressed in [mg reacted/original mg/min] and C_{CO} is in gmole/liter.

The effect of temperature on the reaction rate at various conversions is shown in Fig. 7. The apparent activation energy was 18, 14 and 10 kcal/gmole at 10%, 30% and 50% conversion level, respectively. As conversion increases, the thickness of the TiO_2 product layer in grain particles increases. This increases the diffusional resistance against the CO transport into the grain particles. The higher diffusional resistance causes a decrease in the apparent activation energy.

Isothermal weight loss measurements were performed at 900, 1000 and 1100°C. The temporal profiles of conversion at these three temperatures and

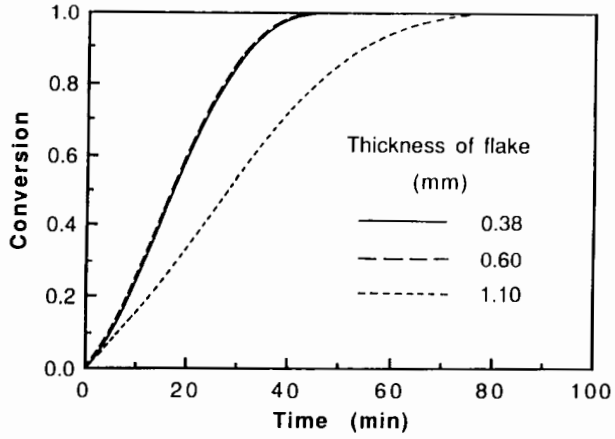


Figure 5. The effect of flake thickness on the reduction rate of ilmenite. [CO]=23%, $T = 1100°C$.

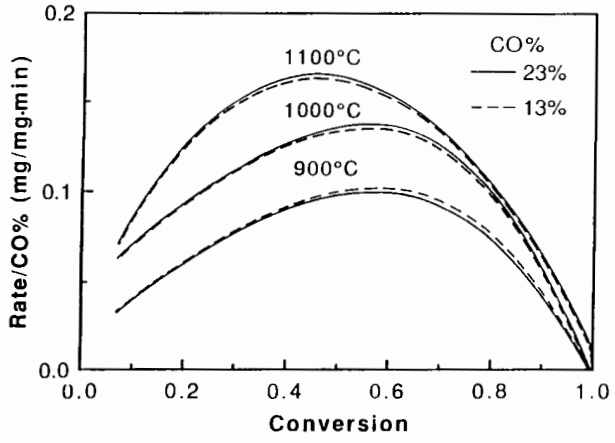


Figure 6. The effect of CO concentration on the reduction rate of ilmenite.

CO concentration of 23% are shown in Fig. 8. These profiles have a sigmoidal shape and indicate the presence of three different stages (induction, acceleration and deceleration) during the reduction reaction.

To gain insight into the mechanism of the ilmenite reduction, particularly in relation to the observed three stages, samples of both completely and partially reduced ilmenite were analyzed using various analytical techniques. In particular, a combination of optical microscopy, SEM, EDX and XRD analyses provided very useful information on the nature and distribution of various phases including the products Fe and TiO_2.

A SEM secondary electron micrograph of the polished cross section of an ilmenite flake after partial reduction at 1000°C is shown in Fig. 9a. The

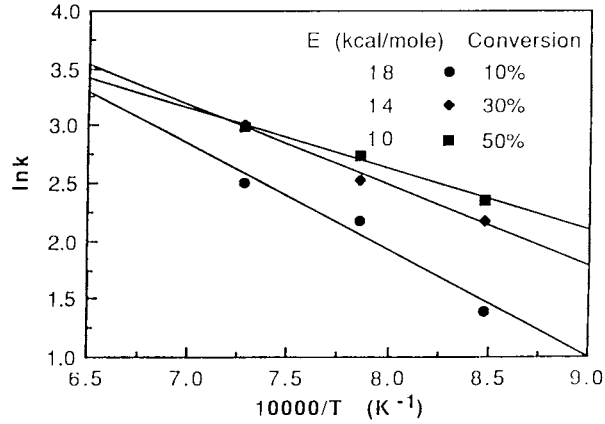


Figure 7. Values of apparent activation energy at 10, 30, and 50% conversion.

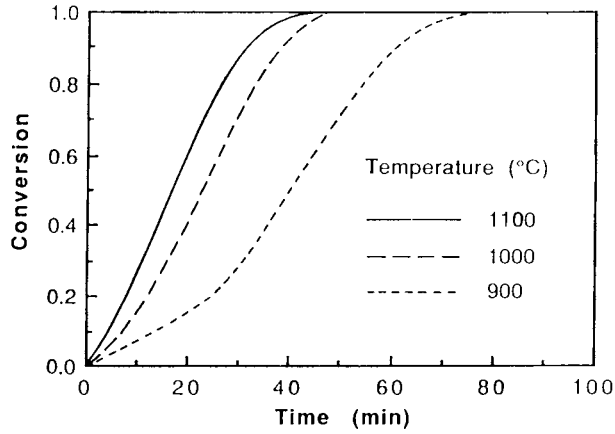


Figure 8. The effect of temperature on the reduction rate of ilmenite. CO%=23.

micrograph reveals three distinct regions which appear as bright, light gray and dark gray phases. In order to identify the phases present, quantitative EDX was performed at spots marked in Fig. 9a. The observations show that the bright phase is primarily iron, the dark gray phase is made up of titanium dioxide and the light gray phase is unreacted $FeTiO_3$. These results suggest that there is a strong tendency towards the segregation of the products iron and titanium dioxide and that iron diffuses to the grain boundaries through the TiO_2 layer during the reduction. This finding has important implications for product separation for recovery of Fe and TiO_2. Because the reduction temperatures are much lower than the melting point of TiO_2, it is expected that the TiO_2 product exists in polycrystal form. As shown in Fig. 9a, it appears that the reaction in the grain particles proceeds according to the shrinking

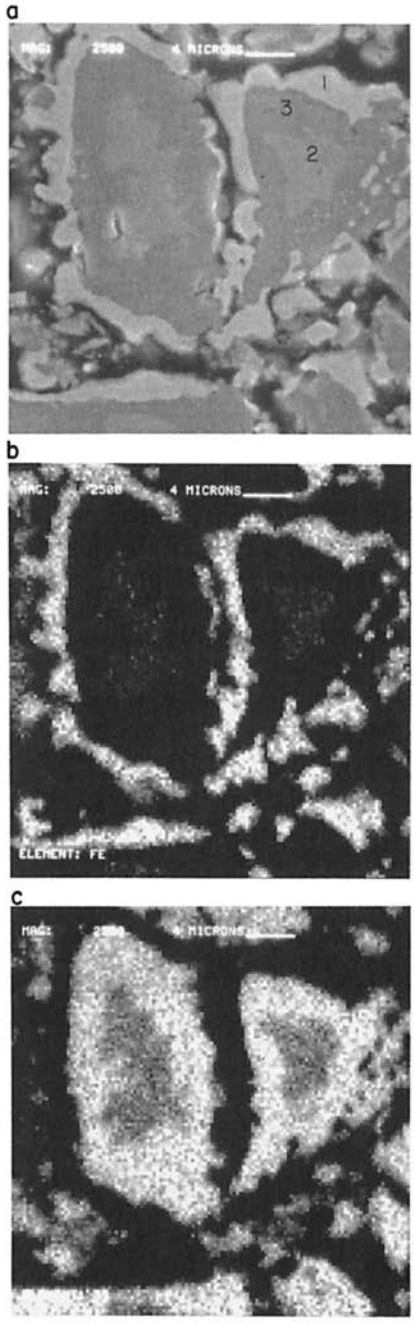


Figure 9. (a) SEM secondary electron micrograph of the polished cross section of ilmenite flake after partial reduction. $T = 1000°C$; magnification=2500X; point 1: 3.1 atom% Ti, 96.9 atom% Fe; point 2: 51.6 atom% Ti, 48.4 atom% Fe; point 3: 98.9 atom% Ti, 1.1 atom% Fe. (b) Fe Kα X-ray map of the cross section shown in 9a. (c) Ti Kα X-ray map of the cross section shown in 9a.

unreacted core model. This result is expected because the grain particles of synthetic ilmenite are nearly nonporous, whereas the product TiO_2 is porous. The corresponding X-ray Kα map of iron and titanium, as shown in Figs. 9b and 9c, confirm the shrinking core configuration.

The polished cross-sections of synthetic ilmenite flakes after complete reduction at 900 and 1100°C were also examined by SEM and EDX. The results obtained, as shown in Figs. 10 and 11, indicate that a similar mechanism of reduction is involved. However, the coalescence of grains is observed in the flakes reduced at 1100°C, which is apparently due to the sintering of iron. The results of EDX analyses on these samples indicate that, the phase enriched in titanium is depleted in iron and vice versa. This confirms the high degree of segregation of product Fe and TiO_2.

The XRD spectra of both partially and completely reduced samples are shown in Fig. 2b and 2c. The phases present after complete reduction at 900, 1000, and 1100°C were iron and titanium dioxide and those present after partial reduction at 1000°C were iron, titanium dioxide and unreacted ilmenite. These findings confirm the data obtained from EDX analysis.

The various observations described here all point to a mechanism consisting of the following main steps for the reaction in each grain:

1. Diffusion of CO through the porous product layer of TiO_2 towards the unreacted core of grain particles;
2. Reaction of CO with the ilmenite core to produce TiO_2 and Fe;
3. Migration of Fe through the TiO_2 layer away from the unreacted core towards the grain boundary;
4. Formation of iron nuclei and their subsequent growth outside and around the reacted grain particles.

Steps 3 and 4 result in almost complete segregation of the two solid products iron and titanium dioxide in the scale of grains.

Using this proposed mechanism, the three stages observed during conversion can be described as follows:

a. Induction stage: this represents the initial stage of the reduction process and corresponds to the formation of iron nuclei. At this stage, most of iron reduced from $FeTiO_3$ is in the matrix of TiO_2. The duration of this period is temperature sensitive, decreasing from 15 minutes at 900°C to five minutes at 1100°C.
b. Acceleration stage: in this stage, the iron nuclei formed during the induction stage as well as those that may form subsequently grow. As more Fe diffuses out and pores in TiO_2 layer open up; this allows more CO to diffuse into the matrix of TiO_2 and react with $FeTiO_3$. The rate of reduction increases and reaches a maximum. Iron cannot be a catalyst leading to rapid rate during this stage of the reaction. If iron does catalyze the reaction, then the induction stage would not exist and the maximum rate would be observed at a very low conversion.

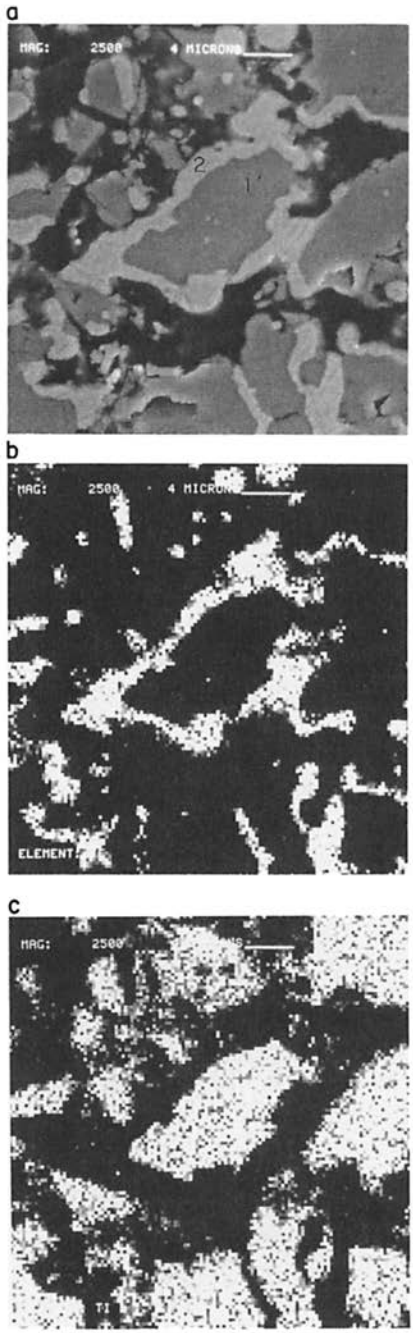


Figure 10. (a) SEM secondary electron micrograph of the polished cross section of ilmenite flake after complete reduction. $T = 900°C$, magnification=2500X. Point 1: 99.0 atom% Ti, 1.0 atom% Fe; point 2: 5.4 atom% Ti, 94.6 atom% Fe (b) Fe Kα X-ray map of the cross section shown in 10a (c) Ti Kα X-ray map of the cross section shown in 10a.

c. Deceleration stage: depletion of $FeTiO_3$ results in a decrease in the rate of reduction.

For a direct observation of the effect of iron nucleation on the rate, in a series of experiments, a known amount of iron powder with particle size $<45\ \mu m$ was added to ilmenite powder before pressing. As shown in Fig. 12, the addition of iron significantly reduced the length of induction period. This confirms the suggestion that the low rate during the initial induction period is due to the absence of sufficient iron nuclei. This causes slow transport of iron away from the reaction front and inhibition of CO contact with the unreacted $FeTiO_3$.

It appears that only Poggi et al. (1973) has examined the CO reduction of ilmenite. The synthetic ilmenite sample in their study is in the form of fused briquette with the porosity of 1%, whereas the sample in our study has porosity of 30%. In addition, the CO disproportionation was not considered in their work. These make it difficult to compare the results. They found the activation energy to be 14.1 kcal per mole in the temperature range 900 to 1100°C.

IV. REDUCTION OF ILMENITE WITH HYDROGEN

A. Experimental Results

Initially, some experiments were conducted to determine the effect of transport resistance in the interphase around the flakes. The experiments were conducted at the highest temperature (1014°C) with 3.4% H_2. The results showed that the interphase resistance is not important if the flow rate is at least 660 std.cc/min. At temperatures below 1014°C, the interphase resistance is even less significant.

The results in Fig. 13 show the effect of H_2 concentration on the reaction at 945°C. The H_2 concentration was varied between 3.4% and 14.7% in N_2 atmosphere. As expected, an increase in the hydrogen concentration results in an increase in the rate and a decrease in the time required to attain certain fractional weight loss, which is defined as the ratio of weight loss of the sample to initial weight of the sample. The reaction order with respect to H_2 is established using the runs which were not influenced by diffusion in the ilmenite flake. The results, shown in Fig. 14, indicate that the reaction order is unity in the H_2 concentration range of 3.4% to 14.7% at 807°C and 876°C. This confirms the first order kinetics assumption presented in the theoretical section.

Isothermal weight loss measurements were performed at 807, 876, 945 and 1014°C. The temporal profiles of conversion at these four temperatures and H_2 concentration of 3.4% are shown in Fig. 15. The profile at 807°C has a sigmoidal shape and indicates the presence of three different stages (induction, acceleration and deceleration) during the reduction reaction. The profiles at 876°C, 945°C and 1014°C do not have these three stages. As

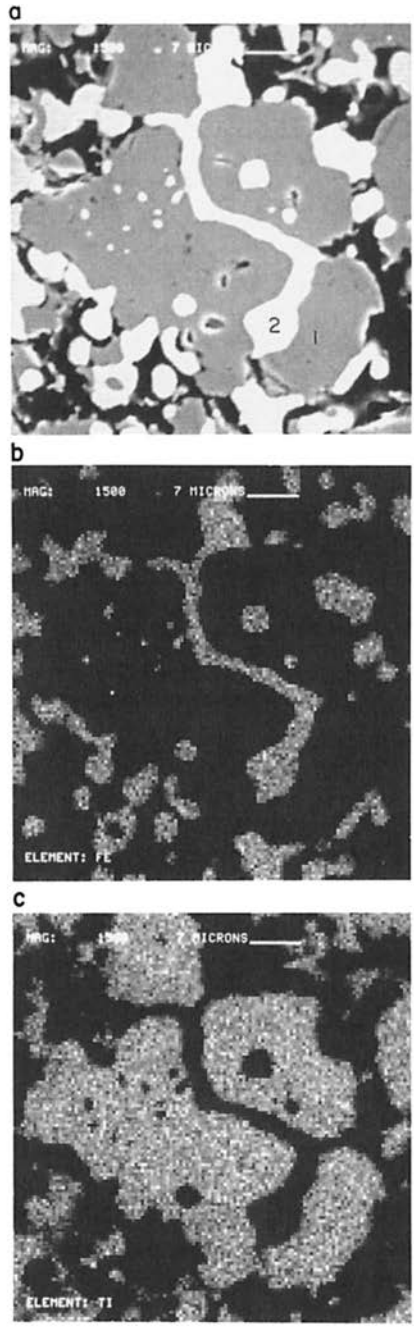


Figure 11. (a) SEM secondary electron micrograph of the polished cross section of ilmenite flake after complete reduction. $T = 1100°C$, magnification=2500X. Point 1: 99.0 atom% Ti, 1.0 atom% Fe; point 2: 2.5 atom% Ti, 97.5% Fe (b) Fe Kα X-ray map of the cross section shown in 11a. (c) Ti Kα X-ray map of the cross section shown in 11a.

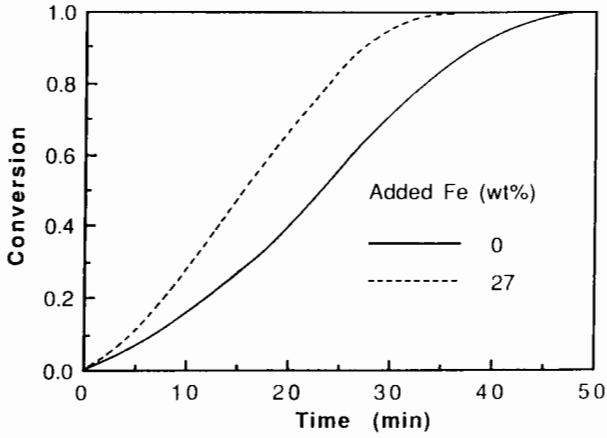


Figure 12. The effect of iron addition on the reduction rate of ilmenite. CO%=23, $T = 1000°C$.

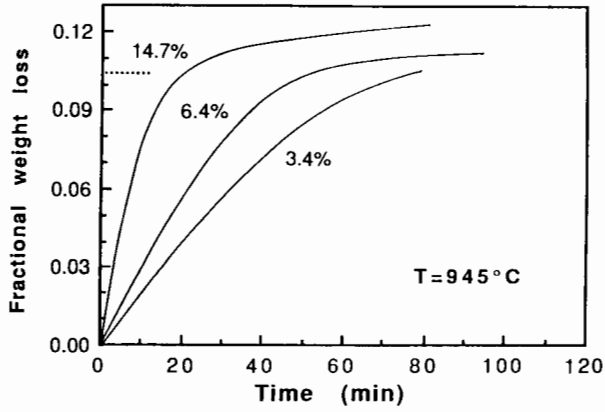


Figure 13. Effect of hydrogen concentration on the reduction of ilmenite; — complete iron metallization.

shown in Fig. 15, the time required to attain the weight loss corresponding to complete iron metallization when all the iron in ilmenite is reduced to metallic iron, is 210 minutes at 807°C and 52 minutes 1014°C with 3.4% H_2. The effect of temperature on the reaction rate is shown in Fig. 16. The apparent activation energy calculated based on initial rates is 22.3 kcal/mole.

SEM backscattered electron micrographs of the polished cross section of ilmenite flake after partial reduction at 1014°C and 807°C are shown in Figs. 17a and 18a. The micrographs reveal three distinct regions which appear as bright, light gray and dark gray phases. In order to identify these phases, quantitative EDX was performed at spots marked in Figs. 17a and 18a. These results and the XRD observations (to be discussed later) show that the bright

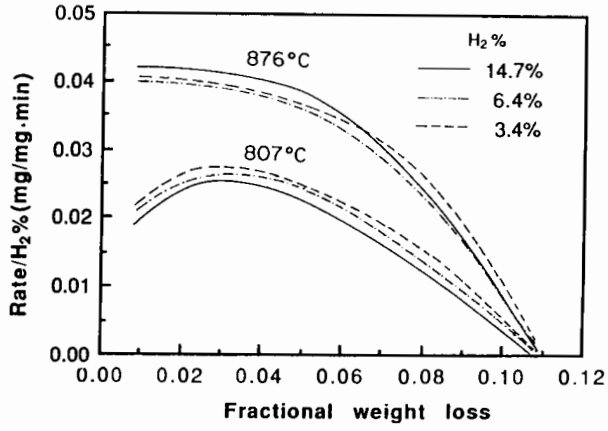


Figure 14. Effect of hydrogen concentration on the reduction rate of ilmenite.

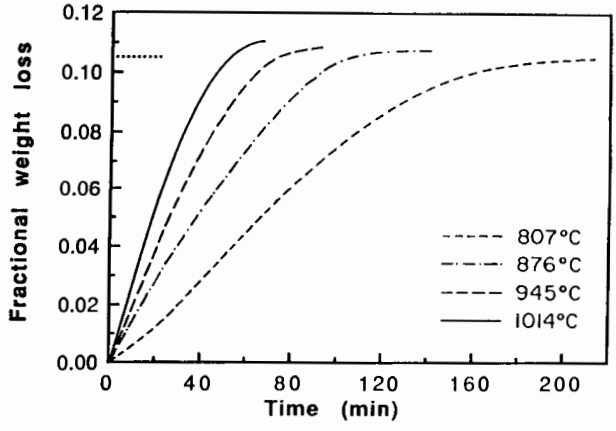


Figure 15. Effect of temperature on the reduction rate of ilmenite: H_2=2.4%; —, complete iron metallization.

phase is primarily iron; the dark gray phase is titanium dioxide and the light gray phase is unreacted $FeTiO_3$.

Figure 17a and 18b indicate that the reaction in the grain particles proceeds according to the shrinking core model. The corresponding X-ray map of iron and titanium, shown in Figs. 17b, 17c, 18b and 18c, confirm the shrinking core configuration, which is very similar to CO reduction.

The results in Sec. III showed that the reduction of TiO_2 by CO did not take place at any appreciable rate in the temperature range between 800°C to 1100°C. However, the experimental results for the ilmenite reduction by hydrogen in Fig. 13 show that the total weight loss of the sample exceeds the weight loss corresponding to the complete iron metallization. This indicates

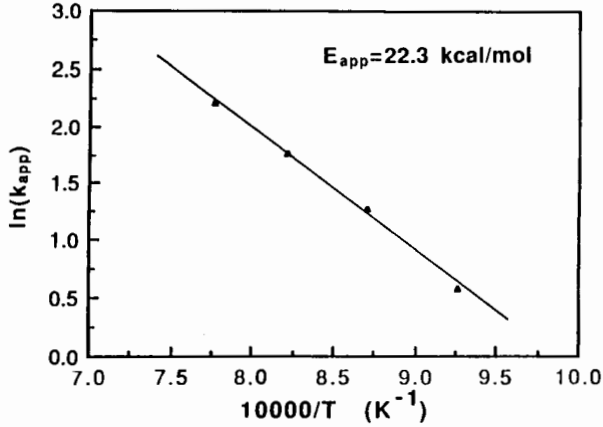


Figure 16. Temperature dependence of the apparent rate coefficient.

that titanium dioxide can be reduced by hydrogen in the temperature range of 876°C and 1014°C. The rate of reduction of titanium dioxide depends on both the hydrogen concentration and the reaction time. The reduction of titanium dioxide has practical significance in the production of oxygen from lunar ilmenite, because 67% of oxygen in ilmenite is bound to titanium.

In order to understand the mechanism of titania reduction by hydrogen, the electron microprobe with wavelength dispersive X-ray analysis was employed to determine the atomic ratio of oxygen to titanium in the reduced titanium dioxide phase. The analyzed area of the grain in the sample reduced at 1014°C is shown in Fig. 19. For each sample, the analyses for titanium and oxygen were performed across the two polished grains. One is located at the edge of the flake; the other is located at the center. The results obtained from these two grains are very similar. The results showing the extent of reduction of TiO_2 are given by Fig. 20 and indicate that the reduction of titania took place throughout the titanium dioxide matrix in each grain. These observations suggest that the reduction of titanium dioxide in each grain and across the flake is kinetically controlled.

Another important point related to the reduction of TiO_2 is its inception relative to iron reduction. To further study this, two partially reduced samples were prepared at 1014°C and 14.7% H_2, the first one at 35% conversion and the second one at 70% conversion. The results of WDX analysis showed that the TiO_2 phases in both samples had not been reduced at either conversions. These results indicate that the reduction of TiO_2 does not occur to any significant extent as long as iron metallization is not completed. It is speculated that this is related to inhibition effect of water vapor as the reaction product. The equilibrium constants of TiO_2 reduction by hydrogen in the temperature range of interest are much smaller than those of ilmenite reduction. During ilmenite reduction (iron metallization), the concentration of water vapor inside the

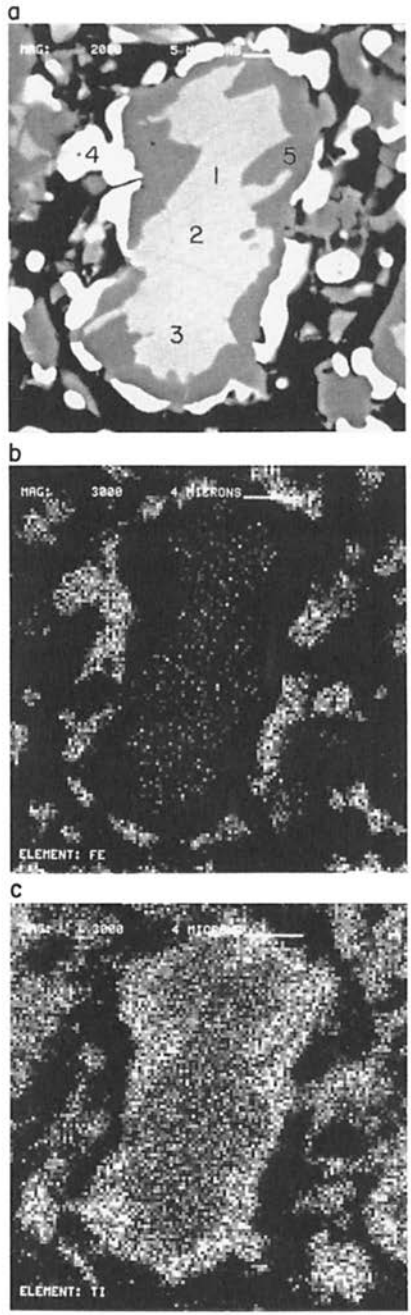


Figure 17. (a) Backscattered electron micrograph of the polished cross section of ilmenite flake after partial reduction. $T = 1014°C$, magnification=2000X. Point concentrations in atom% are as follows: Point 1: Ti 50; Fe 50. Point 2: Ti 51; Fe 49. Point 3: Ti 51; Fe 49. Point 4: Ti 2; Fe 98. Point 5: Ti 96; Fe 4. (b) Fe Kα X-ray map of the cross section shown in 17a. (c) Ti Kα X-ray map of the cross section shown in 17a.

TiO_2 pores is high enough to inhibit the TiO_2 reduction. However, as iron metallization approaches completion, the inhibition effect decreases and TiO_2 reduction starts.

The XRD spectra of partially reduced samples are shown in Fig. 21b. The phases present after partial reduction at 807°C and 1014°C are iron, titanium dioxide and unreacted ilmenite. Figure 21a is the XRD spectrum of sample reduced at 807°C with "complete iron metallization," and indicates the presence of Fe and TiO_2. The samples reduced at 876°C, 945°C and 1014°C with 13% fractional weight loss were also analyzed using XRD. The patterns, shown in Fig. 21c, indicate that all the peaks of TiO_2 phases disappeared, which confirms the reduction of TiO_2 by H_2.

The sigmoidal profiles of conversion and the three stages observed during ilmenite reduction are similar to what is observed in the CO reduction in the Sec. III. However, the induction stage in the H_2 reduction of ilmenite is less significant than that in CO reduction. This is because hydrogen diffusivity in the pores is significantly larger than CO diffusivity. Therefore, the effect of pore blockage on H_2 reduction is less than that on the CO reduction.

B. Theoretical Modelling

A mathematical model is formulated to describe the simultaneous reaction and diffusion that occur in the ilmenite flakes used in the experimental part of this study. The flakes were uniform and thin; therefore, the geometry assumed for mathematical modeling is that of an infinitely long and wide slab. This configuration was selected because it gives the desired information on the kinetics of ilmenite reduction without complications of sample shape. Thin flakes are also suitable for polishing as needed in the microprobe studies. It is assumed that the flakes consist of spherical and equal-sized grain particles of ilmenite. Based on the SEM micrographs and image analysis of grains in a flake, the overall size of a flake and the size of an individual grain do not change significantly during the reaction.

The reactions considered are as follows:

$$FeTiO_3(s) + H_2(g) = H_2O(g) + TiO_2(s) \tag{3}$$

or

$$FeTiO_3(s) + CO(g) = CO_2(g) + Fe(s) + TiO_2(s) \tag{4}$$

The following additional assumptions are made:

1. The pseudosteady state approximation is appropriate for describing the concentration of the gaseous species within a flake;
2. The system is isothermal;
3. The effective diffusivities of gaseous reactant and product are equal and uniform throughout the flake;
4. The reaction is first order with respect to H_2, H_2O, CO, and CO_2;
5. The grain particles have little porosity and react following a shrinking-core mechanism.

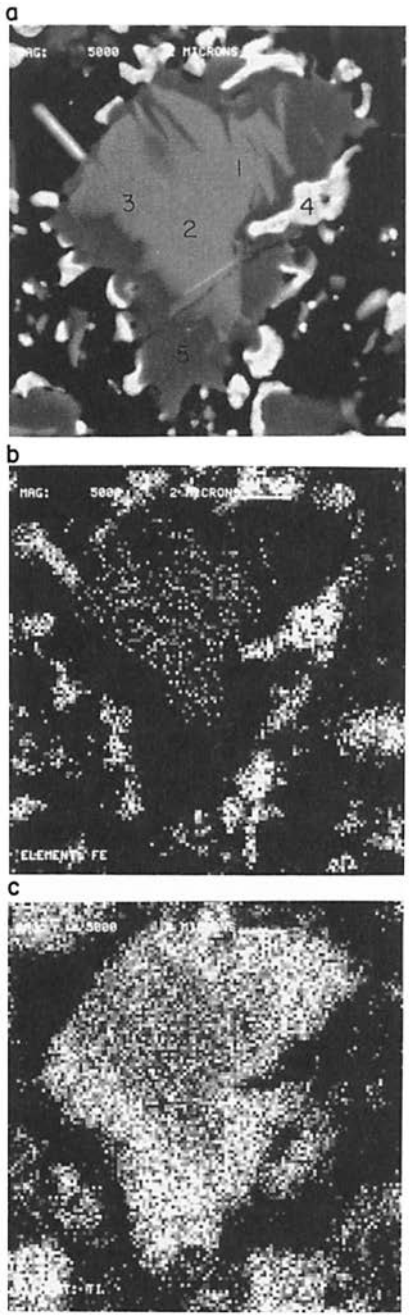


Figure 18. (a) Backscattered electron micrograph of the polished cross section of ilmenite flake after partial reduction. $T = 807°C$, magnification=5000X. Point concentrations in atom% are as follows Point 1: Ti 50; Fe 50. Point 2: Ti 50; Fe 50. Point 3: Ti 51; Fe 49. Point 4: Ti 5; Fe 95. Point 5: Ti 99; Fe 1. (b) Fe Kα X-ray map of the cross section shown in 18a. (c) Ti Kα X-ray map of the cross section shown in 18a.

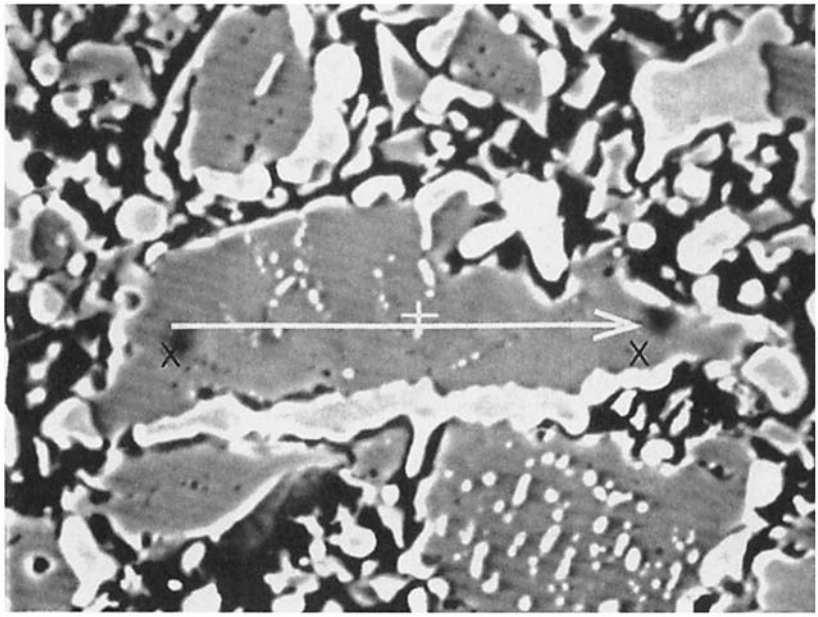


Figure 19. Backscattered electron micrograph of a reduced TiO_2 phase.

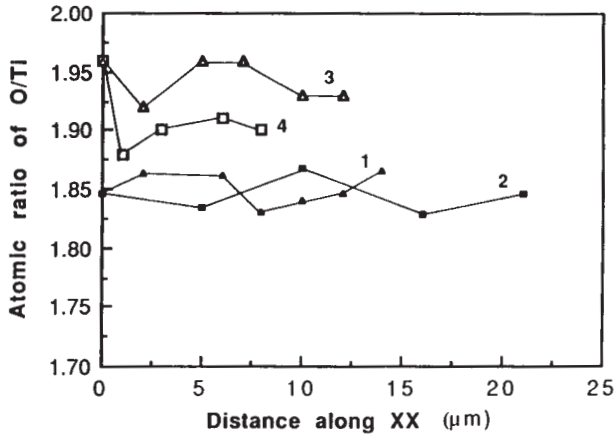


Figure 20. Variation in the oxygen to titanium atomic ratio in flakes. 1: sample 1, edge (Fig. 11), $T = 1014°C$; 2: sample 1, center $T = 1014°C$; 3: sample 2, edge, $T = 876°C$; 4: sample 2, center, $T = 876°C$.

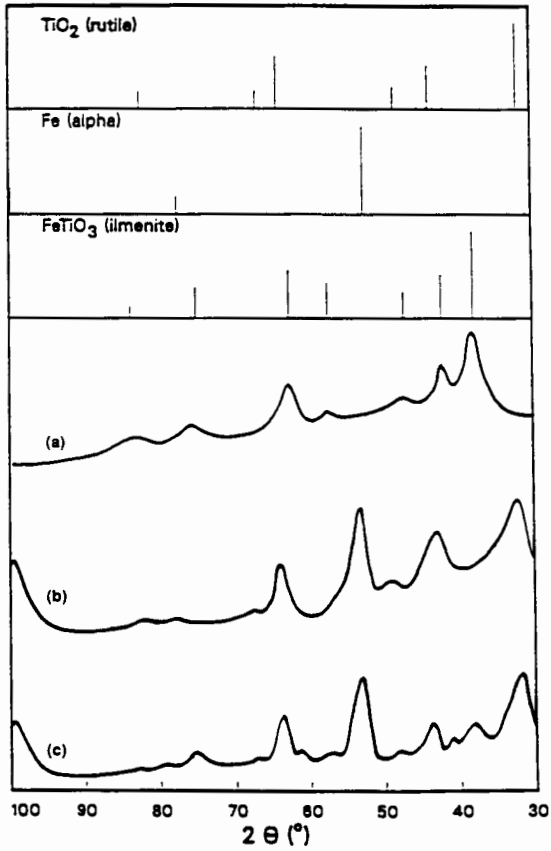


Figure 21. X-ray diffraction spectrum of: (a) ilmenite, complete iron metallization at 807°C; (b) ilmenite, partially reduced at 807°C or 1014°C; (c) ilmenite, complete iron metallization and partial reduction of TiO_2 at 876°C, 945°C and 1014°C.

The conservation equations for hydrogen or CO, A and water or CO_2, C, can be written as follows:

$$D_e \nabla^2 C_A - R_A = 0 \tag{5}$$

$$D_e \nabla^2 C_c + R_A = 0. \tag{6}$$

The local rate of reaction based on the grain particles is given by the standard shrinking core model:

$$-\rho_s \frac{dr_c}{dt} = k \frac{C_A - C_c/K}{1 + \frac{k(1+1/K)}{D_{eg}} \left(1 - \frac{r_c}{r_s}\right) r_c}. \tag{7}$$

Dividing Eq. (6) by K and subtracting from Eq. (5) gives:

$$D_e \nabla^2 (C_A - C_c/K) - R_A(1 + 1/K) = 0. \tag{8}$$

The initial and the boundary conditions for (7) and (8) are: at $t = 0, r_c = r_s$; at $Z = L, C_A - C_c/K = C_{Ab} - C_{cb}/K$; at $Z = 0, \frac{d(C_A - C_c/K)}{dz} = 0$.

An expression can be obtained for the local rate of reaction R_A. For a flake in this study, R_A is given by:

$$R_A = 3\frac{\frac{r_c^2}{r_s^3}(1-\epsilon)k(C_A - C_c/K)}{1 + \frac{k(1+1/K)}{D_{eg}} r_c\left(1 - \frac{r_c}{r_s}\right)}. \tag{9}$$

The local conversion can be related to the unreacted grain radius as follows:

$$X_B = 1 - \left(\frac{r_c}{r_s}\right)^3. \tag{10}$$

The overall conversion for a flake is given by:

$$X = \left(\frac{1}{L}\right)\int_0^L X_B \mathrm{d}Z. \tag{11}$$

It has been shown by Sohn (1974) that the solution to the model can be described by the following relationship:

Time required to attain a certain conversion $\approx$	Time required to attain the same conversion in the absence of resistance due to intrapellet diffusion	+Time required to attain the same conversion under the control of intrapellet diffusion

$$t^* \approx f(X) + \sigma_g^2 \cdot g(X) + \sigma^2 \cdot p(X) \tag{12}$$

$$t^* = \frac{k}{\rho_s r_s}(C_{AB} - C_{Cb}/K)t \tag{13}$$

$$\sigma_g^2 = \frac{k r_s}{6 D_{eg}}(1 + 1/K) \tag{14}$$

$$\sigma^2 = L^2 \frac{(1-\epsilon)k}{2 D_e r_s}(1 + 1/K) \tag{15}$$

$$f(X) = 1 - (1 - X)^{1/3} \tag{16}$$

$$g(X) = 1 - 3(1 - X)^{2/3} + 2(1 - X) \tag{17}$$

$$p(x) = X^2. \tag{18}$$

The terms on the right side of Eq. (12) are the asymptotic expressions which can be related to conversion as shown in Eqs. (16)–(18). The results

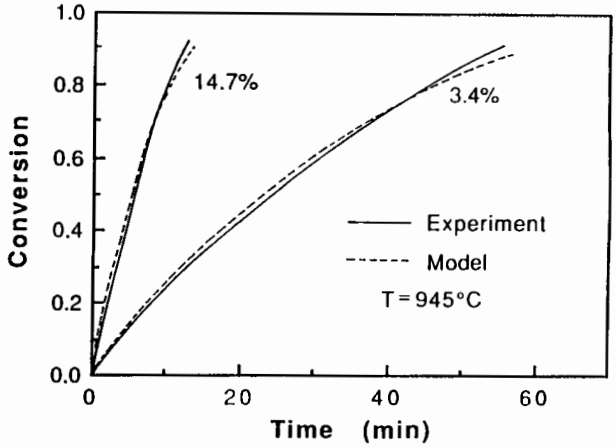


Figure 22. Comparison of experimental data with model predictions: H_2=3.4% and 14.7%.

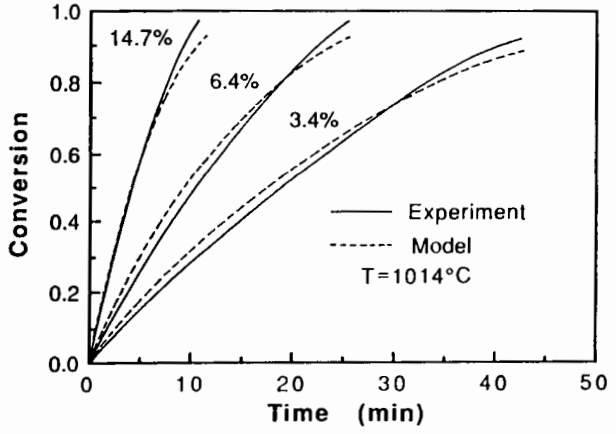


Figure 23. Comparison of experimental data with model predictions: H_2=3.4%, 6.4% and 14.7%.

obtained by this method are very close to the exact solution that must be obtained numerically Sohn (1974). The closed-form solution of this problem is discussed in more detail by Sohn (1978).

The model formulated has been used to extract intrinsic reaction rate constant of H_2 reduction of ilmenite from the experimental measurements. A list of parameters used in the model is given in Table II. The model agrees well with the experimental measurements. Sample results for model predictions at 945°C and 1014°C are shown in Figs. 22 and 23. The Arrhenius plot based on the intrinsic reaction rate constants is shown in Fig. 24. The intrinsic activation energy for the reaction on ilmenite core is 16.9 kcal/mole, which is lower than the apparent activation energy discussed earlier. The reason is

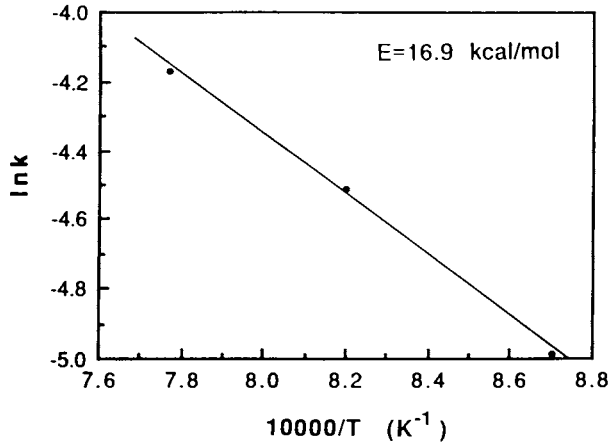


Figure 24. Temperature dependence of the intrinsic rate coefficient.

that the intrinsic rate is influenced by the pore blockage effect particularly at low temperature. As temperature increases, the iron mobility, nucleation and growth are enhanced and the pore blockage effect becomes less significant. This causes an increase in the observed initial rate which is separate and in addition to the usual increase in rate caused by increasing temperature. Due to these dual effects of temperature, the apparent activation energy calculated from the initial rate data is higher than the intrinsic activation energy.

TABLE II
Values of Model Parameters

Parameter (unit)	Value
ϵ	0.33
ρ_s (gmol cm^{-3})	0.032
Average r_s (cm)	7.2×10^{-5}
Thickness of flake (cm)	0.06
D_e (cm^2 s^{-1})	0.95 (945 K)
D_{eg} (cm^2 s^{-1})	0.033 (945 K)
K	0.09 (945 K)

V. COMPARISON OF CO AND H_2 REDUCTION AND A NEW PROCESS FLOWSHEET

Shomate et al. has shown that the theoretically possible conversions for the reduction of ilmenite by CO are 5.1, 6.6 and 7.8% at 827°C, 1027°C and 1227°C, respectively. At the same temperatures, the theoretically possible conversions for the reduction of ilmenite by H_2 are 5.1, 10.5 and 16.7%. Both carbon monoxide and hydrogen would be present in the gaseous stream if the process for the production of oxygen from ilmenite is based on the use

of carbonaceous waste as a carbon source. Therefore, it is very important to compare the reduction of ilmenite by CO with the reduction of ilmenite by H_2.

The mechanisms of ilmenite reduction by H_2 and CO are very similar. Both reactions involve the migration and nucleation of iron, leading to the complete segregation of iron from TiO_2. The main difference between these two reactions is that TiO_2 can be reduced to lower oxides of titanium by hydrogen and the reduction rate of ilmenite by H_2 is much faster than that of ilmenite by CO.

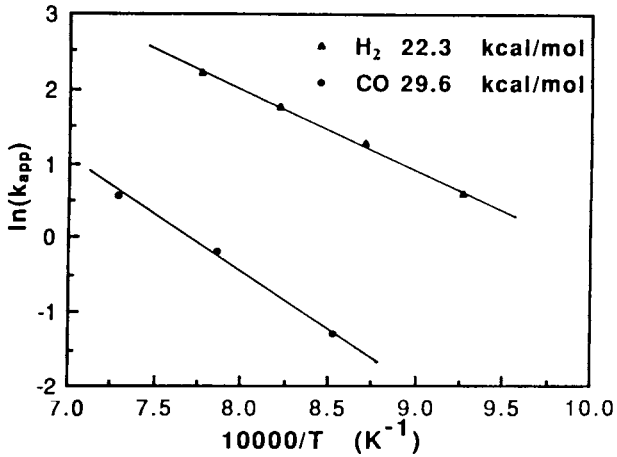


Figure 25. Temperature dependence of the apparent rate coefficient.

The effect of temperature on both reactions is shown in Fig. 25. The apparent activation energy for H_2 reduction of ilmenite is 22.3 kcal/mol, whereas the apparent activation energy for CO reduction of ilmenite is 29.6 kcal/mol. This suggests that the reduction of ilmenite by CO is more sensitive to temperature than that by H_2.

In order to determine the effect of reducing agents on the reaction rate and the time corresponding to the "complete iron metallization," two sets of experiments were conducted at the same condition. The results in Figs. 26 and 27 show that the initial reaction rates of ilmenite reduction by H_2 are 8.6 and 11.3 times larger than those by CO reduction at 1000°C and 900°C, respectively. The times corresponding to the "complete iron metallization" are 12.5 minutes and 34 minutes at 1000°C and 900°C for H_2 reduction of ilmenite, whereas, for the CO reduction of ilmenite, times are 95 minutes and 135 minutes at 1000°C and 900°C.

The flowsheet for a novel carbothermal reduction process has been recently developed and is shown in Fig. 28. The components and their functions in this flowsheet are described as follows:

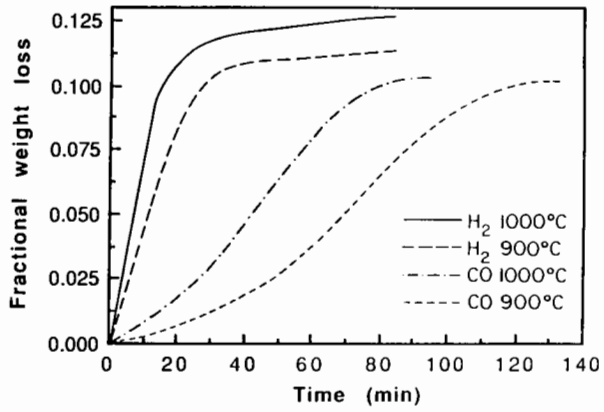


Figure 26. Effect of reducing agents on the reduction of ilmenite: $[H_2]$ or [CO] (10^3 mol/L)—T(°C); 1.15—1000; 1.24—900.

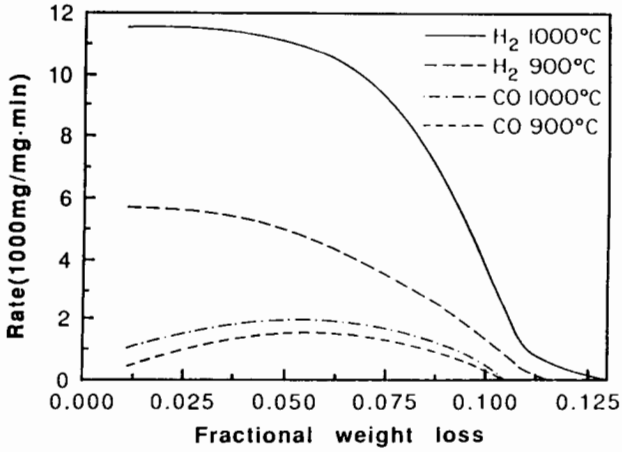


Figure 27. Effect of reducing agents on the reduction rate of ilmenite: $[H_2]$ or [CO] (10^3 mol/L)—T(°C); 1.15—1000; 1.24—900.

1. Carbon monoxide generation section.
2. Reactor with lower operating temperature: this reactor is designed to deposit carbon on lunar ilmenite. The thermodynamic calculation indicates that the deposit of carbon on lunar ilmenite can increase the oxygen yield significantly.
3. Reactor with higher operating temperature: this reactor is to carry out the carbothermal reduction of lunar ilmenite. The major products in this reactor are Fe, TiO_2, CO and CO_2.
4. Oxygen recovery from CO_2 by electrolysis.

The staged reactor system with carbon deposition and reduction occurring

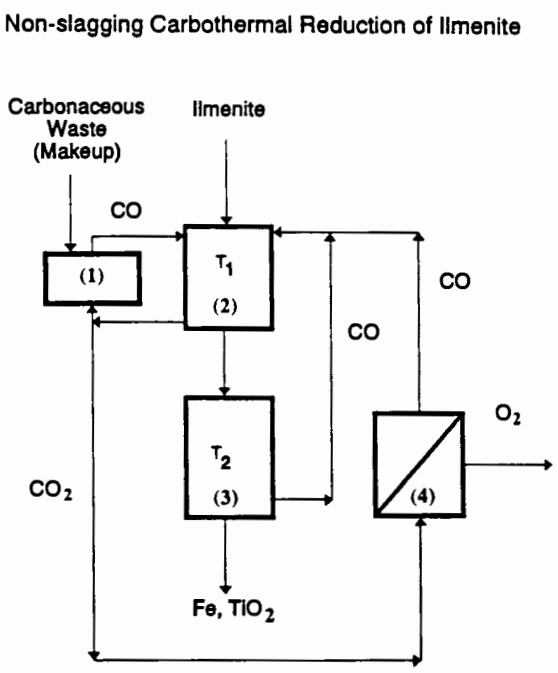


Figure 28. Flowsheet for a novel carbothermal reduction process.

at two different temperatures is a novel scheme that overcomes the inherent yield deficiency in other ilmenite reduction process (Shadman et al. 1991; Zhao and Shadman 1990).

VI. CONCLUSIONS

The reduction of ilmenite grains by H_2 and CO follow a shrinking core configuration and consists of four primary steps: diffusion of H_2 or CO through the product layer TiO_2 reaction of H_2 or CO with the ilmenite, diffusion of iron out of the TiO_2 pores, and nucleation and growth of metallic iron outside the ilmenite grain particles.

The temporal profiles of conversion of ilmenite reduction with H_2 at temperatures below 876°C and by CO at all temperatures exhibit three stages during the reaction: induction, acceleration and deceleration. The induction period is due to the slow transport of iron out of the pores resulting from insufficient iron nuclei at the onset the reduction process.

The apparent activation energy based on initial rate for the ilmenite reduction with H_2 is 22.3 kcal/mole, whereas the apparent activation energy of CO reduction of ilmenite is 29.6 kcal/mole.

The comparison of two reduction processes indicate that the reduction rate of ilmenite by hydrogen is much faster and less sensitive to temperature

than CO reduction. Moreover, TiO_2 can be reduced to lower oxides of titanium by H_2 at temperature higher than 876°C.

The proposed theoretical model agrees well with the experimental measurements and can be used for design and parametric study.

NOMENCLATURE

C_A: concentration of H_2, gmol/cc
C_{Ab}: concentration of H_2 in bulk gas, gmol/cc
C_C: concentration of H_2O, gmol/cc C_{Cb}: concentration of H2O in bulk gas, gmol/cc
D_e: effective diffusivity in macro-pores in flake, cm^2/sec
D_{eg}: effective diffusivity in the micro-pores of product layer in each grain, cm^2/sec
E: intrinsic activation energy, kcal/mole
$f(X)$: function defined by Eq. (15)
k: surface reaction rate coefficient, cm/sec
K: equilibrium constant
L: half thickness of the slab, cm
$g(X)$: function defined by Eq. (16)
$p(X)$: function defined by Eq. (17)
R_A: reaction rate, gmol H_2/cc bulk flake, sec
r_c: radius of unreacted core in each grain, cm
r_s: radius of grain, cm
t: time, sec
t^*: dimensionless time, defined by Eq. (12)
T: temperature, °K
X_B: conversion of each grain
X: overall conversion of flake
Z: distance from the center of the slab, cm

Greek Symbols

ϵ: macro porosity
ρ_s: ilmenite molar density, gmol/cm^3
σ: generalized gas-solid reaction modulus defined by Eq. (13)
σ_g: dimensionless modulus for the reaction of the grain, defined by Eq. (13)

Acknowledgment. This research was supported by NASA/UA Center for Utilization of Local Planetary Resources at the University of Arizona. The NASA/UA Graduate College Fellowship for Y. Zhao is gratefully acknowledged. Discussions with Dr. A. H. Cutler were very helpful in this study.

REFERENCES

Agosto, W. N. 1985. Electrostatic concentration of lunar soil minerals. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 453–464.

Bardi, G., Gozzi, D., and Stranges, S. 1987. High temperature reduction kinetics of ilmenite by hydrogen. *Mater. Chem. Phys.* 17:325–341.

Barnes, C., and Pickles, C. A. 1988. A thermogravimetric studies of the catalytic effect of alkali carbonates on the reduction of ilmenite. *High Temp. Tech.* 6(4):195–201.

Briggs, R. A., and Sacco, J. A. 1988. Oxidation and reduction of ilmenite; application to oxygen production on the Moon. In *Papers Presented to the Symp. on Lunar Bases and Space Activities in the 21st Century*, April 5–7, Houston, Tex., LPI Contrib. 652, p. 34.

Cole, D. M., and Segal, R. 1964. Rocket propellants from the Moon. *Astronaut. Aeronaut.* 2:56–63.

Donnelly, R. P. 1970. Reduction of iron oxide beach sands in ilmenite. *Australian Mining*, March, 1970.

El-Guindy, M. I., and Davenport, W. G. 1970. Kinetics and mechanism of ilmenite reduction with graphite. *Metal. Trans.* 1:1729–1734.

Gupta, S. K., Rajakumar, Y., and Grieveson, P. 1987. Kinetics of reduction of ilmenite with graphite at 1000 to 1100°C. *Metal. Trans. B* 18:713–718.

Jones, D. G. 1975. Kinetics of gaseous reduction of ilmenite. *J. Appl. Chem. Biotechnol.* 25:561–582.

Poggi, D., Charette, G. G., and Rigaud, M. 1973. Reduction of ilmenite and ilmenite ores. In *Titanium Science and Technology*, eds. R. I. Jaffee and H. M. Burte (New York: Plenum), pp. 691–695.

Shadman, F., Zhao, Y., and Massieon, C. 1991. Production of oxygen from lunar ilmenite. In *Annual Progress Report: NASA/University of Arizona Center for Utilization of Local Planetary Resources* (Tucson: UA/NASA SERC), pp. IA-9–IA-68.

Shomate, C. H., Naylor, B. F., and Boericke, F. S. 1946. U. S. Bureau of Mines Report No. 3864 (Washington, D. C.: U. S. Dept. of the Interior.

Sohn, H. Y. 1974. The effect of intragrain diffusion on the reaction between a porous solid and gas. *Chem. Eng. Sci.* 29:630–634.

Sohn, H. Y. 1978. The law of additive reaction times in fluid-solid reactions. *Metal. Trans. B.* 9:89–96.

Wouterlood, H. J. 1979. The reduction of ilmenite with carbon. *J. Chem. Tech. Biotechnol.* 29:603–618.

Zhao, Y., and Shadman, F. 1990. Kinetics and mechanism of ilmenite reduction with carbon monoxide. *AIChE J.* 36(9):1443–1448.

LUNAR OXYGEN PRODUCTION BY PYROLYSIS

CONSTANCE L. SENIOR
PSI Technology Company

Production of oxygen has been identified as a high priority for lunar manufacturing with the primary use for oxygen as a propellant. The economic incentive for oxygen production from lunar material is the savings in transportation costs of oxygen from the Moon as compared to oxygen brought up from Earth. Pyrolysis or vapor-phase reduction involves heating material to temperatures sufficient to allow partial decomposition of metal oxides and vaporization. Some metal oxides give up oxygen upon heating, either in the gas phase to form reduced gaseous species or in the condensed phase to form a metallic phase. Pressures in the range of 0.01 to 0.1 torr are predicted for the pyrolysis step with melt temperatures of 2000 to 2200 K. Metal-containing species in the gas phase can be collected by condensation and thus separated from oxygen. The simplicity and the ability to use unbeneficiated regolith for feedstock make it attractive for lunar manufacturing. This chapter discusses current experimental and theoretical work as well as the process conditions and requirements for feedstock, power, and equipment. Suggestions for future work needed to bring the pyrolysis process to technical maturity are also discussed.

I. INTRODUCTION

Extraction of oxygen from lunar minerals has been studied for over a decade. The processes that have been proposed fall into three general categories: chemical, electrolytic, and pyrolytic, which are discussed in various chapters in this volume. Interest in pyrolysis is based on theoretical arguments concerning vaporization of metal oxides. Many metal oxides vaporize via the formation of reduced oxide or metal atoms in the gas phase (Brewer 1953). The most prominent example is SiO_2 which vaporizes via the reaction

$$SiO_{2(c)} \longrightarrow SiO + 1/2O_2. \qquad (1)$$

SiO will condense as a metastable solid if the gas is rapidly quenched (Nuth and Donn 1983). Because much of the lunar soil and rock is composed of silicates, pyrolysis appears to be a relatively simple way to liberate oxygen.

The evolution of volatiles and more refractory elements has been noted in experiments with lunar and meteoritic materials. Apollo 12 samples were heated in vacuum and the composition of the vaporizing gas was measured (De Maria et al. 1971). Oxygen evolution was observed at temperatures above 1400 K. Meteoritic material has been heated in a solar furnace with the aim of

deducing formation processes for the solar system. Temperatures as high as 3000°C were observed. More than 70% of the material was estimated to have vaporized (Grossman et al. 1982) into the vacuum chamber. Basaltic samples were also processed in the solar furnace, resulting in residues containing only calcium, aluminum, and oxygen (Notsu et al. 1978). Other experiments on vacuum heating of meteoritic materials (Hashimoto et al. 1979) demonstrated that silicon and iron were concentrated in the condensate when the samples were heated in the range 1400° to 1800°C. Condensation of rapidly heated basalt in vacuum showed that the condensed material could be very deficient in oxygen at certain substrate temperatures (Yakovlev et al. 1985).

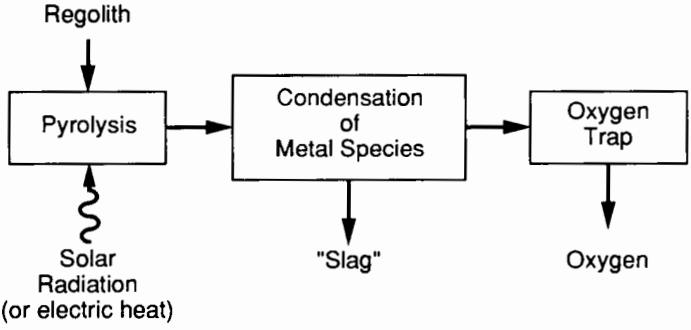


Figure 1. Pyrolysis process.

The pyrolysis process as outlined in Fig. 1 has three major steps: pyrolysis, condensation of metal species, and oxygen recovery. Each of these steps will be discussed in turn with the aim of elucidating the fundamental principles as well as the relevant experimental results which can be used to design a lunar manufacturing process.

Two variations on the pyrolysis process have been proposed (Steurer and Nerad 1983) which take different routes in the separation of oxygen from metal-containing species. In the first, called vapor separation or thermal pyrolysis, lunar material is heated at temperatures up to 3000 K; such temperatures can be obtained using concentrated solar energy. The metal species are condensed from the hot gases to separate oxygen. A second process, called selective ionization or plasma pyrolysis, heats the vapor to very high temperatures (ca. 10,000 deg), generating a plasma. Solar energy cannot be used as a heat source for this process. At temperatures below 9000 K, oxygen atoms are not ionized while metal-containing species are. The resulting plasma is passed through an electrostatic field in which the ionized metals are separated from the neutral oxygen. Considerably less work has been performed to date on the plasma process. Thermal pyrolysis will be considered as a first generation process for production of oxygen, although further in the future, plasma pyrolysis might be implemented.

Pyrolysis was first proposed as a means of producing oxygen in the early

1980s (Steurer and Nerad 1983). At that time, experiments were performed on a single component system; SiO_2 was heated by induction to 2650 K and deposits of silicon metal and suboxide were collected on a cold finger (Steurer 1985). Recently, solar furnace experiments have been conducted with two common lunar minerals (ilmenite and anorthite) in which the pressure increase due to oxygen evolution was observed upon heating of the minerals (Senior 1991*b*). Based on preliminary experiments and equilibrium calculations, the temperatures needed for pyrolysis are expected to be in the range of 2000 to 2500 K, giving total gas pressures of 0.01 to 0.1 torr (Senior 199*a*).

Pyrolysis has several advantages when compared to chemical or electrolytic processes for oxygen production. The analysis of potential advantages or benefits must be seen in the context of the future uses of oxygen and the lunar base environment. The most important products initially from a lunar manufacturing operation will be oxygen and construction materials. Oxygen will be used as a local lunar propellant, but large cost savings will also be realized by using lunar oxygen in low Earth orbit (LEO). Savings in the cost of transporting oxygen from the Earth to LEO are balanced against the cost of transporting processing equipment to the Moon, the cost of supporting personnel for operation and maintenance on the Moon, and the cost of design and testing of the process on Earth.

The primary advantages of pyrolysis are based on the following arguments: (1) simplicity of process steps, (2) use of unbeneficiated feedstock and (3) technological readiness of the process. Each point can be discussed qualitatively since the lack of maturity of lunar oxygen production processes precludes firm quantitative arguments at this time.

Pyrolysis has three process steps and there is no recycling or recovery of reagents, making this process relatively simple compared to chemical reduction processes. Such simplicity reduces the mass of the equipment needed, reduces the labor needed to operate and maintain the equipment, and reduces the time needed on Earth to design and test the process. Furthermore, beneficiation of lunar regolith will not be required, as in the ilmenite reduction process. Mining equipment can, therefore, be reduced in size and there is no need for an extra beneficiation step with its associated equipment and operation requirements.

The technological readiness of the steps in the pyrolysis process is high, because vacuum refining and distillation are practiced on large scales in the production of metals on Earth. Furthermore, high rate vacuum coating operations are used to condense metals and oxides on a variety of substrates industrially. Drawing on terrestrial experience in vacuum processing of materials should reduce the time and cost of designing a lunar process.

Other advantages of pyrolysis which are of secondary importance include: (1) flexibility in primary energy source (solar or electric), (2) no imported chemical reagents, and (3) potential byproducts. The scope of lunar operations as well as the future cost of transporting goods from Earth to Moon determine the importance of these factors.

Potential byproducts of the pyrolysis process include high-temperature refractory glasses or ceramics and metals such as silicon, iron, and magnesium. Pyrolysis can be configured for the selective removal of Na, Mg, Fe, and Si from lunar regolith. The residual material or slag will contain oxides of calcium, aluminum, and titanium which could be used as high-temperature refractory for crucibles, insulation, or vehicle heat shields. Some separation of the condensate can be affected by varying the condensate temperature or by varying the melt temperature. Iron could be vaporized and condensed before silicon, for example.

This chapter discusses the three steps in the pyrolysis process-vaporization, condensation, and oxygen recovery stressing the known experimental results and the theoretical framework needed to model the process. Much of the experimental and theoretical work remains to be done, but based on current knowledge, the process conditions and requirements for feedstock, power, and equipment will be discussed. Finally, suggestions will be advanced for future work needed to bring the pyrolysis process to readiness for lunar manufacturing.

II. PRINCIPLES OF OPERATION

A. Pyrolysis

For equilibrium vaporization at high temperatures, many metal oxides dissociate into reduced species (suboxides or metal atoms). As an example, consider the vaporization of iron which can be described by the following two equations.

$$FeO_{(l)} \Longleftrightarrow FeO_{(g)} \tag{2}$$

$$FeO_{(g)} \Longleftrightarrow Fe_{(g)} + 1/2O_{2(g)}. \tag{3}$$

Using published thermodynamic data, the vapor pressures of the gaseous species can be calculated. Fe is the predominant iron species in the gas phase and the vapor pressure of Fe is significantly higher than that of FeO. Oxygen is also present in the gas phase as a result of the dissociation of FeO. Similar thermodynamic equilibrium calculations for other oxides show that for most oxides, the reduced species are expected to be the predominant gaseous species (Senior 1989). Determination of the equilibrium composition for a single metal oxide system involves a relatively simple calculation which can be done on a hand calculator. The complexity of multicomponent systems requires a more sophisticated computational scheme.

The vapor pressure of a component above a multicomponent liquid requires calculation of the equilibrium composition of the entire chemical system which is commonly accomplished by using an algorithm to minimize the global free energy of the system. Such packages are widely available. For a complex mixture, the total Gibbs free energy is the sum of the free energies for all coexisting phases:

$$G = G_{gas} + G_{liquid} + G_{solid}. \tag{4}$$

The Gibbs free energy of the liquid can be represented as

$$G_{liquid_i} = \Sigma \mu_i^{\circ} n_i + G_{mix} \tag{5}$$

where μ_i° is the standard state chemical potential of the ith component and G_{mix} is the Gibbs free energy of mixing in the liquid relative to the standard state. The standard state that is typically chosen is that of unit activity for all pure substances. The free energy of mixing can be represented as

$$G_{mix_i} = NRT \Sigma X_i \ln X_i + G_{excess}. \tag{6}$$

The first term on the right hand side is the ideal free energy of mixing. The mole fraction of the ith component is given by X_i, equal to n_i/N. The second term on the right-hand side is the excess free energy of mixing which is accounts for the non-ideal behavior of the mixture. The equations for equilibrium can be solved explicitly for the vaporization of single component oxide systems. Multicomponent silicate liquids containing oxides of silicon, calcium, aluminum and iron are known to be non-ideal (Turkdogan 1987).

In this work, we utilize a model known as Ideal Mixing of Complex Components (IMCC). This approach has been used to model the vaporization of alkali species from slags in the MHD process (Hastie and Bonnell 1985) and to model the vaporization that occurred during the formation of the early solar system (Cameron and Fegley 1987). The reason for the non-ideal behavior of these liquids is their ionic character. In the liquid, cations like Ca^{+2} become attracted to anions like SiO_3^{-2}. This attraction in the liquid phase reduces the pressure of the corresponding gaseous components in equilibrium with the liquid. This attraction in the liquid phase can be approximated by including components like $CaSiO_3$ in the liquid mixture. For example, in the CaO-SiO_2 system, the liquid would be a mixture of CaO, SiO_2, and $CaSiO_3$. The correct choice of complex components for a silicate liquid can simplify the calculation of the free energy of mixing. In other words, the non-ideal behavior of the solution can be accounted for entirely by the formation of complex components in the solution.

The IMCC approach has been used to calculate vapor pressures above high temperature melts of simulated regolith composition using the SOLGASMIX program for Gibbs free energy minimization (Eriksson 1975). The thermodynamic data of Cameron and Fegley (1987) were used for the components of the gas and liquid mixtures. This database contains gaseous and liquid species in the Al-Si-Ca-Mg-Fe-Ti-O system. To this database were added the thermodynamic data for gaseous atomic species as well as metallic iron in the liquid phase. Metallic iron has been observed in lunar regolith samples (Allton et al. 1985). Furthermore, the presence of metallic iron dissolved in silicate slags has been shown (Shurygin and Esin 1954).

An example of the application of computational thermodynamic equilibrium to the thermal pyrolysis problem is shown in Fig. 2. Two different

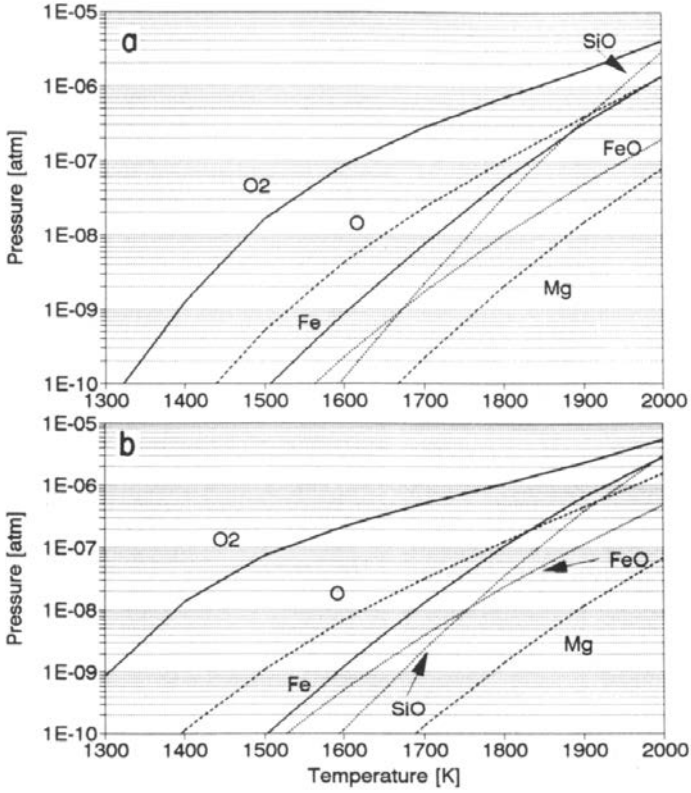


Figure 2. Equilibrium vapor pressures as a function of temperature. (a) Lunar highland soil composiiton; (b) lunar mare soil composition.

starting compositions were used in the calculations (Table I) which are characteristic of lunar highland and lunar mare regolith. The important gaseous species are O_2, SiO, O, Fe, Mg, FeO and SiO_2. Mg is not shown in the figure, but its vapor pressure is approximately equal to that of SiO_2. All other gaseous species are at least 2 orders of magnitude lower in pressure than those in the figure. At temperatures <2000 K, O_2 is the most important gas phase species and is produced by reduction of FeO in the melt to metallic iron. Above 2000 K, the pressure of atomic oxygen approaches that of O_2 and the vapor pressures of metal-containing species, particularly SiO, become significant.

The results of these calculations agree qualitatively with experiments conducted by heating meteoritic, lunar, and terrestrial samples. Samples of the Murchison meteorite were heated in a vacuum chamber (10^{-5} torr chamber pressure) and a mass spectrometer used to identify the gas composition as a function of temperature (Hashimoto et al. 1979). Fe was observed in the gas at 1800 K and SiO began to appear in the gas at temperature >1900 K. Metallic iron was observed in the residue after heating. Vaporization of an Apollo 12

TABLE I
Composition of Lunar Soils used in Equilibrium Calculations

Oxide	Highland Regolith	Mare Regolith
SiO_2	44.8	38.9
TiO_2	0.5	8.3
Al_2O_3	28.1	16.0
FeO	4.2	16.2
MgO	5.5	8.3
CaO	15.7	11.5
K_2O	0.8	0.2
Na_2O	0.4	0.6

soil sample (DeMaria et al. 1971) showed that evolution of Fe was observed to begin at 1400 K, while SiO began to appear at 1700 K.

High temperatures are desirable for producing high vapor pressures and correspondingly high mass flux from the melt. However, the dissociation of oxygen at temperatures above 2000 K makes high temperatures less attractive. Atomic oxygen is a very reactive atom and will probably react with any surface that it strikes. Thus, O atoms should not be considered as part of the oxygen yield because they are more likely to oxidize condensed metallic species than to recombine to form O_2. The dissociation of oxygen was neglected in previous work (Steurer 1985), resulting in the conclusion that temperatures on the order of 3000 K were desirable. However, temperatures in the range of 2000 to 2200 K are preferable because the production of atomic oxygen is reduced. The conclusions from these preliminary equilibrium calculations are that (1) melt temperatures in the range 2000 to 2200 K are desirable and (2) Si, Mg, and Fe are the species that give up oxygen in this temperature range.

Distillation of metal-containing species from molten lunar soil can be modeled at low pressures. According to Langmuir's hypothesis, the evaporation rate (in g cm^{-2} s) at low pressure conditions is related to the equilibrium partial pressure by (Winkler 1971)

$$\frac{dm}{dt} = 0.0583 \epsilon p_e (M/T_s)^{1/2} \tag{7}$$

where ϵ is the condensation coefficient, M is the species molecular weight, and T_s is the melt surface temperature. This equation is valid even when no state of equilibrium exists; that is, the evaporation rate does not depend on the partial pressure in the gas, but only on the concentration of the species and temperature at the surface of the liquid.

The actual rate of evaporation will be limited by one or more of the following processes: (1) heat and mass transport in the liquid, (2) evaporation, (3) mass transport in the gas, and (4) condensation. This problem has been treated extensively theoretically for various regimes of control (see, e.g.,

Krüger 1971). At low enough pressures, the mass transport in the gas phase does not affect the rate. Often in the distillation of metals, the distillation rate is controlled by both liquid mass transfer and evaporation (Krone et al. 1989). In this case, the rate of evaporation (in g s^{-1}) of a species from the melt is expressed as (Krüger 1971)

$$\frac{dM}{dt} = \frac{CA}{\delta/D + C/(0.0583\epsilon p_e)(T_s/M)^{1/2}} \tag{8}$$

where C is the concentration in the liquid, A is the liquid surface area, δ is the liquid boundary layer thickness, and D is the diffusivity in the liquid. Note that the evaporation rate is proportional to the melt surface area.

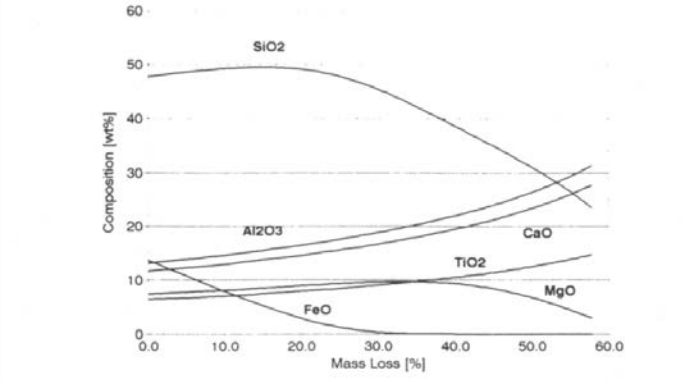


Figure 3. Residue composition after volatilization of mare regolith at 1500°C.

Figure 3 shows an evaporation calculation (a numerical experiment) performed for a starting material which approximates mare regolith. Evaporation at low pressure from an open crucible is assumed with a temperature of 1500°C. Langmuir evaporation (Eq. [7]) has been applied in a step-wise fashion using equilibrium calculations at each time step to compute the gas phase composition and hence the moles of each species removed through distillation. The composition of the residual material as a function of mass loss from the sample is shown in the figure. Iron is lost from the melt first and by the time 30% of the mass has been lost, all the iron has left the melt. The composition of the other major elements in the melt, in contrast, does not change dramatically until after more than 30% of the mass has been lost. Thus, iron is expected to vaporize readily from molten regolith, particularly at low pressures. This calculation agrees qualitatively with the results of vacuum distillation experiments performed by Boschelli and McKay (1987) on lunar simulant glass.

B. Condensation

Generation of oxygen by pyrolysis is half the battle to win oxygen from lunar ores. The metal-containing species generated by pyrolysis must be removed

from the gas phase without recombining with oxygen. Experiments on the deposition of SiO coatings and on vaporization and condensation of basalts in vacuum demonstrate that reduced oxides can be deposited from the gas phase under certain conditions.

Thin films of SiO produced by vacuum evaporation are used widely as coatings for optical components and in microelectronics applications. The refractive index of silicon monoxide coatings depends on the oxygen content of the coating, which can vary with oxygen partial pressure in the gas and substrate temperature. It has been observed (Bradford et al. 1977) that increasing the oxygen content of the gas during deposition of SiO increases the oxygen content of the film. This result is not surprising. However, the observation that increasing the temperature of the substrate during deposition decreases the oxygen content of the films may at first seem counterintuitive. However, the sticking coefficient of oxygen decreases with increasing substrate temperature which will reduce the residence time of oxygen on the surface. Bradford et al. (1977) also suggest that the SiO film is really an atomic-scale mixture of Si in SiO_2. The mobility of silicon atoms increases as substrate temperature is increased which causes larger clusters of Si atoms to be formed which are not oxidized as easily.

Can the observations of SiO condensation be generalized to other metal oxide systems? Yakovlev and coworkers (1985) vaporized basalt samples in vacuum. The starting material was heated to approximately 2500°C which completely vaporized the basalt. The vapor was condensed on a plate held at 25 and 700°C. The elemental composition of the film as a function of distance through the film was determined using secondary ion mass spectrometry. As from Fig. 4, the ratio of oxygen in the film relative to the stoichiometric amount of oxygen has been calculated for substrate temperatures of 25 and 700°C. In this figure, a ratio of less than one indicates that the film is deficient in oxygen.

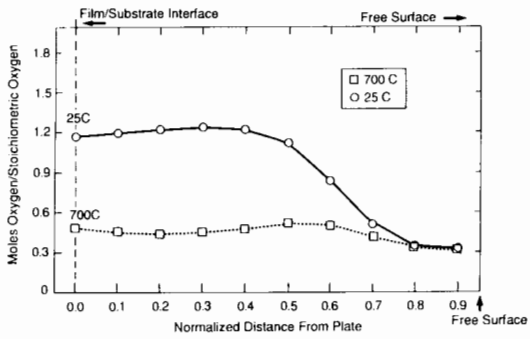


Figure 4. Normalized oxygen content of basalt sample condensed from vacuum as a function of plate temperature (data from Yakovlev et al. 1985).

As with the deposition of SiO films, higher substrate temperatures produce films that are deficient in oxygen. At 25°C, the initial material deposited, which consisted of more volatile elements (Si, Fe, Mg, Na), contained approximately stoichiometric amounts of oxygen. The final material deposited (Al, Ca, Ti) is depleted in oxygen. At a substrate temperature of 700°C, however, the entire deposit contained roughly half the stoichiometric amount of oxygen. The oxygen content of basalt is on the order of 0.4 $gO_2\ g^{-1}$, similar to lunar regolith. These experimental results indicate that it is possible to extract half the oxygen from the original sample by condensation. This corresponds to 0.2 $gO_2\ g^{-1}$. The substrate temperature is a key parameter for condensation.

III. PROCESS DESIGN

A. Industrial Experience

Large scale vacuum processing has been practiced industrially for almost a century to produce high-quality ferrous and non ferrous alloys. Production systems currently in use have charge weights of 1 to 30 tons (Betz et al. 1989). Degassing of metals has been practiced in vacuum systems for many years. Since the late 1940s, the need for alloys with high temperature resistance to creep failure has led to the addition of elements such as Ti and Al to stainless steels and nickel-based alloys (Cremisio 1989). Such alloys cannot be melted in air due to the formation of oxides. As the demand for superalloys in the aerospace industry grows, vacuum induction melting furnaces have been used increasingly for degassing, refining, and casting.

Vacuum melting furnaces are operated in batch mode using a refractory-lined crucible which is filled with an initial charge. Heating is accomplished by induction or arc melting. In some cases, the crucible can be tipped to pour the charge into molds under vacuum conditions. Scale up of vacuum systems is readily accomplished provided that attention is paid to mass transfer considerations. To promote good mass transfer, good mixing in the liquid phase is important. More than half the vacuum induction melting furnaces use electromagnetic stirring devices and others bubble an inert gas through the melt to increase mixing and liquid-to-gas mass transfer (Betz et al. 1989). Melt surface area is the other important factor in design of vacuum degassing and refining systems. Higher surface area reduces the processing time by increasing the rate of evaporation from the melt.

Vacuum distillation of metals, although not widespread, is also used in process metallurgy. Recently, a pilot scale demonstration was made of the distillation of Li and other impurities from an aluminum melt (Krone et al. 1989). Laboratory tests with a 2 kg charge led to the design and operation of a pilot scale unit with charge weights of 100 to 150 kg. The distillation took place at a pressure of 10^{-6} torr and temperatures of 1200 to 1500 K. In the pilot scale tests, the distillation rate of Li was 14 to 20 kg $m^{-2}\ hr^{-1}$. The condensate contained up to 80% Li while the Li content of the melt was reduced by 2 orders of magnitude from 2% to 0.002%. The authors

note that the distillation rate was controlled by both diffusion in the melt and evaporation from the surface.

The application of coatings under vacuum resembles the condensation step for separating metals from oxygen. Coating rates of 10 to 50 kg hr^{-1} are routinely practiced (Schiller et al. 1975). In some applications, scale up to substrates as large as 10 m^2 has been demonstrated (Grubb 1973). Substrates are often moving as, for example, in the continuous application of aluminum to strip steel (Schiller et al. 1975) or of antireflection coatings to plate glass (Grubb 1973).

In industrial vacuum coating, evaporation is usually accomplished by using an electron beam heater; electron beam evaporation sources in the range of 150 to 250 kW have been used in various industries. Multistage pumping systems are used to reduce pressure in the coating chamber to as low as 10^{-6} torr. Current industrial practice, therefore, results in conditions for coating that are similar to the condensation step in pyrolysis.

B. Feedstock Selection

Bulk regolith is the most readily available feedstock for pyrolysis. Table I gives representative major element compositions for highland and mare soils. The character of lunar ores is discussed in more detail elsewhere in this volume; therefore, only the points relevant to pyrolysis will be discussed here. The oxygen yield from various oxides is summarized qualitatively in Table II. The theoretical yields, assuming that the reduction reactions go to completion as written, are given.

TABLE II

Theoretical Total Oxygen Yield for Single and Multicomponent Oxides

Solid Phase	Gas Phase	g O_2 g^{-1} Solid
SiO_2	$SiO + 1/2O_2$	0.27
MgO	$Mg + 1/2O_2$	0.40
FeO	$Fe + 1/2O_2$	0.22
CaO	$Ca + 1/2O_2$	0.29
Al_2O_3	$2AlO + 1/2O_2$ (moderate temperatures)	0.16
Al_2O_3	$Al + 1/2AlO + O_2$ (high temperatures)	0.31
$FeTiO_3$	$Fe + TiO_2 + 1/2O_2$	0.11
$CaAl_2Si_2O_8$	$Ca + SiO + 2AlO + 3/2O_2$	0.23
$CaAl_2Si_2O_8$	$Ca + SiO + AlO + Al + 2O_2$	0.29

Equilibrium calculations point to operating conditions (i.e., temperature and pressure) that maximize oxygen yield. Pyrolyzing lunar materials at low pressures requires relatively moderate temperatures in the range 2000 to 2500 K. Two distinct modes of operation are possible. Silicon, magnesium,

and iron can be separated from calcium and aluminum by operating at the low end of the temperature range, as in the experiments of Boschelli and McKay (1987). This reduces the potential oxygen yield, but a better separation of the metals results. Based on the theoretical yields in Table II, the expected yields of oxygen are 0.16g O_2 g^{-1} soil and 0.20g O_2 g^{-1} soil for the highland and mare soils, respectively. Alternatively, higher temperatures can be used to extract oxygen from calcium and aluminum compounds as well. The expected yields of oxygen are 0.35g O_2 g^{-1} soil and 0.25g O_2 g^{-1} soil for the highland and mare soils, respectively.

The yield of oxygen will not be equal to the maximum yield, of course. The experimental results of Yakovlev et al. showed approximately 0.2 gO_2 g^{-1} at $T > 2500$ K, corresponding to at least 60% of the maximum yield. (This is approximate as the staring basalt composition was not reported.) We can use 60% of the maximum yield as an upper bound. However, it is likely that $<100\%$ of the soil will be vaporized in an actual process. As a lower bound, assume that only 25% of the soil is actually vaporized. When combined with 60% efficiency during condensation, this gives 15% of the maximum O_2 yield as a lower bound. Therefore, the expected yields of O_2 from bulk regolith are in the range of 0.024 to 0.12 gO_2 g^{-1} for low-temperature pyrolysis and 0.063 to 0.21 gO_2 g^{-1} for high-temperature pyrolysis. These yields are comparable to those quoted (see, e.g., Chapter by Taylor and Carrier) for magma electrolysis (0.008 gO_2 g^{-1}) and ilmenite reduction (0.015 gO_2 g^{-1}).

Extraction of oxygen from all the major metal oxide species favors use of highland soil because of the high aluminum content and low titanium content. Extraction of oxygen from only silicon, magnesium, and iron species favors use of mare soils because of their higher iron content.

C. Pyrolysis

The design of the pyrolysis step for lunar oxygen production draws on theoretical considerations and industrial experience on Earth. Refractory-lined crucibles are commonly used in glass and metallurgical melting furnaces. Lunar regolith is more corrosive than commercially produced glasses, however. The high iron content and the correspondingly low SiO_2 and Al_2O_3 contents of regolith ensure that most refractory oxide materials will dissolve at pyrolysis temperatures. Equilibrium calculations (Fig. 5) were performed with excess of Al_2O_3, $MgAl_2O_4$ and Mo to simulate the crucible materials in equilibrium with a high Ti mare basalt. Both Al_2O_3 and spinel ($MgAl_2O_4$) dissolve in the melt appreciably at 1600 K. Molybdenum is not soluble in the melt, but the formation of gaseous Mo oxides severely restricts the oxygen production which using Mo as a crucible. Careful consideration must be given to the selection of the crucible material.

Various options exist for heating the lunar regolith to temperatures in the range of 2000 to 2500 K. Concentrated solar energy is the most obvious source for energy, although fogging of windows will be a problem. A radiator, as proposed by Henson and Drexler (1978), could be used to transfer heat

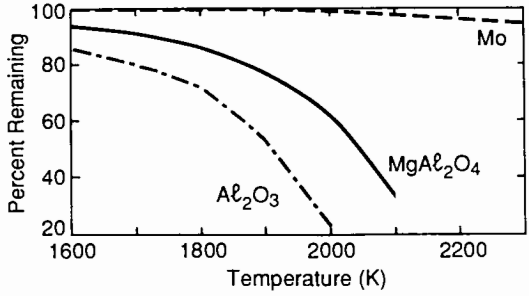


Figure 5. Equilibrium between high Ti mare basalt and crucible composed of Al_2O_3, $MgAl_2O_4$ or Mo. Percent of original crucible remaining.

to the melt efficiently. Figure 6 shows a high surface area radiator which could be made of graphite backed with an oxidation-resistant material such as silicon carbide. An inert gas is cycled through the window cavity to prevent condensation of carbon vapor on the window.

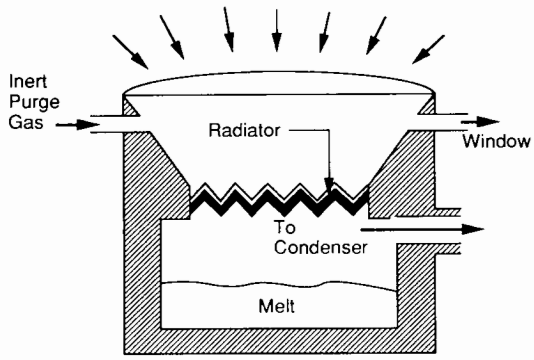


Figure 6. Solar heater for thermal pyrolysis of lunar regolith.

Pyrolysis can be accomplished in either batch or continuous mode. Batch processing of glass and metals is common on Earth and design of batch vacuum processing is well understood. Laboratory experiments of a semi-batch process, Rayleigh distillation, applied to lunar simulants has been demonstrated (Boschelli and McKay 1987). In Rayleigh distillation, the condensed phase is treated batch-wise, while the gas that is evolved from the liquid is removed continuously, giving a better final separation between components of the initial mixture than can be attained by equilibrium distillation. This mode of operation would be well suited for the selective pyrolysis of the more volatile oxides of Fe, Mg, and Si. The remaining liquid phase would be enriched in Ca, Al, and Ti. Figure 7 illustrates the batch approach to pyrolysis.

Limitations on the batch approach arise from the increasing viscosity of

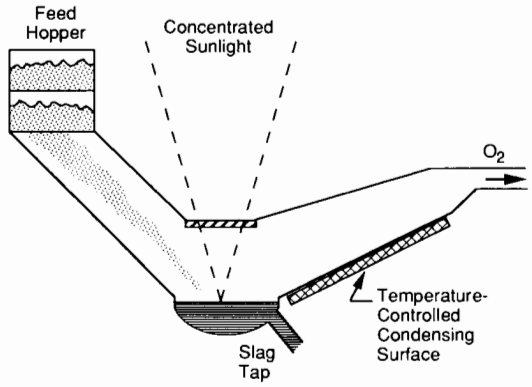


Figure 7. Schematic for pyrolysis process for production of oxygen from lunar regolith.

the melt as Fe and Mg are removed. For good mass transfer, the liquid must be stirred; an electromagnetic stirring device could be used. As the composition of the liquid changes, either the temperature of the melt must be increased or the energy to the stirrer must be increased to compensate for the viscosity changes.

The viscosities of lunar soil simulants are surprisingly low. Measurements of the viscosity of a glass with the composition of an Apollo 11 soil sample have been made (Murase and McBirney 1970). The reported viscosities range from 7.1 poise at 1668 K to 5 poise at 1723 K and are similar to the viscosity of heavy motor oil at room temperature. At melt temperatures approaching 2000 K, the viscosity will be even lower which should provide good mass transfer in the melt with stirring.

The removal of elements such as Fe increases the viscosity of silicate glasses (Williamson et al. 1968). As the more volatile elements are distilled from the melt during the pyrolysis process, the melt viscosity will increase, making it more difficult to promote good mixing in the liquid. A continuous process would solve this problem by maintaining the melt viscosity at a constant value. Regolith would be added and slag would be withdrawn from the melt on a continuous basis. Good mixing of the liquid during continuous pyrolysis remains as important as in the batch process.

D. Condensation

The condensation step must meet several objectives. Metal-containing species must be removed without recombining with oxygen in the gas or on the surface. The deposit must be easily removed, preferably on a continuous basis. Coating moving substrates is routinely practiced in terrestrial processing and this technology might be adapted to the pyrolysis process. The substrate temperature should be kept high enough to reduce sticking of oxygen. However, if temperatures are too high, the condensed material may form a dense deposit

that is difficult to remove. Alternatively, the deposit could be removed by localized melting.

By varying the temperature of the condensing surface, metal species might be segregated. This could lead, for example, to separate recovery of metallic iron and SiO. Further refining of SiO is possible: when heated above 500°C, SiO reverts to a mixture of silicon and SiO_2. The SiO_2 could be separated from Si and recycled into the feed or used to make optical quality glass.

E. Oxygen Recovery

The anticipated pressures during the pyrolysis step are in the range of 0.01 to 0.1 torr. Assuming that most of the gas is oxygen (Fig. 2) and that some cooling takes place during the condensation step, the residual oxygen pressure after condensation might be an order of magnitude lower, 0.001 to 0.01 torr. Thus, the pressure of the oxygen stream must first be raised. Because the desired product is liquefied oxygen, the oxygen will probably need to be compressed to high pressures to achieve efficient liquefaction.

The most obvious method for producing high pressure oxygen is to use a compressor. Multiple stage vacuum pumps are used on an industrial scale on Earth to operate between 1 atm pressure and pressures as low as 10^{-6} torr. The lack of heat sinks on the Moon may complicate the design of such pumps, but this is a challenge faced by many different pieces of equipment needed for the lunar environment.

A novel method for increasing the oxygen pressure would be to adsorb the oxygen reversibly at low pressures. A molecular sieve could be used to trap oxygen at low pressure. Thermal regeneration of the trap could then be carried out at high pressure. The advantage of the oxygen trap is that solar energy or possibly waste process heat could be used to produce high pressure oxygen instead of electricity. Compression might still be needed to reach the final pressure, but some of the mechanical work would be saved.

IV. SUMMARY

In summary, a process for oxygen production has been proposed. Preliminary experimental work and terrestrial process experience provides insight into how the process might be configured. Pressures in the range of 0.01 to 0.1 torr are predicted for the pyrolysis step with melt temperatures of 2000 to 2200 K. Yields of oxygen are estimated to be in the range of 0.02 to 0.2 $gO_2\ g^{-1}$ soil.

The temperatures that have been suggested for pyrolysis could be attained by using either solar power or electrical power. Induction heating, electrodes or an electron beam could be used in place of concentrated solar radiation. The ability to use either solar or electrical power means greater flexibility for the pyrolysis process.

Condensation of metal-containing species without oxidation is the key to good oxygen yields. The temperature of the condensing surface determines the

degree of oxidation of the deposit as well as the porosity of the deposit. Very dense deposits may be difficult to remove. Finding the optimum condensation temperature will mean balancing the ease of removal of the deposit with its oxygen content.

After condensation, the oxygen stream will have to be compressed to high enough pressures so that liquefaction can be carried out efficiently. Multistage compressors might be used. Adsorption techniques, while not proven technology, might be used to produce oxygen at moderate pressures.

The process could be run in either a batch or continuous mode. Batch distillation would produce good separation between more volatile elements (Si, Fe, Mg) and the more refractory elements (Al, Ca, Ti), but the increasing melt viscosity might be result in mixing problems. The pyrolysis process will become better defined as more work is carried out as discussed in this chapter. Research is needed in several key areas in order to advance the level of technical maturity of the process.

Thermodynamic and transport properties of molten lunar regolith are critical to the design of the process. Fortunately, the literature in the iron and steelmaking literature already contains much of the data and models needed for prediction of liquid activity, heat capacity, viscosity, and diffusion coefficients. One key difference between molten regolith and industrial slags is the oxidation state of iron. In the former case, the iron is all Fe^{+2} state, while for most terrestrial slags, the iron is predominantly Fe^{+3}. The coordination state of iron in silicate melts depends on the oxidation state and this can have a dramatic effect on viscosity. Viscosity measurements on a glass with the composition of lunar sample 15555 (Cukierman and Uhlmann 1974) revealed that the viscosity increased *3 orders of magnitude* when the glass was oxidized from a ratio of Fe^{+2}/Fe of 0.76 to a ratio of 0.2. The activity of iron in silicate melts as well as its vaporization behavior will depend on its oxidation state. This is an area where further experimental and theoretical work is needed.

If current terrestrial practice is followed, a crucible will be used to contain the molten regolith during pyrolysis. The materials of construction must withstand higher temperatures than are routinely used in terrestrial glass processing. Good mixing in the liquid is necessary. Electromagnetic stirring devices are current practice on Earth; their adaptation to a lunar environment needs to be considered. Removal of the molten slag must be accomplished without premature crystallization. Plugging of slag taps remains a problem in coal-fired boilers on Earth and there is good reason to believe this will be a problem on the Moon, too.

Condensation of reduced species is vital to this process, yet relatively little experimental or theoretical work has been done that is directly related to pyrolysis process conditions. The effect of substrate temperature on the oxygen content and adhesion of the deposit needs to be studied. Continuous removal of condensate will ultimately be practiced. The vacuum coating industry may provide lessons in continuous coating techniques.

Oxygen recovery and liquefaction are problems that is common to all

oxygen production processes, yet comparatively little has been written on the subject. Recovery of oxygen at low pressures might be accomplished by adsorption instead of pumping using a molecular sieve, a zeolite, for example. Capture and regeneration of oxygen needs to be tested to determine the feasibility of this idea.

The sequence for process development can be outlined briefly. Ultimately, a demonstration at the pilot scale should be the goal of this activity. Along the way to this goal, there is much to be learned about the properties of lunar ores and processing on the Moon.

1. Development of process models for vaporization and condensation of lunar regolith. This effort starts with the prediction of fundamental properties such as activity and viscosity and ends with a model for distillation and oxygen recovery.
2. Calculation of mass and energy balances in the individual process steps. In the current state of knowledge, choices need to be made as to the heat source, crucible geometry, etc. These questions cannot be answered until the options can be compared quantitatively.
3. Laboratory demonstration of the operation of subsystems. Pyrolysis, condensation, and possibly oxygen recovery must be demonstrated on a bench scale. Such testing will also serve to verify process models.
4. Pilot scale demonstration. An integrated and continuous test of the operation must be made on Earth first. Success of this test will lead to the specification of equipment for a lunar pilot scale module.

REFERENCES

Allton, J. H., Calindo, C., and Watts, L. A. 1985. Guide to using lunar soil and simulants for experimentation. In *Lunar Bases and Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.) pp. 497–506.

Betz, U., Kremmer, H., Schlebusch, D., and Schwartz, W. 1989. Vacuum induction melting—new developments in equipment design and process control. In *Special Melting and Processing Technologies*, ed. G. K. Bhat (Park Ridge, N. J.: Noyes Publ.), pp. 183–210.

Boschelli, L. J., and McKay, D. S. 1987. Differential volatilization of lunar impact glass using Rayleigh fraction modeling. In *Proc. Lunar Planet. Sci. Conf.* 18:109–110.

Bradford, A.P., Hass, G., and McFarland, M. 1977. The effect of the substrate temperature on the optical properties of reactively evaporated silicon oxide films. *Thin Solid Films* 42:361–367.

Brewer, L. 1953. The thermodynamic properties of the oxides and their vaporization processes. *Chem. Rev.* 52:1–75.

Cameron, B. and Fegley, A. G. W. 1987. A vaporization model for the iron/silicate fractionation in the Mercury protoplanet. *Earth Planet. Sci. Lett.* 82:207–222.

Cremisio, R. S. 1989. The evolution of vacuum melting process routes as a function of superalloy property requirements. In *Special Melting and Processing Technologies*, ed. G. K. Bhat (Park Ridge: Noyes Publ.), pp. 98–125.

Cukierman, M., and Uhlmann, D. R. 1974. Effects of iron oxidation state on viscosity, lunar composition 15555. *J. Geophy. Res.* 79:1594–1598.

De Maria, G., Balducci, G., Guido, M., and Piacente, V. 1971. Mass spectrometric investigation of the vaporization process of Apollo 12 lunar samples. In *Proc. Lunar Sci. Conf.* 2:1367–1379.

Eriksson, G. 1975. Thermodynamic studies of high temperature equilibria. *Chemica Scripta* 8:100–103.

Grossman, L., Kawabe, I., and Ekambaram, V. 1982. Chemical studies of evaporation residues produced in a solar furnace. *Meteoritics* 17:4.

Grubb, A. D. 1973. Design and operation of a large scale semicontinuous electron beam evaporator. *J. Vac. Sci. Technol.* 10:53–57.

Hashimoto, A., Kumazaw, M., and Onumua, N. 1979. Evaporation metamorphism of primitive dust material in the early solar nebula. *Earth Planet. Sci. Lett.* 43:13–21.

Hastie, J. W., and Bonnell, D. W. 1985. A predictive phase equilibrium model for multicomponent oxide mixtures—Part II. Oxides of Na-K-Ca-Mg-Al-Si. *High Temp. Sci.* 19:275–306.

Henson, H. K., and Drexler, K. E. 1978. Vapor-phase fabrication of massive structures in space. In *Space Manufacturing Facilities II*, ed. J. Grey (New York: AIAA), pp. 207–213.

Krone, K., Rupp, H., Seebaur, C., and Wilson, W. R. 1989. Vacuum distillation in a induction furnace—a method for recycling aluminum-lithium scrap. In *Special Melting and Processing Technologies*, ed. G. K. Bhat (Park Ridge, N. J.: Noyes Publ.), pp. 222–250.

Krüëger, J. 1971. Use of vacuum techniques in extractive metallurgy and refining of metals. In *Vacuum Metallurgy*, eds. O. Winkler and R. Bakish (New York: Elsevier), pp. 145–321.

Murase, T., and McBirney, A. R. 1970. Viscosity of lunar lavas. *Science* 167:1491–1493.

Notsu, K., Onuma, N., Nismida, N., and Nagasawa, H. 1978. High temperature heating of Allende meteorite. *Geochim. Cosmochim. Acta* 42:903–907.

Nuth, J. A., and Donn, B. 1983. A note on the composition and structure of interstellar grains. *Astrophys. Space Sci.* 95:175–178.

Schiller, S., Foerster, H., and Jaesch, G. 1975. Possibilities and limitations of large-scale electron beam evaporation. *J. Vac. Sci. Technol.* 12:800–806.

Senior, C. L. 1989. Production of Oxygen and Other Products By Pyrolysis of Lunar Materials. Final Report under NASA Contract NAS9-18102, PSI Report TR-948.

Senior, C. L. 199*a*. Lunar oxygen production by pyrolysis. In *Space Manufacturing 8, Proc. of the Tenth Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 331–341.

Senior, C. L. 1991*b*. Solar heating of common lunar minerals for the production of oxygen. *J. Brit. Interplanet. Soc.* 44:579–588.

Shurygin, P. M., and Esin, O. A. 1954. *On the Solubility Of Iron In Liquid Slags.* Brutcher Trans. No. 3415. Trans. from *Dokl. Akad. Nauk SSR* 95:1043–1045.

Steurer, W. H. 1985. Lunar oxygen production by vapor phase pyrolysis. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSI Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnam and G. Maryniak (New York: AIAA), pp. 123–131.

Steurer, W. H., and Nerad, B. A. 1983. Vapor phase reduction. In *Research on the Use of Space Resources*, ed. W. F. Carroll, JPL Publ. 83-36.

Turkdogan, E. T. 1987. *Physicochemical Properties of Molten Slags and Glasses* (London: The Metals Society), pp. 93–125.

Williamson, J., Tipple, A. J., and Rogers, P. S. 1968. Influence of iron oxides on kinetics of crystal growth in CaO-MgO-Al2O3-SiO2 glasses. *J. Iron Steel Inst.* 206:898–903.

Winkler, O. 1971. Thermodynamics and kinetics in vacuum metallurgy. In *Vacuum Metallurgy*, eds. O. Winkler and R. Bakish (New York: Elsevier), pp. 1–93.

Yakovlev, O. I., Fainberg, V. S., Shapkin, A. I., and Ramendik, G. I. 1985. Melt evaporation under fast heating conditions. *Lunar Planet. Sci.* XVI:924–925 (abstract).

COST AND BENEFITS OF LUNAR OXYGEN: ECONOMICS, ENGINEERING AND OPERATIONS

BRENT SHERWOOD and GORDON R. WOODCOCK
Boeing Defense and Space Group

The practicality of producing oxygen from indigenous lunar resources is addressed from three facets: technical, economic, and evolutionary. Technical complications of integrating oxygen production into the function of a lunar base are summarized, based on generalizing conclusions from the comprehensive point-design and analysis of one early, oxygen-producing lunar base concept. Economic viability is assessed using three methods: parametric equations, return-on-investment analysis, and input-output modeling. Requirements for technology advancement and the marginal economic return resulting from these analyses are evaluated in the light of long-term expansion of lunar capabilities. The development of lunar oxygen production is found to appear feasible and guardedly advisable, albeit only as a government-funded venture.

I. INTRODUCTION

Human expansion into deep space will, like all expansion has on Earth, depend on an ability to use local materials. High transportation costs assure this. Because it is likely that at least early cis-lunar space transportation will be based on cryogenic chemical (lunar O_2/lunar H_2) propulsion, and because most of the mass of such a system is propellant, and 6/7 of *that* is oxygen, it seems that oxygen would be a good candidate for early indigenous space materials utilization (ISMU). As is well known, the Moon itself contains an enormous amount of oxygen. The fundamental problem is its accessibility as a resource. To our knowledge, oxygen exists on the Moon in two forms: in the regolith as adsorbed gas originating in the solar wind, and bound up in minerals in the regolith and parent rock itself. Neither source would be considered at all practical by terrestrial resource recovery standards: the solar wind oxygen concentration is low enough to force the processing of huge amounts of regolith to get it in large quantities, and oxygen is comparatively hard to extract directly from rock. Compared even to Mars, where oxygen can be derived both directly and indirectly from the tenuous atmosphere, or from trapped and frozen water in the ground, the Moon will not give up its oxygen easily for human use.

Many processes continue to be proposed and studied for producing lunar-derived oxygen (LLOX) in quantities useful for propulsive consumption. Progress in understanding some of the candidates is described elsewhere in this

volume. The purpose of this chapter is to provide the perspective of a realistic program context for that scientific inquiry. We examine some basic issues of how useful LLOX would reasonably be, and how complex an entire, emplaced LLOX operation would realistically be. By advancing our understanding of both the basic processes *and* the implications of full-scale implementation of those processes—that is, by balancing science with systems engineering—we can discover over the next decade the highest-leverage research directions.

The question we are trying to answer is this: is what we get out of a LLOX production capability worth the difficulty? There are several facets to this question. First (see Sec. II) is the *technical* side: what are the hidden technological hurdles to be overcome, and the ancillary engineering costs that have to be paid, in order to achieve LLOX production in the first place? If indeed there are more failed mining operations on Earth than successful ones, how much robustness must we build in and how much inefficiency must we be willing to tolerate to achieve reliable production on the Moon? The second facet (see Sec. III) is the economic side: is the return on investment positive and sufficiently large over reasonable payback times to warrant LLOX development? Several sensitivities are key: where in the program architecture is it sensible to use LLOX? How much oxygen does the program have to consume before LLOX production makes sense? How expensive can the LLOX production equipment afford to be? How heavy can the delivery technology afford it to be? How much human tending can the program afford it to need? How long can the program afford to take before LLOX production starts? How cheap does Earth-to-orbit launch have to get before LLOX production loses its attractiveness? A third facet (Sec. IV) is the *evolutionary* side: what other, perhaps more necessary ISMU operations will be enabled by having achieved LLOX production technology? What capabilities will we need to develop anyway, whether we recover oxygen or not? What are the opportunity costs of not learning how to get oxygen out of lunar rocks, or of not learning how to run ISMU operations on the Moon at all? We will examine each of these three facets in turn.

II. TECHNOLOGY OF LLOX PRODUCTION

The issues of LLOX production engineering divide loosely but conveniently into three categories: transportation constraints, environmental factors and operations.

A. Transportation Constraints

Life support, while clearly requiring oxygen, is not a driving demand for its production. During the early exploration phase, life support requirements are much too small to warrant LLOX production. Oxygen recycling will be baselined for all surface missions exceeding a few days' duration regardless of crew size. Should a large lunar settlement experience substantial air makeup needs due to leakage and airlock cycling, the driving concern will be nitrogen

(70% of a 10.2 psi atmosphere), which is much scarcer on the Moon than oxygen.

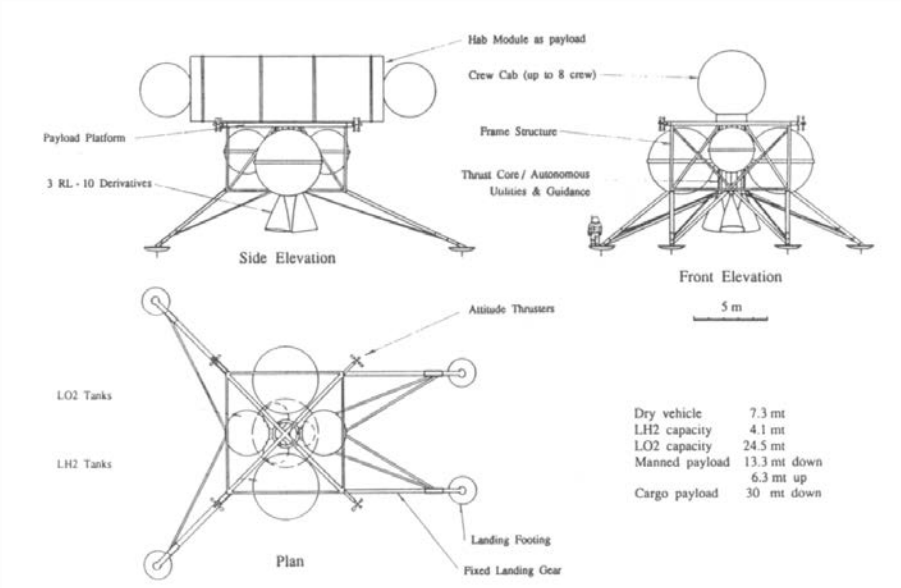


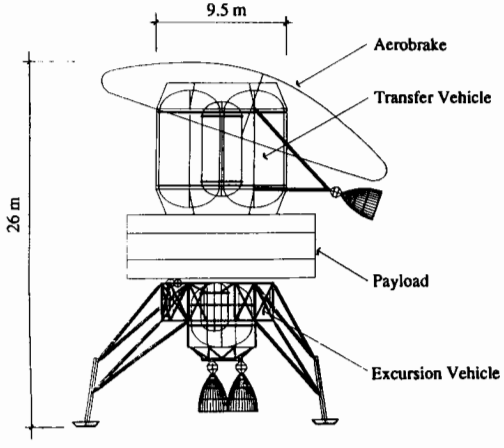
Figure 1. Cryogenic lunar lander with 30 tonne down-cargo capacity.

Transportation, on the other hand, introduces a real demand. Roving surface transportation may use oxygen/hydrogen fuel cells for mobile power, but these also can be regenerable, conserving systems. However, propulsive vehicles (jetpacks, hoppers, flyers, landers, space transfer vehicles) consume and expel large amounts of mass. The quantitative engineering results presented in this section are all based on a LLOX production requirement of 100 tonne yr^{-1}. One hundred tonnes is sufficient oxidizer for retanking a typically sized lunar lander (Fig. 1) four times. Full tanks enable such a lander to ascend without payload to low lunar orbit (LLO), and land again with about 30 tonne cargo and sufficient hydrogen fuel (LH_2) for the next ascent, both transferred to the lander from a transfer vehicle arriving from Earth orbit. Or, the same lander could launch several crew in a crew cab to LLO, and then land a new crew together with over 10 tonne of supplies and equipment, and LH_2 for the next ascent. Four lander flights per year, sustained over several years, enable a rich menu of basic and applied lunar science investigations to be performed. If required by the program plan, and especially if augmented by ISMU, this flight rate can even enable a substantial lunar population to accumulate (Boeing 1991). The propulsive use of about 100 tonne 3 yr^{-1} of LLOX is a practical, reference transportation demand requirement.

Incidentally, a capability to produce indigenous LLOX *and* LH_2 (by recovering the solar-wind-implanted gas from vast quantities of regolith)

would enhance lunar exploration operations greatly. A lander able to retank both cryopropellants on the lunar surface could perform round-trip, suborbital ballistic excursions ranging up to 500 km from the retanking base. That opens up 785,000 km^2, or 2%, of the Moon's surface area around any such location to direct human exploration. Such a vehicle could also rescue and return a rover stranded several hundred kilometers away from the home base, significantly enhancing the safety of human operations on the Moon.

LLOX production is as intimately tied to space transportation constraints as it is to transportation demand. *Sizing* a lunar lander depends on many interrelated factors, only one of which is the desirable size of its payload. For a repeating carrier, logistical concerns tend to dominate. In particular, matching the vehicle hardware, resupply needs, and payload manifests to integral numbers of Earth-to-orbit flights is important for efficient operation. Compatibility, or even commonality, with other elements of the entire transportation architecture may also drive lander sizing. Cargo capacities ranging from 20 to 60 tonne continue to be studied, and the optimal size cannot be specified yet. However, only two down-cargo capacities are likely to be available finally (one for a reusable mode and a larger one for a delivery-only mode), and all surface system payloads will conform to them. The surface system examples shown in this chapter assume the delivery system can land 30 tonne at once.



Mass Statement (LEV Delivered)

Transfer Vehicle Inerts	10,638
Transfer Propellant	107,719
Aerobrake	3,000
Excursion Vehicle Inerts	5,592
EV Propellant	23,151
Payload	24,000
IMLEO (in kg)	174,100

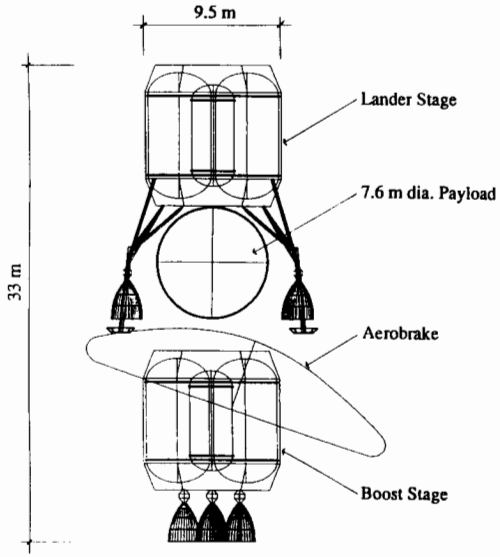
Mass Statement (LEV Refueled)

Transfer Vehicle Inerts	10,638
Transfer Propellant	107,100
Aerobrake	3,000
Resupply Tanks	2,377
Resupply Propellant	23,765
Payload	25,000
IMLEO (in kg)	171,880

Figure 2. Lander configuration with top-mounted payload.

Configuring a lander also depends on many factors besides the preferred geometry of its payload. Propulsive considerations are paramount. The payload must be predictably, and perhaps adjustably, mass-balanced to insure stable flight. Available thrust in nominal and contingency cases, throttling rates and depths, and feasible gimbal authority all affect the number of engines on the vehicle and their positioning. How this is resolved in turn affects payload

placement. Top-mounted payloads (Fig. 2) are less envelope-constrained, and generally allow the use of fewer engines, but require tall, straddling surface mobility vehicles for practical offloading. Underslung payloads (Fig. 3) are closer to the surface upon touchdown, but risk damage from bad landings and engine blast effects, and require very low, trailer-type offloading vehicles that cannot handle unprepared lunar soil conditions. Surface systems, including LLOX production equipment, will have to accommodate one or the other set of constraints.



Mass Statement

Boost Stage Inerts	10,638
Boost Propellant	108,950
Aerobrake	3,000
Lander Stage Inerts	11,113
Lander Propellant	92,603
Payload	60,000
IMLEO (in kg)	**286,304**

Figure 3. Lander configuration with underslung payload.

Payload envelope dimensions as large as 10 m in diameter and 25 m in length may be permissible for some restricted mission modes. The work shown here presumes maximum envelope dimensions of 7.6×16 m. Together, mass and volume constraints limit the size of individual pieces of process equipment able to be brought to the Moon. Anything larger requires assembly on the lunar surface.

B. Environmental Factors

The lunar surface is a difficult place to conduct operations. Many terrestrial industrial processes deal with, and in fact require, challenging physical and chemical environments. But the Moon's unique set of challenging conditions dominates all phases of industry, from emplacing equipment through mining, processing, production, storage, distribution, use and disposal. Accommodating the six major lunar environmental factors—gravity, radiation, vacuum, diurnal cycle, temperature and dust—will be a major technological accomplishment with extensive leverage for other planetary activities.

Lunar gravity is an equivocal complication. The acceleration and dynamic loads encountered during Earth-to-orbit launch set lower limits on the intrinsic strength and stiffness of equipment bound for the Moon. However, the Moon's 1/6 Earth-normal surface gravity does in general permit fully configured structures to be lighter and more slender than we are accustomed to. The presence of gravity is beneficial for stability in many assembly operations, and simplifies loading and gauging operations as well. However, lunar gravity is low enough that weight cannot be used as well as on Earth for stabilizing foundations and providing traction.

The two significant, naturally occurring kinds of radiation bombarding the lunar surface are energetic protons originating in solar flares, and the heavy ions of galactic cosmic radiation. These can disrupt electronic systems, interfering with control operations. Over time they can also degrade many engineered materials, especially polymeric substances.

The native lunar atmosphere is so tenuous (as low as 10^{-12} torr at night) as to constitute an extremely hard vacuum. Several complications follow for lunar systems, as they do for orbital space systems. First, there is no atmospheric scattering to temper the harshness of direct solar light. The contrast between sunlit surfaces and shadows is very high. This is a severe complication for intensity-based vision (especially human and saturation-prone machine systems). Second, permanent outgassing in hard vacuum changes the properties of many materials. Some may even become useless after only a short time, so designing long-life seals is a challenge. Third, heat rejection can only occur through radiation to cold space (the vacuum makes even lunar regolith an extremely poor thermal conductor). This complicates the design of motors and electronics. Finally, lubrication is a problem. Metals may cold-weld, greases turn to glue, liquids evaporate and intercalation fails. Space system mechanisms that are small and designed for limited use can skirt the lubrication issue through sacrificial lubricants. Long-lived, reliable, industrial scale operations, however, will require robust solutions.

The lunar synodic cycle is 29.53 days long. On the surface, sunlight is inescapable for two weeks at a time, and unavailable for another two week period, as the Moon turns. Reflected sunlight (Earthshine, peaking at 80 times brighter than a full Moon does on Earth) is available at nearside sites, but is not practical for power generation. The 15 day lunar night, and the required scale of surface operations, together drive selection of lunar power production and storage systems. The choice of space-beamed power, nuclear power, or solar power has profound implications for the types of processes practical for a lunar base: light-based, heat-based, or electricity-based; continuous or batch. The work presented here used solar-photovoltaic electricity with a daytime-only production schedule, and hydrogen/oxygen regenerable fuel cells for minimal "keep-alive" power at night.

The lunar vacuum and diurnal cycle together cause an extreme variation in surface temperature. At night the ground gets as cold as $-160°C$, and each day it reaches 110°C in sunlight. Insisting on continuous production

operations throughout this entire range is an unprecedented requirement for materials systems (in terms of dimensional stability and tolerance as well as sheer survivability), lubricants, seals, sensors, transducers and electronics. Heat-tolerant electronics, specialized and separate equipment for day and night, sheltered operations and warmed machinery are all possibilities—all with penalties in mass and mobile power. Lunar "trucks" will not be simple.

Lunar gravity, vacuum and temperature extremes represent opportunities for well-conceived processes, equipment and operations as much as they represent problems for poorly designed ones. However, the most consistently frustrating environmental constraint on lunar systems will probably be regolithic dust. Because the complete lunar dust problem is unprecedented in terrestrial operations, we discuss it in depth here. Fully half of the regolith by mass consists of particles finer than the human eye can resolve ($<70\,\mu$m). These particles are largely abrasive minerals, and in the dessicated lunar conditions are electrostatically "sticky." Even mesoscopic regolith particles clump together because of the jagged, interlocking shapes of their agglutinate component. From an engineering standpoint therefore, lunar soil will coat everything it touches, penetrate every crevice, and abrade every joint.

Some dust countermeasures are prophylactic. Critical components like photovoltaic surfaces and thermal radiators should be kept off the ground. Sources of lofted particles should be minimized (locally, by dust-stabilizing paving and slow locomotion) and intercepted (by debris shields surrounding landing pads). Exposed sensor lenses and windows may be protected by multiple-layer optical films, peeled away one at a time to remain clear. Electronics, pump units, and some motors and sensors can be hermetically sealed. Maintenance would be effected by replacing the entire unit robotically, and bringing the defective one inside a pressurized lab where it could be cleaned, opened, repaired and resealed by human crews. This scenario illustrates the complex interaction of people and machines required to make a productive lunar operation robust.

Other dust countermeasures are compensatory. Electrostatic precipitators may prove useful for periodic, robotic cleaning. But many gears and joints just cannot be kept dust-free, because they are either part of dust-handling systems, located under dust-shedding mechanisms, or even operated directly in the regolith. Mobility, mining and beneficiation equipment will suffer most. Openly configured mechanisms will let all but the inevitable dust film fall through, preventing macroscopic binding. Due to space transportation constraints, over-massive and over-powered approaches are less favorable than those based on advanced materials.

Treating critical bearing surfaces specifically for surface hardness will mitigate abrasive wear. Promising options include plasma deposition of diamond-like carbon and real diamond films on hard alloys. Appropriate joint and bearing designs will feature such specially treated alloys in robotically replaceable mechanism inserts. Rack-and-pinion drives may find wide usage for high-torque mechanisms requiring a large range of motion, as they

can be configured openly and actuated by electric motors. It may even be possible to develop practical, space-qualified, hydraulic actuators as well. The equipment concepts shown here are based on rack-and-pinion actuation.

Relatively long-lived, low-friction surfaces can be made using tough applied coatings (including outgas-resistant polymers like polyfluorotetraethylene) impregnated with chalcogenide compounds like MoS_2. Relubrication of dry-lubricant systems means part replacement, so critical surfaces treated in this way should be robotically replaceable inserts also.

In summary, industrial mechanisms optimized for lunar use will tend to have slender members with open, "knobby" joints consisting of standardized parts with replaceable, specially treated, active-surface inserts. Advances in lunar tribology (the science of friction, lubrication and wear) are enabling for reliable ISMU production operations. Simple adaptation of terrestrial methods and standards will probably not prove either technically successful or cost-effective.

C. Operations

By itself, the use of LLOX in ascent/descent vehicles has fairly low leverage on transportation operations requirements, because other architecture choices tend to be dominant drivers. For example, the problem of maintaining a low lunar orbit based lander during its orbital dormancy overshadows the problem of LLOX storage onboard during that interval. And for a surface-based scenario, the problem of long-term surface storage of liquid hydrogen dominates. Insuring reliable, available systems for crew use when needed has much more to do with propellant-combination, flight rate, mission-mode abort robustness and robotics capability than it has to do with whether or not LLOX is used. However, the use of LLOX does tend to drive a transportation architecture naturally toward surface basing. Surface-based retanking facilities would require extensive inspection, service and repair capabilities, robotic and human skills, and spares stockpiles at the base. Clearly this all complicates the surface base.

The most direct operations costs and benefits, however, derive from the surface mining, production and storage infrastructure required to run a LLOX industry. The data presented here are based on a thorough systems engineering study (Boeing 1990), whose intention was indeed to uncover operations requirements of (among other things) LLOX production, and to develop credible concepts for meeting them. Because of its prevalence in the ISMU literature at that time, the hydrogen reduction of ilmenite was chosen over other LLOX processes as the study reference. The results reflect not just the commonly quoted process parameters, but also the entire context within which that process would actually be embedded. Some top-level quantities are helpful for perspective. Equipment specifically associated with the LLOX industry approximately doubled the delivered mass of a human-tended, early base. About 190 tonne of equipment was required to support a LLOX production rate of 100 tonne yr^{-1}, for an equipment-to-product mass

ratio of 1.9 tonne per tonne yr^{-1}. Four-hundred twenty kWe was required, for a power-to-product-mass ratio of 4.2 kWe per tonne yr^{-1}. With all operations losses accounted for, the actual production yield, or ratio of product stored to regolith moved, was 0.6%. Nonetheless, the production of LLOX changed primarily the extent, not the type, of lunar base capabilities.

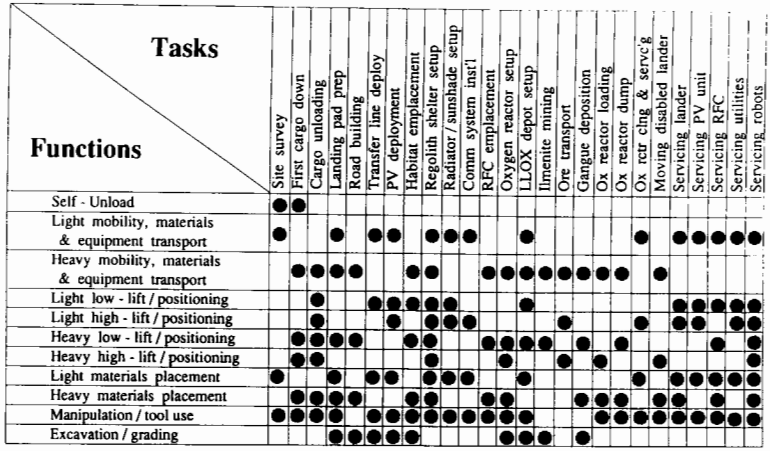


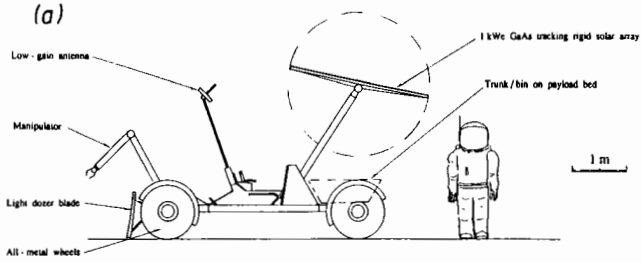
Figure 4. Range of robotic functions required to support lunar base tasks.

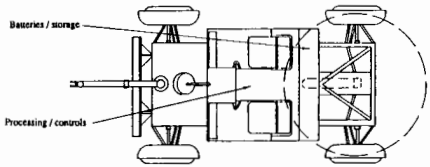
From the automation and robotics (A&R) standpoint, LLOX production introduces a quantitative increase in requirements, but not a qualitative upgrade. That is, as can be seen from the task-function matrix in Fig. 4, activities specifically associated with LLOX production do not introduce requirements for new types of tasks. Rather, they reinforce the already-present need for transportation and lifting of heavy payloads, and for the positioning and manipulation needed by inspection, adjustment and maintenance operations.

The A&R sophistication required to run the actual LLOX processing equipment is closer to current, terrestrial industrial capabilities than it is to that required for coordinating the action of mobile lunar base machines required regardless of LLOX production. Sequencing, loading, monitoring, cycling and unloading are all simpler robotic tasks than the planning, diagnosing, navigating, and manipulating capabilities required by other tasks around a lunar base.

All the base functions requiring complex mobile robots for their accomplishment can be accommodated by three classes of device, shown in Fig. 5a–e: light rovers, for autonomous ground surveys and suited crew transport; medium-weight utility trucks with a suite of special-purpose trailers and attachments, including a high-reach boom and manipulators; and large, straddling gantries with manipulators and excavation attachments. Redundancy requires a minimum of two of each for a practical base. Most of the

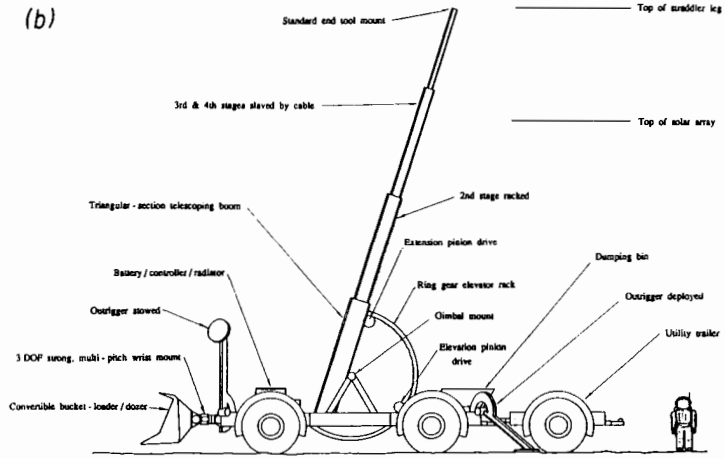
(a)



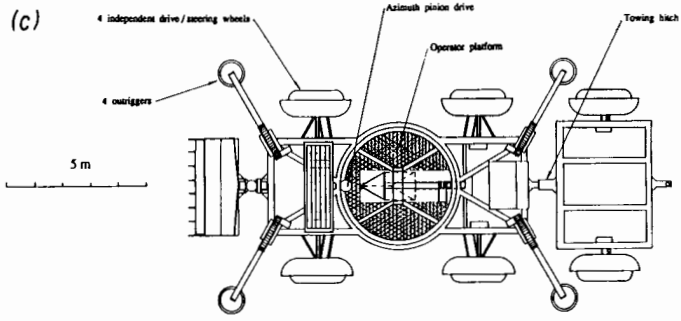


-- Based closely on successful LRV
-- Driven robotically or manually
-- Precursor surface & subsurface survey
-- Packages for Atlas-Centaur flight
-- 450 kg total

(b)



(c)



-- Operated manned or robotically
-- Reach envelope: from highest site element to below - grade
-- 6 kWe average power, 30 kWe peak
-- NaS batteries or RFC, 10 hr nominal charge
-- Can place payloads in pressurized workshop via garage
-- Towed: utility trailer, bucket - wheel excavator, vibrating compactor, lander LOX fill - line spool - cart
-- High - reach end tools: crew bucket, fine manipulator pair, sensors, rock grapple, hoist, small excavator - bucket, oxygen - reactor maintenance set, forklift

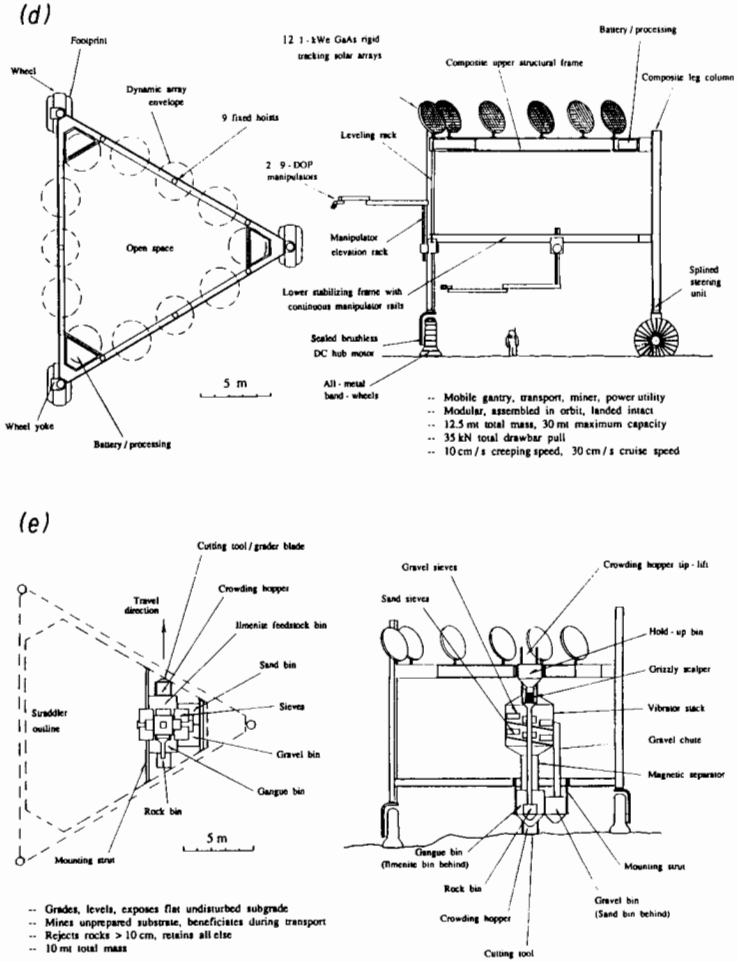


Figure 5. Three classes of mobile robot for lunar base functions: light rover, mid-weight utility truck with multiple end effectors and hitched accessories, and mobile gantry with mining unit.

material-transportation and manipulation complexity of a LLOX operation can be offloaded to these robots, keeping the process equipment itself comparatively simple. Even with their other duties, a generous allowance for downtime, and a feedstock-intensive process, these robots were found to be sufficient to enable 100 tonne yr^{-1} LLOX production rates.

Even for a small base which minimizes road-building and supports human crews only intermittently (Fig. 6), the largest operations driver—measured both by time expended and material processed—turns out not to be LLOX production. Rather it is base construction: the excavation of foundations for equipment siting, and the paving of roads and work areas for easier mobility and for dust control. The presence of the LLOX industry itself about doubles

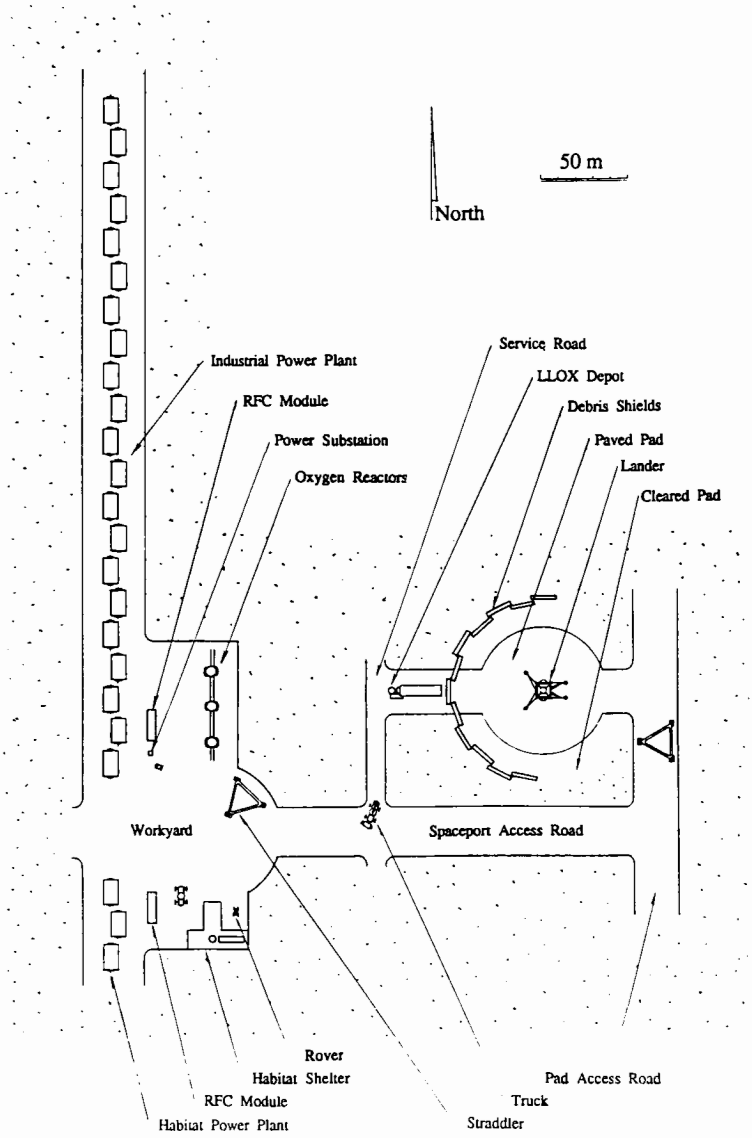


Figure 6. Small, solar-powered, oxygen-producing lunar base defined by integrated operations requirements.

the required construction time, and increases the paving material requirement by about half. If the vehicles are sized to support 100 tonne yr^{-1} LLOX production rates, a 2.5 yr stockpile of beneficiated LLOX feedstock (properly sized, 55% enriched ilmenite-bearing regolith) awaits processing by the time the base infrastructure is emplaced. Even modest expansion—more landing

areas and more habitation facilities—insures that construction operations will drive the excavation requirement for several years.

However, inspecting and repairing the LLOX plant would probably dominate the maintenance schedule of such a base. No LLOX plant concept is sufficiently detailed, nor do we know enough about the behavior of complex mechanical systems in the lunar environment, to let us make firm estimates of maintenance activity. An ilmenite plant concept was designed in sufficient detail to match the rest of the reference scenario, allowing determination of feedstock requirements as well as delivered mass and reliability estimates. But these data are conceptual only. The practical usability of regolithic ilmenite, techniques for keeping 10 atm of 900°C hydrogen from leaking past dusty seals, performance of large fluidized beds in 1/6 g, and the effect of dust on electrolyzer cells all require development work.

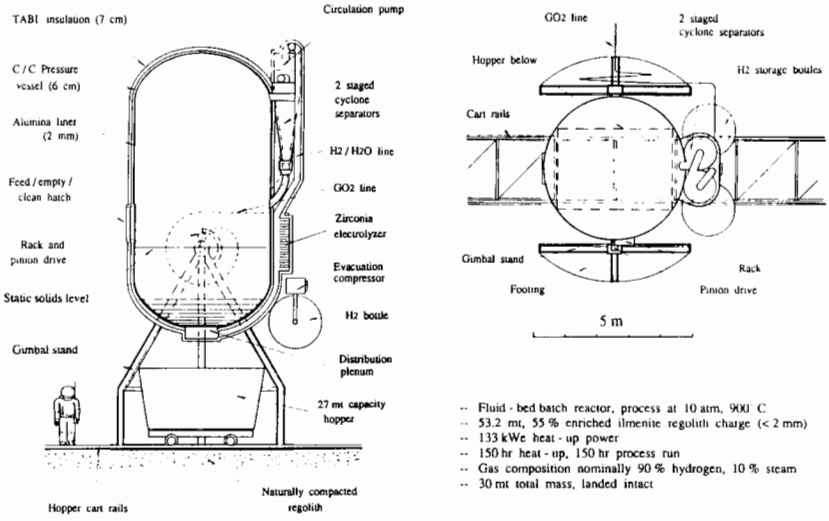


Figure 7. Single-stage fluidized bed ilmenite reduction reactor to produce oxygen.

A single-stage, batch-mode, fluidized-bed concept was developed (Fig. 7) which minimized the number of seals and moving parts, and configured them for easy replacement. No heat recovery would be used; cooldown would occur via passive radiation at night. Staged cyclone separators would remove most of the dust from the circulating gas, and solid-state zirconia cells would separate oxygen from the product water vapor. This oxygen would be piped through a thin, buried line to a liquefaction, storage and pumping depot (Fig. 8). Introduction of the feedstock, and removal of the slag, would occur through the same hatch. No provision was made for "pre-processing" of the feedstock grains to evert the ilmenite or remove sulfur. The rutile and iron-rich slag was assumed discarded. The plant design was kept extremely simple to enhance reliability even at the expense of power use,

process "efficiency," spares provisioning, hydrogen resupply and run-time. We estimated component reliability by adapting reliability data of analogous terrestrial components to lunar environmental conditions. The initial mean-time-between-failure calculated for this simple assemblage of subsystems was 5640 hr, a value sure to worsen with increased modeling fidelity.

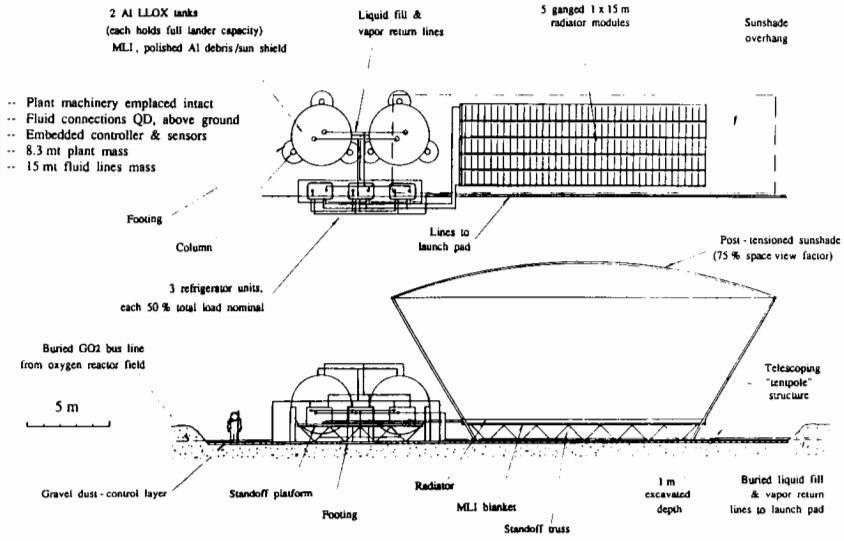


Figure 8. Oxygen liquefaction, storage and pumping facility.

In the development of a credible operations scenario, it is necessary to know where to buy important dividends at modest cost. Running a power-rich operation wastefully may be a good trade when compared to complex energy recovery methods, for instance. In the reference study case, emplacing a large solar power system (Fig. 9) was a complex but repetitive operation, judged easier than maintaining the more complicated LLOX plant design a smaller power capacity would force. Plant reliability was not the only issue; a more complex plant would have been heavier, which because of the 30 tonne lander capacity would have meant using four plants instead of three to meet the production rate. This would have added a flight to the manifest, as well as increased the number of separate plant subsystems requiring maintenance. The high mass cost of storing power overnight on the Moon (about 0.6 to 1 tonne kWe^{-1} using regenerable fuel cell systems like the one depicted in Fig. 10) constrained the scenario to daytime-only production, but the batch-mode operation used the dark time productively for passive cooldown and for equipment repair. A beamed-power scheme (spectrally matched photovoltaic collection of laser light aimed from power stations on Earth) would require less lunar array setup, could limit the extreme cooldown of equipment at night, and

might allow more-continuous production and therefore less process equipment for the same production rate. A nuclear power system would introduce these same benefits, but would require more site preparation for either shielding or distancing.

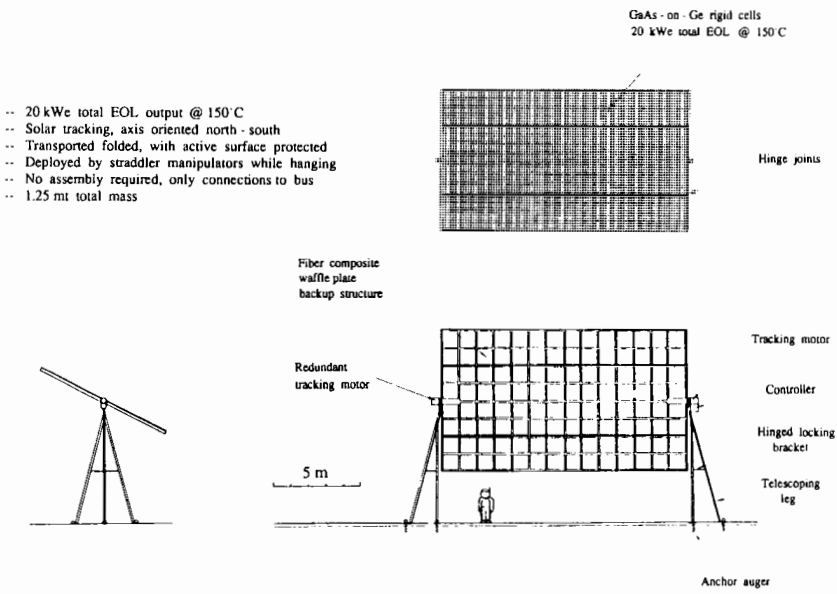


Figure 9. Rigid-panel, tracking photovoltaic power unit (24 deployed for 480 kW need.

The overall operations costs, not the process chemistry efficiency alone, nor the mass of catalyst required from Earth, determine how various candidates for LLOX production compare. A true understanding of operations comes only by examining the details. So far, an end-to-end scenario (including equipment concepts, base design, robotic control specifications, delivery manifests, buildup schedules, construction sequences, operations timelines, contingency analyses, and reliability calculations) has been generated for only one process, the hydrogen reduction of ilmenite. No real comparison is possible until similarly complete analyses have been done for other process candidates.

III. ECONOMICS OF LLOX PRODUCTION

The potential applications of local planetary resources are endless for a long-range space program. For this chapter, lunar oxygen (LLOX) for propulsion system use was selected as a vehicle for representative economics analysis for several reasons: (1) a general treatment, applicable to any and all uses would be vague; (2) the specific methods presented here are easily extrapolated

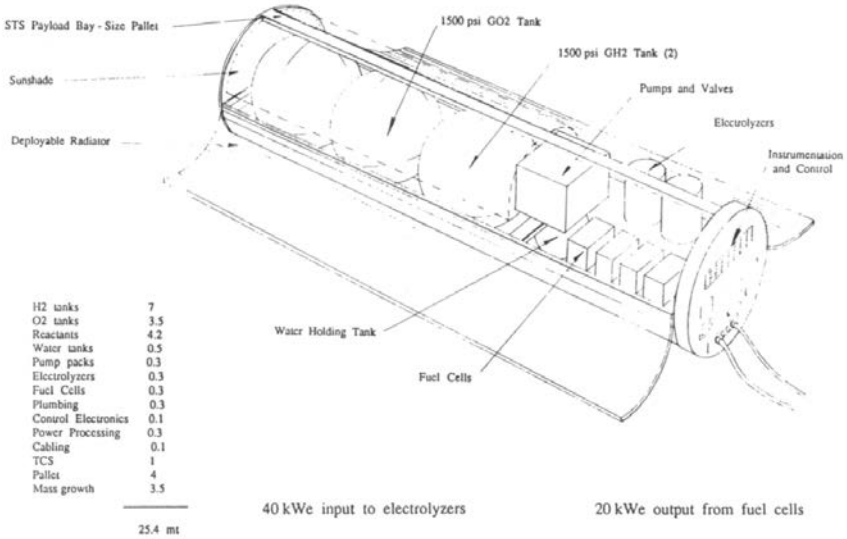


Figure 10. 20 kW regenerable fuel cell module for nighttime power continuity.

to other products; (3) lunar oxygen has a potential significant, early, and relatively quantifiable payoff; (4) lunar oxygen production technology is the subject of numerous ongoing research projects, so economics analysis has current relevancy; and (5) estimates of the technical characteristics of lunar oxygen production systems, while quite uncertain, are more certain than for most other products.

Economics analysis attempts to evaluate an investment in economic terms. It is not enough for an investment to "pay for itself." It must pay for itself with an adequate return on investment. The point is this: emplacing a production capability for LLOX will presumably cost more in the early years than not doing so. Having emplaced it, however, the operations cost for space transportation in later years will be lower. Because the investment occurs before the payback, the "cost of money" is central to economic feasibility.

There are two economic scenarios of interest. The first involves an investment which will have a payoff in reduced operating cost later on; the second involves lunar oxygen as one of several alternatives for improving lunar system performance, such that the development of lunar oxygen is traded against other developments in terms of the usual cost, risk, schedule, and performance criteria.

In this section, three example analyses are presented to describe and illustrate the basic methods of economics analysis. The first method (Sec. III.A) derives parametric equations. It involves major simplifying assumptions such as constant annual cost for elements of the program, and is best suited to evaluations of initial economic feasibility. The simplifying assumptions are necessary to permit integrals of discounted net present value cost streams

to be analytically evaluated. The loan payment equation is such an integral for constant annual cost and was used herein. Integrals are also easily obtained for ramp functions (constant annual increase or decrease) and, for that matter, any function that can be represented as a polynomial in time, e.g., $F = a + bt + ct^2$, etc.

The second example (Sec. III.B) simply extends the approach by numerically integrating to obtain the discounted net present value. Whereas the parametric equation approach usually permits directly solving for the discount rate that causes the overall discounted net present value to be zero (this discount rate is the internal rate of return), the numerical integration approach requires iteration to find the correct discount rate. This is a bit too tedious to do by hand but is readily done on a small computer and can, as a numerical integral is nothing more than a sum of things, be done with a spread sheet. The most straightforward spread sheet way to do the iteration is to write a suitable macro.

The third example (Sec. III.C) does not deal with discount rates and returns. It demonstrates the use of an input-output model to balance out a lunar base "economy" where many things are interdependent. Input-output models divide a lunar economy into sectors, such as electric power generation, that have inputs from other sectors such as operator time, maintenance, spares resupply, and thermal control, and produce an output needed by other sectors. The internetted system must have at least one external output such as lunar oxygen for transportation or crew time for science, and can have many outputs. A crude sketch of the idea that an input-output model can be dynamic (time-dependent) is also given.

All of these examples are simplified compared to what will ultimately be required, but the methods are applicable to more detailed analyses.

A. Parametric Equations

The basic economic potential of early utilization of lunar oxygen is analyzed, following a method presented in Woodcock (1989). LLOX supplied to the lander/ascent vehicle of a lunar transportation system is indicated as one of the highest-leverage uses of lunar resources because the oxygen is used at the point of production. The alternative to lunar oxygen is to supply it from Earth. It is generally true that the highest leverage for planetary resource production occurs when the resource is used at the point of production rather than requiring space transportation to the point of use.

In the early years of lunar operations, cargo Earth-to-orbit launch costs are expected to be about \$1000 per lb, perhaps more. Here \$1000 per lb is used as a representative value. (Space transportation cost is most commonly quoted in units of \$ per lb, although the total masses in the rest of the analysis are discussed in metric tonnes.) The cost to deliver cargo to the lunar surface will be about 10 times higher, mainly because about 6 tonnes of propellant must be delivered to low Earth orbit for every tonne of cargo delivered to the Moon.

The cost of oxygen made on the Moon includes amortization of the production equipment, support equipment and crew accommodations, and their transport to the Moon, plus the cost of operating the production system. The amortization factor is selected to include the cost of money as described below. Infrastructure and operations may be calculated separately, added together, and related to the cost of space transportation to low Earth orbit as follows.

The amortization equation estimates the cost of emplacing the production system and amortizes it at an annual rate that provides a reasonable return on investment. The total mass delivered per tonne per year of oxygen production is $NMa + M_h$ (symbols are defined below). The cost of launch is $C_t B$. To this is added the cost of owning and operating the lunar transfer vehicle (LTV) fleet. If it is reusable and low-Earth-orbit -based, the ratio of LTV fleet cost to launch cost is on the order of 0.25. Here the cost of launching propellant for the LTV fleet has been covered in the factor B. Thus the total transportation cost is given by $C_t B + C_t B(C_s/C_t)$ or $C_t B(1 + C_s/C_t)$. To this cost must be added the cost of producing the lunar production plant and the crew support facilities. This is done with a ratio C_h/C_t. Adding this and applying the annual amortization factor generates the following parametric equation:

$$\text{Amortization} = C_t A(NM_a + M_h)[C_h/C_t + B(1 + C_s/C_t)] \qquad (1)$$

where: C_t is the cost to low-Earth orbit, taken as \$1000 per lb; A is the annual amortization factor 0.2, discussed below; C_h/C_t is the ratio of production hardware cost to low-Earth orbit cost, taken as 2, i.e., production hardware costs \$2000 per lb to make; B is a "burden factor," ratio of total mass launched to mass delivered to the Moon, 7; C_s/C_t is the ratio of the cost of owning and operating the space transfer fleet to the cost to low-Earth orbit (a typical value is 0.25); N is the number of crew needed on the Moon per tonne yr^{-1} of lunar oxygen production, taken as nominally 0.02, M_a is the mass of crew accommodations in tonne per person, taken as nominally 10; and M_h is the mass of production hardware in tonne per tonne yr^{-1} of production, taken nominally as 1.5 (the point-design discussed in the last section resulted in $M_h = 1.9$).

A similar procedure results in the operations cost equation that follows:

$$\text{Operations cost} = C_t[(B'M_c/S + BM_r)N + BM_s] \qquad (2)$$

where B' is a "burden factor" for transport of crew cabs to the Moon; M_c is the mass of a crew cab in tonne per person, nominally 1; S is crew stay time in years, nominally 1; M_r is annual crew resupply in tonne per person-yr, nominally 2; and M_s is the annual resupply for the production system in tonne per tonne of production, nominally 0.1.

None of these figures is accurately known. The most uncertain ones were allowed to vary by a factor of 2 from the nominal values listed above to obtain

TABLE I

Parametric LLOX Production Cost Analysis Results[a]

Parameter	Nominal	Low			High		
	Value	Amort	Ops	Total	Amort	Ops	Total
1. Hardware/lunar cost ratio	2.0	3315	1475	4790	4335	1475	5810
2. STV/launch cost ratio	0.25	3358	1327	4685	4250	1770	6020
3. People per tonne	0.02	3440	1175	4615	4085	2075	6160
4. Crew accomm. tonne/per.	10	3440	1475	4915	4085	1475	5560
5. Prod. hdwe tonne/tonne	1.5	2043	1475	3518	6880	1475	8355
6. Crew stay time, yr	1	3655	1350	5005	3655	1725	5380
7. Crew resupply, tonne/per.	2	3655	1300	4955	3655	1825	5480
8. Prod. resupply, tonne/tonne	0.1	3655	1038	4693	3655	2350	6005

[a] Nominal cost: Amort—3655, Ops—1475, Total—5130. Values are in dollars per pound of oxygen produced.

an indication of cost sensitivity. Results are summarized in Table I and Fig. 11. The nominal cost for LLOX production of $5130 per lb amounts to about half the cost of shipping the oxygen from Earth instead. Lunar oxygen production at a level suitable for use in lunar lander/ascent systems for a lunar base thus appears to be a good investment. The sensitivities are modest, except for the mass of production hardware.

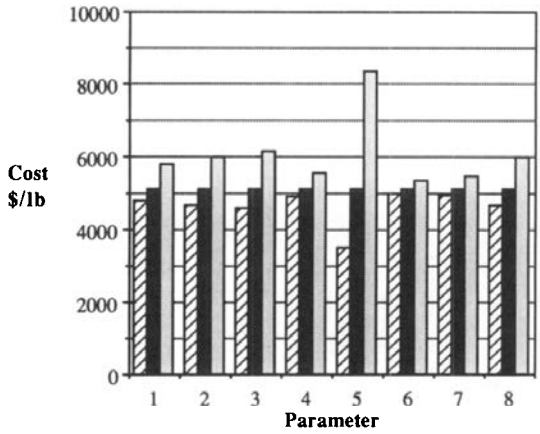


Figure 11. Parametric LLOX cost sensitivities (see Table I for parameter descriptions).

More than 2/3 of the nominal cost is amortization. This calls into question the value used for amortization. It is based on a 12% cost of money, and 10 yr to recover the investment (the actual value of A for these conditions is 0.18; it was rounded up to 0.2 for the analysis). A commercial venturer would want to recover costs sooner, with a return perhaps as high as 0.4 annually. Nominal LLOX production costs for higher amortization are: $6957 per lb for $A = 0.3$; $8785 per lb for $A = 0.4$. These are close enough to the cost for delivery from Earth to make the realistic financing of an initial lunar oxygen production system by venture capital doubtful. $A = 0.2$ is, however, adequate for government financing. Ventures that entail significant risk but have long-range and far-reaching benefits, such as the development of space resources, are most appropriate for government financing. Such ventures as the Tennessee Valley Authority could not have been privately financed, but have had enormous long-range social and economic benefits. An early lunar base operation will probably be government funded anyway. Hence, lunar oxygen production is indicated as a desirable and economically beneficial option.

The parametric estimate for production hardware is based on, but assumes a somewhat less feedstock-intensive process than, the point-design integrated system discussed in the last section. A large part of total plant mass is the power system. If reliable nuclear power were available, plant mass could be

considerably less. The total cost could well be more, depending on the cost and life of space nuclear power systems. For example, if the development cost for a 300 to 500 kWe nuclear powerplant is 10^9 dollars, this is equivalent to the delivery cost for 50 tonne to the lunar surface. Nuclear plants would have to be replaced every several years, and more than one might be required for assurance of power availability.

Another potential source of lunar surface power is laser beaming from Earth (Rather 1991). Like nuclear power, this would provide a continuous power supply. The lunar surface power system mass can be less than for the solar array plant used as a reference, because the power per unit array area is up to several times greater. Its life would be indefinite. The leverage on oxygen production cost is significant: a factor of 2 on oxygen plant delivery mass because of continuous operation independent of diurnal cycle, and a reduction in the lunar surface powerplant mass. If laser beam power were available for lunar use, a commercially financed initial lunar oxygen production system would probably be feasible.

With lunar oxygen production in place, the "burden" values B and B' both drop to $\sim$5. This improves the economics for continued investment in lunar resources development.

B. Program Life-Cycle Return-on-Investment Analysis

While the cost of money can become an involved subject, the life-cycle return on investment issue may be simply stated: is it best for the government to avoid a particular investment cost and thereby avoid the interest cost for financing that investment (remember, the government is running a deficit), or does the investment have enough payoff to recover the interest cost and more? To make this determination, we must evaluate scenarios both with and without the investment, and compare them in a return-on-investment analysis.

The method described here is apropos to scenario evaluation where annual funding requirements have been estimated in detail. Figure 12 defines net present value and diagrams a typical problem. One funding stream (the "test" case) incurs added early costs to improve an operation so that the operating costs downstream can be reduced. The question is whether this is a wise investment. The answer is obtained by computing the return on investment (sometimes called "internal rate of return" because the alternative funding scenarios are internal to the problem at hand, and do not involve external investments) as illustrated in the figure. The calculation is iterative, and proceeds by adjusting the discount rate until the net present value of the difference between the streams is zero. If that resulting discount rate (in percent) is greater than the cost of money, the investment is worthwhile.

It is important to be consistent in the treatment of inflation. Cost analyses are commonly done in constant dollars, such as 1991 dollars. In constant dollars, the cost of money is the inflationary cost of money (actual borrowing costs in the real economy) less the annual average rate of inflation in percent. That is, if the cost analyses used for return-on-investment calculations were

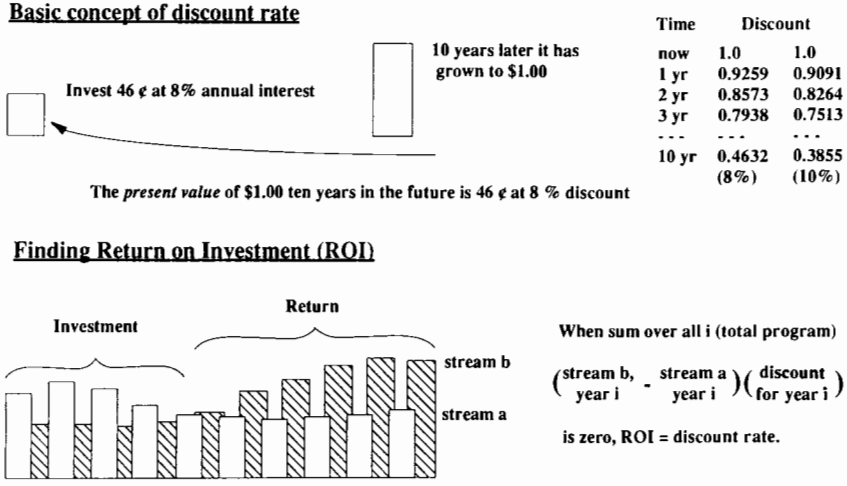


Figure 12. Net present value and return on investment

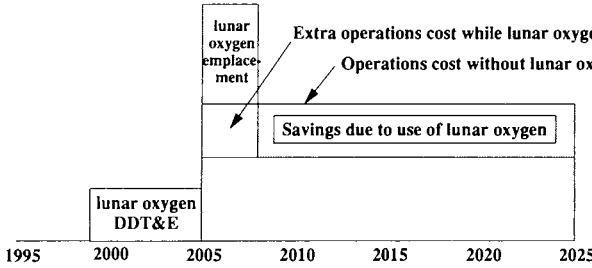
adjusted for inflation, the computed internal rate of return would be greater by the rate of inflation.

It is also important to recognize that government agencies such as NASA do not pay interest on their invested funds; interest on the national debt is a separate budget item handled by the U. S. Treasury. If NASA considers the investment value of a capability (like lunar oxygen production) based on the NASA budget, the break-even discount rate is about zero and can even be negative. For example, if a lunar program is budgeted at a fixed annual level in then-year dollars, the real annual value of this funding stream declines with inflation and the apparent cost of money to the program is negative.

A simple investment analysis example for lunar oxygen is presented in Fig. 13. The cost figures are hypothetical; they are not derived from actual cost analyses for lunar oxygen production systems. Investments and savings assumptions are shown in the figure. Investments include both the design, development, test and engineering costs (DDT&E) for lunar oxygen production including any dedicated support equipment, and emplacement of the production system on the Moon. Savings result from reduced space transportation cost.

At low utilization rates, the research and development cost for the oxygen production system dominates the investment analysis. In this analysis assumptions were made independent of the size of the production system. While this is excessively conservative, much of the engineering design, development and test cost will tend to be independent of plant size.

Figure 14 shows the estimated return on investment for lunar oxygen, for three lunar flight rates, which is not attractive at low flight rates (unless zero or negative cost of money is considered). The assumed size of the oxygen



Cost values

- Lunar oxygen DDT&E $550M for 6 years
- Lunar oxygen emplacement 3 cargo deliveries
  - 2 HLV each @ $300M
  - 1 expended LTV @ $300M
  - 1 reused LTV @ $50M
  - Space operations $100M
- Sensitivity case: HLV cost reduced by half

- Operations for 1 lunar mission:
  - $122M for lunar transportation hardware
  - $300M per HLV flight
  - 2 HLV flights per lunar mission without lunar oxygen.
  - 1 HLV flight per lunar mission with lunar oxygen
- Examined 1, 2, and 4 lunar missions per year.

Figure 13. Simple investment analysis example for LLOX.

production system in this case study was enough to support four lunar lander trips per year, corresponding to the 100 tonne yr^{-1} point-design example of the last section. At this moderate flight rate the return on investment does become attractive. This example is conservative in that the cost of emplacing the plant was assumed independent of size, whereas smaller plants would certainly cost less to emplace (but probably not much less to develop).

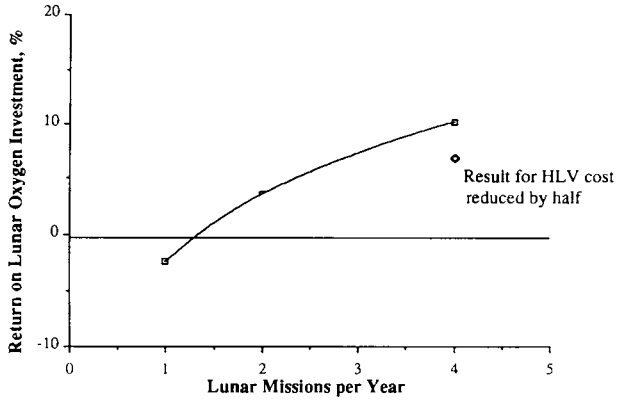


Figure 14. Return on investment for lunar oxygen example, parameterized by flight rate.

The example presented here should not be used to derive conclusions as to the economic benefit of lunar oxygen because the input data were

assumed, not based on actual estimates. In particular, the costs to develop and emplace the lunar oxygen system were not varied with scale of use. Also, the only benefit credited was a reduction in operating cost. Because lunar oxygen can stretch the performance of the lunar transportation system, it is quite likely that its timely introduction will avert other development costs for upgrading the system such as development of an aerobrake for lunar transfer vehicle recovery. Elementary performance calculations show that lunar oxygen, used for return from the Moon and propulsive capture in low Earth orbit, has greater leverage than aerobraking for upgrading performance enough to achieve reusable operations.

C. Input-Output Analysis

The first, parametric analysis method example reflected some of the operations and support costs for a lunar oxygen production system; the second, life-cycle cost method example did not. Program life-cycle cost analysis can include these costs if properly identified. As all of the lunar operations are interdependent, accurate estimates of total costs can only be obtained through an input-output analysis. The examples given here are very simple, with only about ten sectors comprising the lunar economy. More thorough analyses would characterize from dozens to about a hundred sectors. Because the characterization of each sector includes quantitative estimates of its interactions with other sectors, a thorough analysis requires a major study effort to obtain quantitative estimates for hundreds of input-output coefficients.

This example illustrates some potential benefits of space settlements. There is as yet no formal distinction between the terms settlement and base in the context of space exploration and development. It is necessary for the purpose here to define the difference, so the following are adopted.

Base implies: (1) Human crews stay from six months to two years; (2) life support technology ranges from physico-chemical closure on water and oxygen to ecological closure on water, oxygen, and food; (3) most facilities are delivered from Earth.

Settlement implies (1) Human crews stay from two years to indefinitely; (2) life support is ecologically closed; (3) facilities are largely to mainly constructed from indigenous materials.

To get an idea of the benefits of lunar settlement as compared to lunar basing, one can perform a simple mass-oriented input-output analysis. The key parameters for a base requiring resupply from Earth are listed in Fig. 15, which also shows the full input-output matrix and identifies its nonzero values. Values in the matrix are simply the coefficients of a set of linear algebraic equations in terms of the variables listed (Sherwood 1992). For example, the first equation simply sets the mass of the production plant equal to 150 tonne. (Again, these results are based on a LLOX production rate of 100 tonne yr^{-1}.) Simultaneous solution of the equations yields values for the variables with all identified interrelationships taken into account.

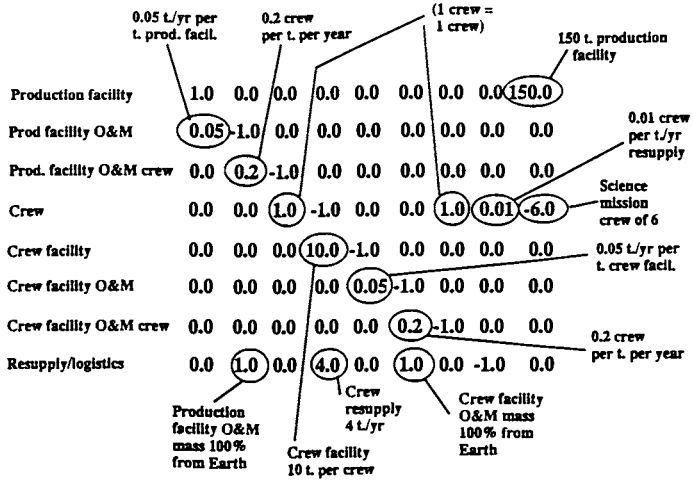


Figure 15. Input-output matrix for resupplied lunar base.

Coefficients (for example the annual 5% spares rate) related to operations and maintenance (O&M) were taken from representative aerospace experience. Space Station *Freedom* is using a higher figure than this, but it is presumed that technology maturation will limit the spares resupply requirement for a planetary base. Typical aerospace experience (for example on shuttle flights) is about four failures per day. Assuming an average of 4 hr per maintenance action and standard 8 hr shifts, two people are required dedicated full-time to maintenance on a system as complex as shuttle or space station. The coefficient of 0.2 people allocated per tonne of remove-and-replace equipment per year (O&M), multiplied by 5% spares and a 100 tonne system yields a more optimistic average of just one person. Again, a necessary assumption is improvement in both the technology and design for maintenance.

For the self-sufficient settlement case, it is assumed that 90% of removed equipment can be repaired at a lunar workshop for replacement and further use but that this doubles the maintenance labor per tonne of remove-and-replace hardware. Spares resupply from Earth is greatly reduced, to just 10% of the remove-and-replace mass in our example. Field repair should be able to restore circuit cards by replacing failed electronic components, and electromechanical devices by the adjustment, decontamination, or replacement of individual components such as drive motors and switches. However, the actual repair labor burden doubles the maintenance time devoted to each failed item, because shopwork is required as well as the remove-and-replace activity.

The other significant item in the self-sufficient system is ecological (biological) closure of life support supplanting electrochemical closure. This has two benefits: (1) importation of food is essentially eliminated; and (2) com-

plex, high-tech electrochemical equipment is replaced by plants, the means for their efficient growth and processing, and the means for recycling agricultural waste. On-site repair and service of the latter systems appear much more feasible. Preliminary studies of plant growth and processing needs for ecological closure suggest that a doubling of the crew-support facilities mass and volume per person is roughly correct. Present conceptual design estimates, supported by exploratory experimentation, are considerably more optimistic than this. Doubling seems very conservative. The input-output matrix for the self-sufficient case is shown in Fig. 16.

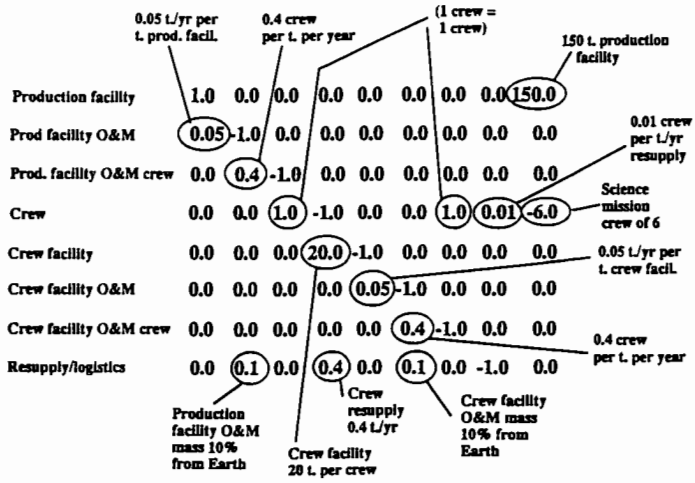


Figure 16. Input-output matrix for self-sufficient lunar settlement.

Table II compares the input-output model results for the two cases. In the "base" example, a total lunar crew of 18 can provide 100 tonne yr^{-1} of oxygen with an average of 6 people devoted to exploration and science. The base resupply requirement is about 90 tonne yr^{-1}. Roughly half of the resupply supports the people and facilities that do oxygen production, and the oxygen cost is about half the cost for delivery from Earth, as was the case in the first parametric example.

In the "settlement" case, the resupply requirement is reduced by about a factor of 6, and lunar oxygen production is economically attractive. However, the facilities needed on the Moon are roughly tripled. Whereas the base requires about 4 yr for delivery of the base facilities at the resupply rate, the settlement would require nearly 50 yr. This shows that, to be viable, a settlement scenario must include the local construction of its crew-support and production facilities *as well as* the production of foodstuffs typically described in closed ecological life support scenarios. The question of rates of growth from indigenous resources is central.

Modifying the input-output model to include internal growth is straightforward. An annual growth rate is selected. This determines a rate for construction of facilities. In this simplified model, annual growth rates for crew-support and production facilities are assumed equal. The annual growth rate determines an annual facilities-mass addition rate. Crew requirements, measured in people per tonne yr^{-1}, were nominally assumed the same as for remove-and-replace activity: 0.2 people per tonne yr^{-1}. Variations were investigated: crew requirements for facilities construction between 0.1 and 0.4 people per tonne yr^{-1}, and annual growth rates between 5 and 20%. For a nominal oxygen production rate of 100 tonne yr^{-1}, various solutions are shown in Table III.

TABLE II
Comparison of Input-Output Model Results for Lunar Base and Settlement Cases

Parameter	Resupplied Base Case	Self-Sufficient Settlement Case
Production facility mass (tonne)	150.0	150.0
Production facility O&M (tonne)	7.5	7.5
Crew for prod. fac. O&M	1.5	3.0
Total crew	17.63 (18)	27.74 (28)
Crew-support facility mass (tonne)	176.3	554.9
Crew-support facility O&M (tonne)	8.8	27.7
Crew for crew-sup. fac. O&M	1.8	11.1
Resupply mass (tonne yr^{-1})	86.8	14.6

From these results it is clear that the lunar crew size, and in fact the feasibility of the entire settlement enterprise, is critically dependent on the efficiency of construction and the desired growth rate. Early LLOX production at about 100 tonne yr^{-1} rates can be emplaced using facilities brought entirely from Earth if the operations and maintenance workload can be kept low, e.g., by robotics and teleoperation.

More ambitious activities need indigenous resources for facilities production. For example, the University of Wisconsin ^{3}He mining scenario (Kulcinski et al. 1992; NASA 1989) estimates that collection of enough solar-wind-implanted ^{3}He from lunar regolith for 1 GW of electrical generation in terrestrial fusion power plants would require about 100 tonne of facilities on the Moon, similar to the amount for early lunar oxygen production.

An input-output analysis was performed scaled for lunar surface operations to produce sufficient lunar ^{3}He to generate 300 GW-yr per year—most of the present U. S. demand for electrical generation. Annual growth from an early industrial capacity to the level needed for ^{3}He production—or for other major lunar-based manufacturing operations—must be on the order of 20% per year to achieve the needed growth. A viable ^{3}He scenario would typically

TABLE III
Effects of Varying Settlement Growth Parameters

Construction crew per tonne yr^{-1}	0.4	0.2	0.2	0.1
Growth rate (%)	5	5	10	20
Lunar crew	101	28	46	44
Total facilities mass (tonne)	2162	705	1064	1037

increase from 1 GW in 2020 to 250 GW in 2050, resulting in 25,000 tonne of facilities and 5000 people on the Moon.

Twenty percent annual growth is unachievable without a *very* high rate of productivity. 0.1 crew per tonne yr^{-1} of construction and O&M equipment yields favorable results in a simple input-output model, but this is about ten times the productivity in man-years-per-unit-mass-of-hardware achieved on commercial jetliner production lines. Attaining such high productivity is likely to require advanced automation and robotics; human crews would mainly supervise and maintain a robotic "labor force" directly involved in mining, manufacturing and production operations.

IV. EVOLUTION OF INDIGENOUS SPACE MATERIALS UTILIZATION (ISMU) CAPABILITY

LLOX production is commonly held to be the "first" ISMU likely to be emplaced. The previous sections show that this claim is neither technically correct nor economically secure. In fact, to support either human habitation or any production-rate ISMU activities, sitework construction will be required first. Excavating, sieving, separating, and consolidating lunar regolith is the first set of ISMU problems to be solved. Configuring foundations, work areas, lasting roads and stabilized landing pads comes next. These are all preconditions for reliable, repeated base operations of any kind, regardless of the presence of any industrial production, and their difficulty should not be underestimated. Next in complexity are resources yielding a chemical which is the final product (like LLOX and other volatiles). Finally come industrial products requiring further fabrication processing (like metal and glass construction materials).

As we learn more about the operations complexities associated with various ISMU processes for making LLOX, enabling direct comparison of their costs, we will be able also to compare their synergetic benefits. Just as a less efficient but simpler process may win out over a more efficient one, a more complex process may win out over a simpler one if it produces more than just oxygen, or if its energy usage and "waste" production integrate better into an overall base scenario. At the present time, understanding of these synergies is far from sufficient to justify rejecting or endorsing particular process candidates.

To successfully achieve lunar industrialization, many lessons must be learned about how to make mining, process and factory machinery work on

the Moon. LLOX production shares some of those lessons with many other ISMU activities, and results in a simple product of uncontested usefulness in any lunar operations. The argument may be made that a marginal return of investment for modest-scale LLOX production is moot, given an assumption of indefinite growth in lunar capability. However, if the most important, far-reaching benefits are really the ISMU lessons learned, it becomes important to see if other ISMU activities have a better or quicker return. Not enough is known yet. A strong argument can be made that the greatest leverage of all can come from planning an early lunar base for flexible experimentation with many prototype processes, allowing the results of lunar science and operations experience to guide more focused selections.

REFERENCES

The Boeing Company. 1990. Robotic Lunar Surface Operations: Engineering Analysis for the Design, Emplacement, Checkout and Performance of Robotic Lunar Surface Systems. Work performed for NASA Ames Research Center under Contract NAS2-12108, Boeing D615–11901.

The Boeing Company. 1991. *Space Transfer Concepts and Analysis for Exploration Missions, Phase I Final Report.* Work performed for NASA Marshall Space Flight Center under Contract NAS8-37857, Boeing D615-10030-2.

Kulcinski, G. L., Cameron, E. N., Santarius, J. F., Sviatoslavky, I. N., Wittenberg, L. J., and Schmitt, H. H. 1992. Fusion energy from the Moon for the twenty-first century. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 2, pp. 459–474.

NASA. 1989. *Report of NASA Lunar Energy Enterprise Case Study Task Force*, NASA TM-101652.

Rather, J. 1991. Selene concept for terrestrial-based lunar surface power. Presented at *Technology Workshop on Laser Beamed Power*, Feb. 5, Cleveland, Oh., NASA Lewis Research Center.

Sherwood, B. 1992. Lunar materials processing system integration. Lunar Materials Technology: Proc. Third Annual Symp. UA/NASA SERC, Feb. 20–22, Tucson, Ariz., Abstract book, pp. III-67–III-79.

Woodcock, G. R. 1989. Parametric analysis of lunar resources for space energy systems. In *Space Manufacturing 7, Proc. of the Ninth Princeton/AIAA/SSI Conf.: Space Resources to Improve Life on Earth*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 109–121.

MISSION AND TRANSPORTATION SYSTEM APPLICATIONS OF IN-SITU-DERIVED PROPELLANTS

E. REPIC, R. WALDRON, W. McCLURE and H. WOO
Rockwell International Corporation

This chapter discusses the mission and transportation applications of *in-situ*-derived propellants derived from lunar materials. First, a brief summary of available materials and associated propellants is presented. Details of the processing of lunar propellants has been given in earlier chapters. Next, some typical lunar transportation vehicle concepts are discussed to show the types of chemically fueled designs being considered today and the benefits of using *in-situ*-derived propellants. Finally, time-phased mission applications are presented and some mass and cost trades discussed. The results clearly show that the benefits obtainable from the use of *in-situ*-derived propellants for lunar and Mars operations are considerable if Earth to orbit launch costs stay as high as they are now. If a factor of 10 reduction in these launch costs is forthcoming (the goal of the Advanced Launch System and its successors), then the benefits of using lunar propellants becomes minimal.

I. AVAILABLE LUNAR PROPELLANTS

The principal compositional constraints on propellants solely or predominantly derived from lunar materials are the general lack of significant amounts of major constituents of volatiles or fluid compounds other than oxygen from lunar sources. There are only 7 elements commonly found at levels exceeding 1% by weight. Table I shows the mean compositions of lunar soils found in samples gathered during the Apollo 12 and 16 missions representing typical mare and highland sites, respectively (NASA/JSC 1989). Six additional elements may be classified as minor constituents (occurring at levels between 0.1 and 1.0% by weight) of which sulfur, sodium and potassium are potentially extractable by purely thermal methods and could be considered for propellant manufacture. The accessible lunar elements can be supplemented by small import fractions of deficient lunar elements primarily as propellants.

Candidate propellant systems for partial or total *in-situ* lunar manufacture may be classified according to fuel or oxidizer storability as cryogenic, soft cryogenic or fully storable or according to the extent of import mass, if any, required for manufacture or use. Soft cryogenics such as silanes and even oxygen can be stored indefinitely on the lunar surface or in lunar orbit using geometrically shielded, high-technology passive cooling devices in

conjunction with high-performance insulation. Hydrogen is probably the only propellant requiring re-liquefaction or boiloff management for storage.

TABLE I
Compositions of Lunar Soil From Apollo 12 and 16 Missions[a]

Component	Weight Percent	
	Apollo 12	Apollo 16
Al_2O_3	13.71	27.18
CaO	10.55	15.79
FeO	15.41	5.18
MgO	9.91	5.84
SiO_2	46.17	45.09
TiO_2	3.07	0.56
Na_2O	0.48	0.47
K_2O	0.27	0.11
	Parts Per Million	
H	45	56
F	132	72
N	84	89
Na	3600	3500
S	1000	640

[a] Table from NASA/JSC 1989.

Table II presents a matrix of candidate propellant systems classified according to storability and extent of lunar versus import mass accessibility along with approximate specific impulse values at optimum mixture ratios. The maximum specific impulse (I_{sp}) values may not represent the most favorable overall conditions for propellant use, however. Because for many cases, the true cost of fuel may be substantially higher than that of oxidizer, lunar liquid oxygen, (LLOX), total import mass could be reduced by operating at fuel-lean (off optimum) conditions. Of the candidate systems shown in Table II, the most promising include LH_2/LLOX, SiH_4/LLOX and SiH_4/H_2O_2. The silane-hydrogen peroxide system containing the soft cryogen SiH_4, can be considered indefinitely storable with modest technical requirements. Of those propellants solely derivable from lunar materials, Al/LLOX and Mg/LLOX are probably the best selections.

TABLE II
Candidate Propellant Systems Properties

A. Solely Derivable From Lunar Sources		
Storability	Propellant System	Specific Impulse (s)
Cryogenics	Al/LLOX	280–300
	Mg/LLOX	260–290
	Mg(NaK)/LLOX	250–280
Storables	Mg, CrO_3	Low
	Mg(NaK), KO_2, CaO_2	Low
B. Derivable From Lunar Sources Plus Earth Imports		
Storability	Propellant System	Specific Impulse (s)
Cryogenics	Earth LH_2/LLOX	460–480
	SiH_4/LLOX	380–400
	Si_2H_6/LLOX	380–400
	AlH_3/LLOX (Hybrid)	395–415
Soft Cryogenics	SiH_4/H_2O_2	365–385
	SiH_4/H_2O_2	370–380
Storables	H_2O_2 (Monoprop.) 165–180	

II. PROPERTIES AND PRODUCTION METHODS

A. Cryogenic Volatiles

Table III lists the physical properties of candidate cryogenic fuels and oxidizers. Storage at temperatures between the triple points and critical points is normally required. Higher pressures require heavier tankage but some systems could use pressure-fed engines rather than pump-fed, thus simplifying design and operations.

Oxygen. This element is the most abundant in lunar soils and rocks, typically constituting about 40% by weight. It has relatively high fluid density and gives the highest specific impulse obtainable with most fuels. A number of processing methods for production from lunar sources have been proposed and are described in earlier chapters.

Hydrogen. This element is present in only trace amounts in lunar soils, generally below 100 ppm by weight, and is virtually completely absent from lunar rocks. Recovery on a modest scale is possible in principle, but would require mining ore in excess of 20,000 times the mass of the recoverable hydrogen and would require an enormous expenditure of energy. Hydrogen and oxygen provide the highest specific impulse of all potential propellant combinations, even when operating at off-optimum mixture ratios, so that the import fraction of hydrogen can be minimized.

Silane. The Silanes are silicon hydrogen compounds analogous to hydrocarbons. The first member, SiH_4, has physical properties somewhat similar to methane, CH_4, but with appreciably higher triple and critical points. The spe-

TABLE III
Physical Properties of Cryogens

Cryogen	Triple Point			Boiling Point			Critical Point		
	T(K)	P(bar)	D(1)	T(K)	P(bar)	D	T(K)	P(bar)	D(g cm^{-3})
Oxygen	54.4	0.0015	1.31	90.2	1.01	1.14	154.6	50.4	0.436
Hydrogen	13.8	0.0704	0.077	20.27	1.01	0.071	32.94	12.84	0.031
SiH_4	88	0.0002	68.71	161.6	1.01	0.6	269.7	48.4	0.242

cific impulse of SiH_4/LLOX propellants is lower for any given mixture ratio than LH_2/LLOX but silane is only 12.5% by weight hydrogen, so manufacture of the fuel from lunar silicon and imported hydrogen becomes feasible.

Metallic Solids. Rocket propulsion can be provided by combustion of reactive metals, such as aluminum and magnesium, with LLOX and using the released heat to expand excess oxygen through a suitable nozzle. This can provide propellant systems derived totally from lunar materials. Of the two metals, magnesium is easier to recover from lunar soil. However, the specific impulse of Mg/LLOX is less than for Al/LLOX and the density is less so the volumetric requirement for Mg is twice as high as for Al per unit mass of LLOX.

Although it is entirely feasible to derive propellants totally from lunar sources, LH_2/LLOX remains the propellant combination of choice for most applications, at least for the lunar base scenarios studied thus far. If a scenario is chosen in the future which contains the requirement for a permanent base, then combinations such as silane/LLOX, which requires only a small import mass of hydrogen, and aluminum or magnesium and LLOX become of interest.

III. CONCEPTS FOR A LUNAR MISSION VEHICLE

Option studies of lunar mission scenarios and vehicle conceptual designs that can use lunar *in-situ*-derived propellants are based on requirements for the Lunar Outpost Mission defined in the 90 Day Study (NASA 1989). Lunar transfer is performed by two types of common vehicles: the lunar transfer vehicle (LTV) and the lunar excursion vehicle (LEV). Each vehicle has a cargo-only configuration and a piloted configuration that adds a crew module for personnel transfer to the common vehicle.

The LTV provides for transport between low Earth orbit (LEO) and low lunar orbit (LLO). The piloted LTV is serviced in the LEO node for reuse, while the cargo LTV is expended. The LEV is reusable and provides for ascent descent transportation between LLO and the lunar surface. The LEV can either be loaded with Earth propellants by the LTV in LLO or on the lunar surface with the *in-situ* propellants. Based on analyses of surface system requirements, the LTV and LEV are required to deliver at least 33 tons of cargo to the lunar surface in a single cargo-only flight and at least 16 tons of cargo (plus crew module) in a single piloted flight. Also required are a crew size of at least 4 and a return cargo of at least 0.5 tons on piloted missions. The ΔVs and sizing parameters assumed for the lunar mission are shown in Table IV.

A. Design for a Lunar Transfer Vehicle

Chemical propulsion using liquid oxygen and liquid hydrogen (LOX/LH_2) is the reference LTV concept for *in-situ* lunar-derived propellant utilization. A reusable, chemical LTV will require significant mass in LEO, a large

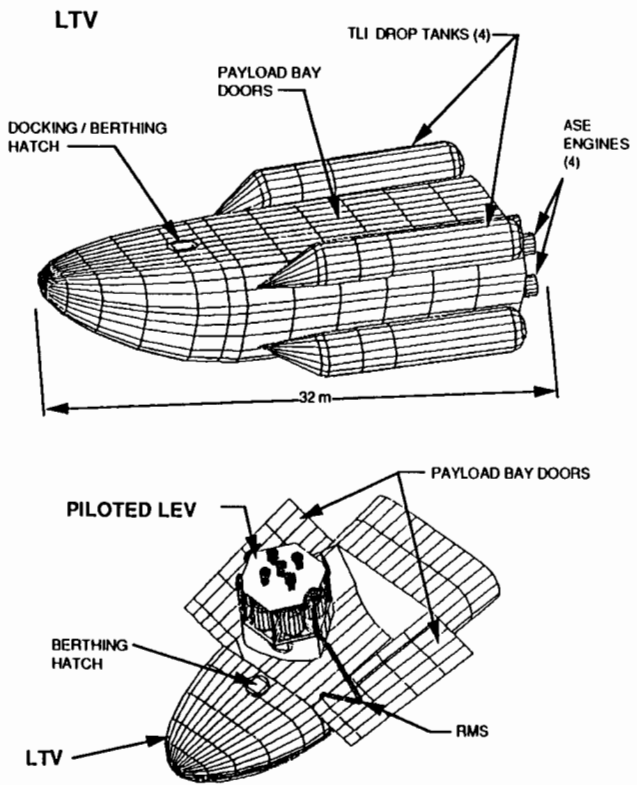
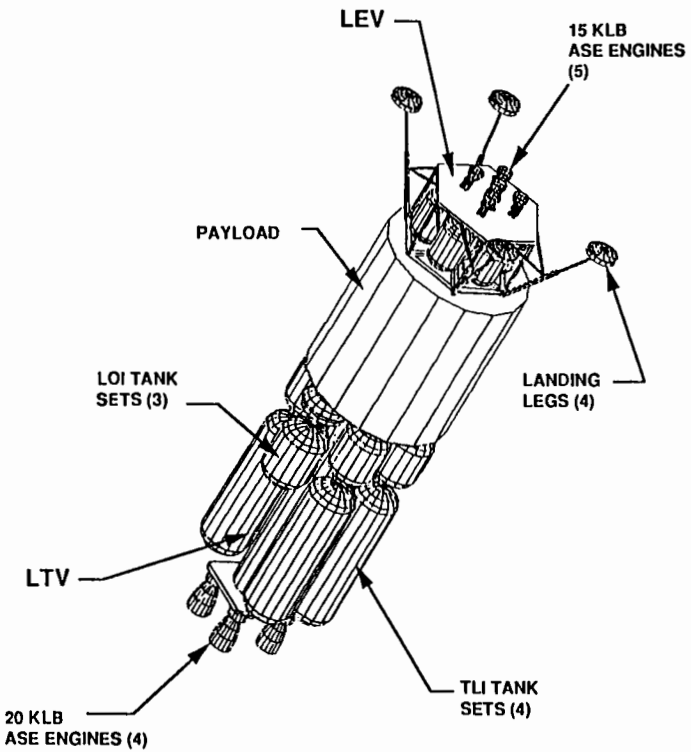


Figure 1. Piloted lunar transfer vehicle.

Component	First Trip I	Steady State ELOX	Steady State LLOX
LTV Crew Module	7.69	7.69	7.69
Aeroshell	5.23	5.23	5.23
LOI & TEI Propellant Tanks	1.43	1.43	1.43
Engines (4)	2.62	2.62	2.62
Structure & Avionics	12.79	12.79	12.79
TOTAL LTV INERT MASS	29.76	29.76	29.76
Return Payload	1.00	1.00	1.00
Crew (5)	0.44	0.44	0.44
RCS Propellant	0.20	0.20	0.20
TEI Propellant:			
From Earth	8.36	8.36	1.19
LLOX	N/A	N/A	7.17
LUNAR DEPARTURE MASS	39.76	39.76	39.76
Less Return Payload	-1.00	-1.00	-1.00
Less Crew	-0.44	N/A	N/A
Less TEI LLOX	N/A	N/A	-7.17
LTV Delivered Payload:			
LEV CM (W/Crew)	6.00	N/A	N/A
LEV Inert (w/GN&C)	6.71	N/A	N/A
LEV Cargo	24.32	25.10	20.21
LEV Propellants	29.67	30.07	5.33
LEV Propellant Tanks	N/A	1.10	0.96
LOI Propellant	27.96	25.32	15.47
LUNAR ARRIVAL MASS	132.98	120.35	73.56
TLI Tanks and Fairings (Staged)	7.90	7.90	7.90
TLI Propellant	131.27	118.90	72.61
LTV (IMLEO)	272.15*	247.15*	154.07

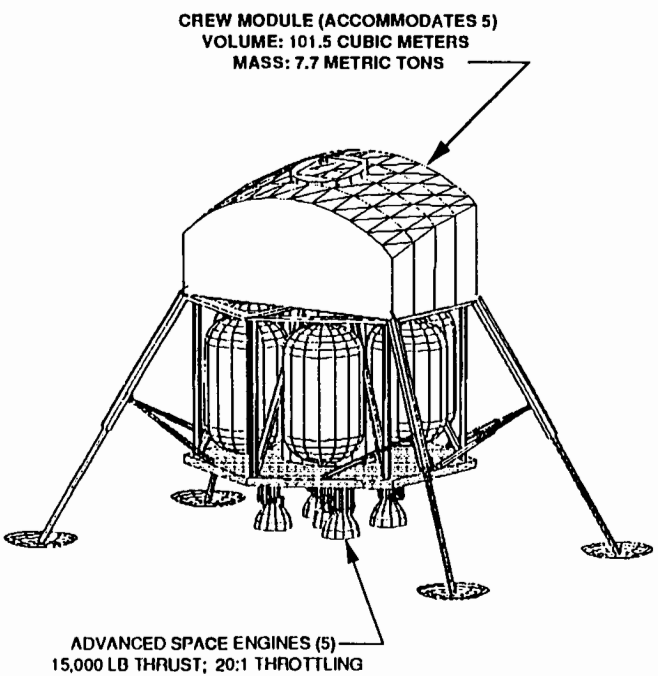
* Requires two HLLV launches for each LTV mission (136t Payload each)
- First Trip (All Earth LOX)
- Steady State LEV Operations Mode (All Earth LOX)
- Steady State LEV Operations Mode (Lunar LOX)
- Cargo capability limited by LEV, not LTV for steady-state cases



Component	First Trip t	Steady State ELOX	Steady State LLOX
LOI Propellant Tanks	1.40	1.40	1.40
Engines (4)	2.62	2.62	2.62
Structure & Avionics	12.82	12.82	12.82
TOTAL TMIS PAYLOAD	16.84	16.84	16.84
LTV Delivered Payload:			
LEV Inert (w/GN&C)	11.10	N/A	N/A
LEV Cargo	36.57	42.98	42.98**
LEV Propellants	33.06	36.33	6.87
Tanks for LEV Prop.	N/A	1.43	1.23
LOI Propellant	35.41	35.41	24.62
LUNAR ARRIVAL MASS	132.98	132.98	92.54
TLI Tanks etc. (Staged)	7.90	7.90	7.90
TLI Propellant	131.27	131.27	91.28
LTV (IMLEO)	272.16*	272.16*	191.72*

* Requires two HLLV launches for each LTV mission (136t Payload each)
** LEV descent cargo limited
- First Trip (All Earth LOX)
- Steady State LEV Operations Mode (All Earth LOX)

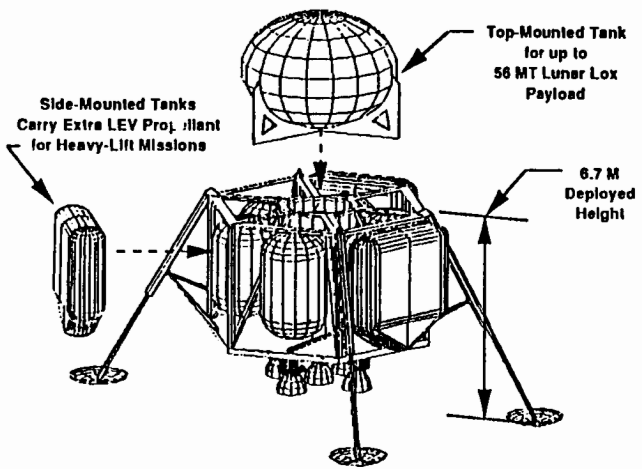
Figure 2. Cargo lunar transfer vehicle.



Component	First Trip (ELOX) Mass (t)	Steady State (ELOX) Mass (t)	Steady State (LLOX) Mass (t)
Crew Module	5.56	5.56	5.56
Structure	3.65	3.65	3.65
Propulsion			
Main	1.42	1.42	1.42
Propellant Tanks	1.34	1.34	1.34
Avionics	0.30	0.30	0.30
INERT MASS	12.27	12.27	12.27
Crew	0.44	0.44	0.44
Ascent Payload:			
Return Cargo	1.00	1.00	1.00
LLOX for LTV	N/A	N/A	7.17
LLOX for LEV Descent	N/A	N/A	15.74
Tanks for LTV LLOX	N/A	N/A	0.11
Ascent LOX/LLOX	6.09	6.09	16.46
ASCENT LH2	1.01	1.01	2.74
ASCENT STAGE MASS	20.81	20.81	55.93
Less Ascent Payload	-1.00	-1.00	-23.91
Less Ascent LLOX	N/A	N/A	-16.46
Descent Cargo	24.30	25.10	20.21
Descent Propellant	22.55	22.39	18.36
LEV MASS AT LLO	66.66	67.30	54.13

First Trip (ELOX) - LEV delivered on LTV to LLO. Uses all earth supplied propellants
Steady State (ELOX) - LEV in LLO. Uses all earth supplied propellants
Steady State (LLOX) - LEV in LLO. Uses earth supplied LH2 and lunar produced LOX
(Maximum LEV descent payload capability depicted in steady state cases)

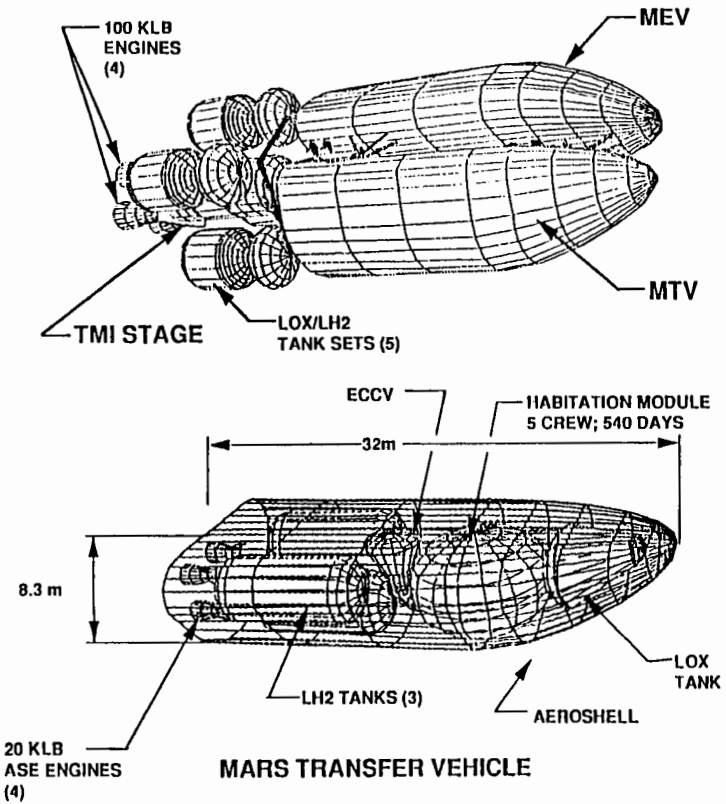
Figure 3. Piloted lunar excursion vehicle.



Component	First Trip (ELOX) Mass (t)	Steady State (ELOX) Mass (t)	Steady State (LLOX) Mass (t)
Structure	7.45	7.45	7.45
Propulsion			
Main	1.42	1.42	1.42
Propellant Tanks	1.93	1.93	1.93
Avionics	0.30	0.30	0.30
INERT MASS	11.10	11.10	11.10
Ascent Payload:			
Cargo (LLOX for MTS)	N/A	N/A	56.56
LLOX for LEV Descent	N/A	N/A	7.87
Tanks for MTS LLOX	N/A	N/A	0.91
Ascent LOX/LLOX	5.07	5.07	34.11
ASCENT LH2	0.84	0.84	5.69
ASCENT STAGE MASS	17.01	17.01	116.24
Less Ascent Payload	N/A	N/A	-64.43
Less Ascent LLOX	N/A	N/A	-34.11
Descent Cargo	36.32	42.98	N/A
Descent Propellant	27.39	30.80	9.19
LEV MASS AT LLO	60.72	90.79	26.89

First Trip (ELOX) - LEV delivered on LTV. Earth supplied Propellants; Delivers Cargo to lunar surface and returns empty to LLO.
Steady State (ELOX) - LEV in steady state mode in LLO. All earth supplied Propellants
Steady State (LLOX) - LEV in steady state mode in LLO. Uses earth supplied LH2 and lunar produced LOX. Delivers maximum payload of LLOX to LLO for MTS
(Steady state cases depict maximum LEV payload capabilities)

Figure 4. Cargo lunar excursion vehicle.



Component	2016 Feb 6 t	2018 Feb 25 t	2020 Jan 16 t
MTV Crew Module	46.9	46.9	49.2
Structure & Avionics	14.5	14.5	14.5
Tether	3.0	3.0	3.0
Power (Solar Panels)	1.2	1.2	1.2
Propellant Tanks	9.8	9.8	9.8
Engines (4)	0.7	0.7	0.7
RCS	0.8	0.8	0.8
ECCV	7.5	7.5	7.5
MTV INERT MASS	84.4	84.4	86.7
Crew & Return Payload	1.4	1.4	1.4
Bolloff	3.7	0.9	1.1
EOI Propellant	38.6	9.0	9.5
TEI Propellant	166.5	91.7	80.0
MARS DEPARTURE MASS	294.6	187.4	178.7
Less Return Payload & Crew	-1.4	-1.4	-1.4
Aeroshell	59.1	59.1	59.1
Bolloff	5.0	2.5	2.2
MOI Propellant	7.7	5.3	5.1
MARS ARRIVAL MASS	365.0	252.9	243.7
Bolloff	17.3	7.6	9.3
MTV Spin Propellant	5.7	3.8	3.7
MTV	388.0	264.3	256.7
MEV	94.9	93.0	92.9
COMSAT	3.0	3.0	3.0
TMIS Inert	48.5	51.5	66.1
TMI Propellant	248.4	398.0	718.2
MTS LLO DEPARTURE	782.8	809.8	1136.9
Less LLOX	-453.0	-474.1	-739.8
LEV LH2	61.2	64.0	99.9
LEV & LOI TANKS	13.8	14.4	21.6
Bolloff	0.2	0.2	0.4
LOI Propellant	109.1	111.6	139.6
LUNAR ARRIVAL MASS	514.1	525.9	658.6
TLI Tanks	13.6	13.9	17.4
Bolloff	0.2	0.2	0.3
TLI Propellant	528.9	540.7	676.9
MTS IMLEO	1056.8	1080.7	1353.2

Figure 5. Piloted Mars transfer system.

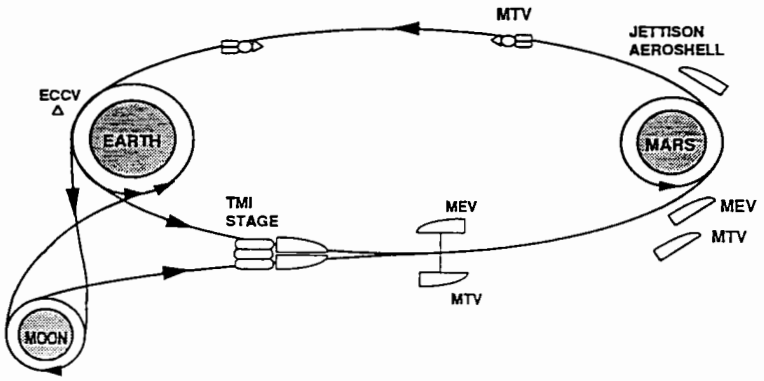


Figure 6. Typical profile of Mars mission using LLOX. Profile characteristics: chemical propulsion using LLOX; artificial gravity (outbound); aerobraking.

aerobrake, a large assembly facility, and fluid transfer for every mission. A piloted LTV, such as the one shown in Fig. 1, with a core stage, drop tanks, advanced space engines, and a remote manipulator system to deploy the LEV offers good performance and reduces on-orbit operation (Schmidt 1991). The LTV can dock with an LEV for either cargo or personnel transfer to or from the surface of the Moon. The reduction in initial-mass low Earth orbit (IMLEO) as the operations transition to steady-state LLOX utilization from Earth LOX is about 93 tons per flight as shown in the table accompanying Fig. 1. The cargo performance exceeds the 4-crew and 16-ton cargo requirement because full advantage of two heavy-lift launch vehicle launches is taken to place the LTV in LEO. Cargo capability with an LTV designed to be launched on one heavy-lift launch (assumed at 136 tons each payload capability) was inadequate.

The cargo LTV uses the same propulsion as the piloted configuration, but is expendable in order to maximize cargo performance. Because it is expendable, the aerobrake is not included and the appearance could be quite different, as shown in Fig. 2. The reduction in IMLEO when using LLOX in steady-state operations instead of Earth LOX is 80 tons per flight as shown in the accompanying table.

B. Design for a Lunar Excursion Vehicle

LOX/LH_2 is also the reference propulsion system for the LEV because of its high performance and adaptability to *in-situ*-derived LLOX. Possible configurations of the piloted and cargo LEV are shown in Figs. 3 and 4. The size of the LEV is based on full utilization of the LTV capability as discussed above. In addition, the cargo LEV is designed to deliver the maximum amount of LLOX to LLO for the Mars transfer system while meeting all the constraints of the lunar transfer.

An alternative fuel for the LEV is lunar *in-situ*-derived fuel silane. Table V shows a comparison of the intial-mass LEOs resulting from the use

TABLE IV
Lunar Transfer Performance Analysis Ground Rules

Crew Module(CM), Crew, and Returned Masses: (Not Part of Delivered Cargo)	
LTV CM (w/5 Crew)	8.13 ton (w/Crew = 0.443 ton)
LEV CM (w/5 Crew)	6.0 ton (w/Crew = 0.443 ton)
Returned Cargo	1.0 ton (Crew Missions)
Specific Impulses (LOX/LH_2):	
LTV	475 s
LEV	470 s
Delta-Vs, km s^{-1}:	
Trans Lunar Injection (TLI)	3.2
Lunar Orbit Insertion (LOI)	1.1
Trans Earth Injection (TEI)	1.1
Earth Orbit Insertion (EOI)	0.0 (Aerobraked)
Lunar Descent & Landing	1.903
Lunar Ascent & Rendezvous	1.927
Inert Masses:	
LTV (Tanks, Engines, and Structure)	10% of LTV Initial Propellant and Payload
LTV Aerobrake Shield	20% of Entry Mass (w/o Aerobrake)
LTV RCS	0.20 t
Piloted LEV Inert	6.71 t
Cargo LEV Inert	11.1 t
LTV LEO Node Operations	— Reusable LTV Structure is Only Mass at LEO Node Prior to Launch — One-Way LTV: No Mass at LEO Node Prior to Launch
LEV/LTV LLO Node Operations	— At LLO Node When LTV Arrives and Departs — LTV Supplies LH_2 — Crew Module Not Used for Cargo Only Missions

of either Earth LH_2 or lunar-derived silane (SiH_4) to fuel the LEV. There is about a 6% reduction in initial-mass LEO when the silane fuel is used because the required mass of LH_2 from Earth is lower for the silane-fueled LEV. The cargo performance of the LEV was about the same for both propellant combinations because of the different staging of the Earth H_2 and the lunar silane. Descent was performance limiting for Earth H_2, while ascent was limiting for

TABLE V
Comparison of Lunar Mission IMLEOs Using Silane

Configurations:
— Steady-State LEV Operations Mode Using LLOX
Piloted and Cargo Only
Earth LH_2 and Lunar Silane LEV Fuel Alternatives
— Piloted LEV Inert Mass = 6.71 tons
— LEV Crew Module w/Crew Mass = 6.00 tons
— Cargo LEV Inert Mass = 11.1 tons
— I_{sp} of LLOX/Silane = 370 s

Lunar Transfer Vehicle	Piloted E-LH_2	Piloted L-Silane	Cargo E-LH_2	Cargo L-Silane
Component	tons	tons	tons	tons
LTV Crew Module	7.69	7.69		
Aeroshell	5.23	5.23		
LOI & TEI Propellant Tanks	1.43	1.43	1.40	1.40
Engines (4)	2.62	2.62	2.62	2.62
Structure & Avionics	12.79	12.79	12.82	12.82
TOTAL LTV INERT MASS	(29.76)	(29.76)	(16.84)	(16.84)
Return Payload	1.00	1.00		
Crew (5)	0.44	0.44		
RCS Propellant	0.20	0.20		
TEI Propellant:				
From Earth	1.19	1.19		
LLOX	7.17	7.17		
LUNAR DEPARTURE MASS	(39.76)	(39.76)	(16.84)	(16.84)
Less Return Payload	−1.00	−1.00		
Less TEI LLOX	−7.17	−7.17		
LTV Delivered Payload:				
LEV Cargo[a]	20.21	20.21	42.98	42.98
Hydrogen	5.33	2.25	6.80	2.99
Tank for Hydrogen	0.96	0.41	1.22	0.54
LOI Propellant	15.47	14.43	24.62	22.99
LUNAR ARRIVAL MASS	(73.55)	(68.65)	(92.46)	(86.33)
TLI Tanks and Fairings (Staged)	7.90	7.90	7.90	7.90
TLI Propellant	72.61	67.77	91.28	85.23
HLLV Payload (IMLEO)	154.07	144.33	191.64	179.46

TABLE V (cont.)
Comparison of Lunar Mission IMLEOs Using Silane

Percent IMLEO savings	6.32%	6.36%

[a] The cargo performance of the LEV was about the same for both propellant combinations because of the different staging of the Earth H_2 and the lunar silane. Descent was performance limiting for Earth H_2, while ascent was limiting for lunar silane, both at about the same cargo capability.

lunar silane, both at about the same cargo capability. While the performance is about the same, there are design differences between an LEV fueled by silane/LLOX and the one fueled by LH_2/LLOX because of the difference in I_{sp} and fuel density. Also, the exhaust products (SiO_2, etc.) of silane combustion are very harsh on the engines; the life of the LEV may be limited relative to that of the LH_2/LLOX-fueled LEV.

IV. VEHICLE CONCEPTS FOR A MARS MISSION

As with the lunar mission, Mars transfer is performed by two types of common vehicles: the Mars transfer vehicle (MTV) and the Mars excursion vehicle (MEV) (Schmidt 1991). In addition, the Mars transfer system includes a trans-Mars injection stage that is sized according to mission ΔV requirements as shown in Fig. 5. Each MTV and MEV has a cargo-only configuration and a piloted configuration that adds a crew module for personnel transfer to the common vehicle.

A scenario for a Mars mission that can take advantage of lunar *in-situ*-derived propellant is shown in Fig. 6. In this scenario, the piloted Mars transfer system stops in low-lunar orbit to load LLOX delivered by the LEV and to deliver the LH_2 required by the LEV. The basic requirements from the 90 Day Study (NASA 1989) include at least 4 crew and delivery of at least 25 tons to the Martian surface with the crew and return of 1 ton of cargo with the crew. An unpiloted cargo mission will have preceeded the piloted mission to emplace habitats and supplies. The first human landing on Mars would be in 2016. The ΔVs and trip times for several piloted Mars mission opportunities (via Moon and direct from LEO) are shown in Table VI and the ground rules for a typical vehicle design are shown in Table VII. Although not required by the 90 Day Study, the ground rules include artificial gravity using a tether between the MEV and the piloted MTV and an associated spin-up/spin-down reaction control system. Figure 5 from Bienhoff and Skruch (1990) shows a piloted Mars transfer system that can meet all requirements and even uses a common aeroshell design for the MTV and MEV vehicles. The mass breakdown for three mission opportunities accompanies the figure.

A comparison of initial-mass LEOs between lunar staging (using LLOX) and direct (not using LLOX) missions is made in Table VIII for the vehicle design described in Fig. 5. In this table, the MTV mass was variable for comparison purposes, while the design shown in Fig. 6 has a fixed (common)

MTV that only changes slightly due to consumables loaded for different mission durations. Table VIII shows that the reduction in initial-mass LEO due to the use of LLOX for piloted missions is between 142 tons and 1386 tons depending on the opportunity.

TABLE VI
Mars Round-Trip Mission Delta-Vs[a]

Departure Date	2015 Nov 8	2016 Feb 6	2018 Feb 25	2020 Jan 16	2022 Feb 4	2024 Jan 25
Outbound Time, Days	270	190	170	230	210	210
Inbound Time, Days	230	230	250	270	270	270
Lunar Staging Earth-Mars Delta-Vs km s^{-1}						
TLI	3.20	3.20	3.20	3.20	3.20	3.20
LOI	1.10	1.10	1.10	1.10	1.10	1.10
TMI	2.91	1.76	3.12	4.61	6.78	9.45
MOI	0.10	0.10	0.10	0.10	0.10	0.10
TEI	2.03	3.94	3.20	2.83	2.68	2.30
EOI	1.59	1.78	0.48	0.49	0.20	0.20
Direct Earth-Mars Delta-Vs, km s^{-1}						
TMI	4.61	3.66	4.82	6.11	8.18	10.65
MOI	0.10	0.10	0.10	0.10	0.10	0.10
TEI	2.03	3.94	3.20	2.83	2.68	2.30
EOI	1.59	1.78	0.48	0.49	0.20	0.20
Entry Velocities km s^{-1}						
at Mars	6.67	7.50	9.23	7.29	8.20	8.77
at Earth	14.50	14.50	14.50	14.50	13.83	12.65

[a] Criteria: Stay time at Mars = 60 days.
Round-trip time $\leq$1.53 yr.
Includes post-capture trim ΔV (MOI) = 100 m s^{-1}.
Includes post-capture trim ΔV (EOI) = 200 m s^{-1}.
Note: Outbound times do not include lunar staging time.

However, the same conclusion cannot be reached for cargo only Mars missions. The propellant requirements for the cargo only mission are based only on the trans-Mars injection maneuver, which requires a ΔV of only 3.6 to 4.2 km s^{-1} for opportunities every two years. Because this one-way ΔV is small relative to the round trip ΔVs, using LLOX provides a much smaller benefit. The benefit is so small that it does not make up for the ΔV required to transfer from the Earth to the Moon to acquire the LLOX.

The scenario used for these comparisons is one of many including LEO staging, lunar surface staging, or Earth-Moon libration point staging. The

TABLE VII
Mars Vehicle Sizing Ground Rules

Mission Payload	Piloted MEV: 89.22 ton + spin propellant; 1 ton return cargo Cargo MEV: 2 @ 86.02 ton (one-way)
Number of crew	5
Crew mass	443 kg
Specific Impulses:	
TLI maneuver	470 s
TMI (Direct Earth-Mars only)	470 s
All Other MTS Maneuvers	475 s
MTS IMLEO includes:	– LH_2 for LEV –All required MTS LOX tanks
LLOX required for LEV operation Based on:	81% of LLOX delivered to MTS in LLO
LEV inert mass	15% of LLOX delivered per trip
LEV mass ratios (M_i/M_f)	
Descent	1.518 (I_{sp}=465 s)
Ascent	1.526 (I_{sp}=465 s)
Earth-Moon transit time	2 days
Time in low-lunar orbit	4 days
Propellant boiloff rate	1.2% per month
TLI thrust acceleration (T/W)	0.15 g
TMIS tanks & structure mass	10% of TMI propellant (or TLI+LOI prop., if greater) + 10% of LH_2
TMIS engine mass	Thrust/($g \times$ (T/W))
MTV inert mass (tanks, structure, & engines)	10% of MTV Propellant + 1% of ECCV* + 1% of Crew Module
Additional MTV masses:	
Crew module	46,910–49,170 kg (varies with mission (duration)
RCS hardware	831 kg
Solar panels	1,224 kg
ECCV for crew return	7,500 kg
Artificial g tether & equipment	3.902 kg/m
MTV Mars aerobrake mass	20% of Mars entry mass – Aerobrake staged prior to TEI maneuver
Comsat mass	3 ton (Staged prior to MOI)
Artificial g on Earth-Mars Trip (None on Mars-Earth Trip)	0.7 g (Piloted MTV) 2 RPM

* Earth Capture Crew Vehicle

TABLE VIII

Performance of Mars Transfer System Using Lunar Liquid Oxygen

Departure Date	2015 Nov 8	2016 Feb 6	2018 Feb 25	2020 Jan 16	2022 Feb 4	2024 Feb 24
Mars via Moon, Variable MTV:						
IMLEO, ton	877	1,057	868	1,056*	1,630	4,191
MTV Inert, ton	104	143	99	91*	92	89
TMIS Inert, ton	65	76	64	83*	153	480
Direct from LEO (no LLOX), Variable MTV:						
IMLEO, ton	1,019	1,230	1,032	1,495	3,016	n/a**
MTV Inert, ton	104	143	99	99	92	n/a
TMIS Inert, ton	67	70	69	113	261	n/a

* Dec. 17, 2019.

** n/a: IMLEO much too great for practical consideration.

characteristics and advantages of the various basing options have been described by Woodcock (1988) His conclusion was that L2 was the best location, saving about 0.5 km s^{-1} relative to LLO for the Mars mission. Thus, the initial-mass LEOs based on the scenario shown in Fig. 6 may not be the lowest possible using *in-situ*-derived propellants.

V. COST BENEFIT OF UTILIZING IN-SITU-DERIVED PROPELLANTS

Figure 7 indicates a time-phased set of mission requirements for a typical lunar base, derived from the NASA 90-day study (NASA 1989). A permanent lunar surface outpost is established, with crew rotation and resupply every 6 months. In this baseline case, it is assumed that all mass emplaced on the lunar surface or returning to the Earth originates on the Earth's surface.

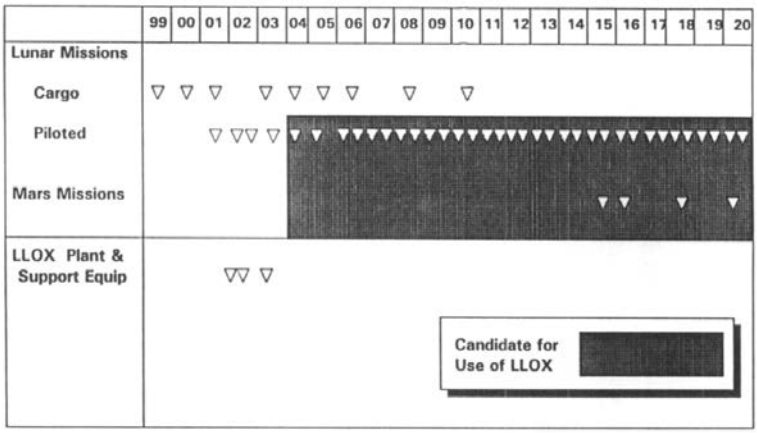


Figure 7. Typical Lunar/Mars mission model from the NASA 90-day study (NASA 1989).

Using the typical LTVs and LEVs, described in the previous section, these requirements may be translated into a total mass flow per mission, and into the amount of initial mass placed in LEO. Transportation from LEO to LLO and back is supplied by the LTV, and from LLO to the lunar surface is supplied by the LEVs. Note that operation of returning LLOX to LEO via the LTV has been added to the operations described in the previous section. As indicated in Fig. 8, requirements of typical lunar missions are very strongly driven by the need to continually provide propellant in LEO to transfer people and hardware to the lunar surface, and rotate crew from the lunar surface through LEO to the Earth.

Figure 8 also shows that missions to support Mars surface operations are even more strongly driven by the need to provide large masses or propellant in LEO, if chemical propulsion systems are used. Again, a time-phased set of mission requirements is used derived from the NASA 90-Day study (NASA

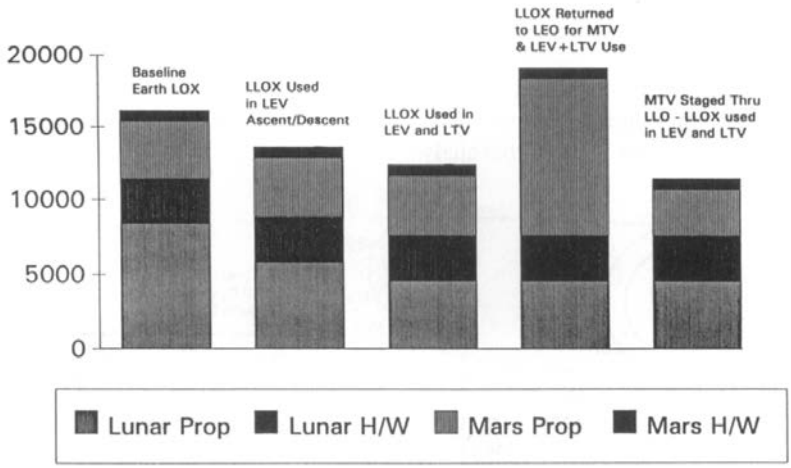


Figure 8. Diagram showing LLOX usage which reduces mass requirements for lunar and Mars missions.

1989). As shown in Fig. 9, for a combined Mars and Lunar operations scenario, the vast majority of the mass needed in LEO is propellant.

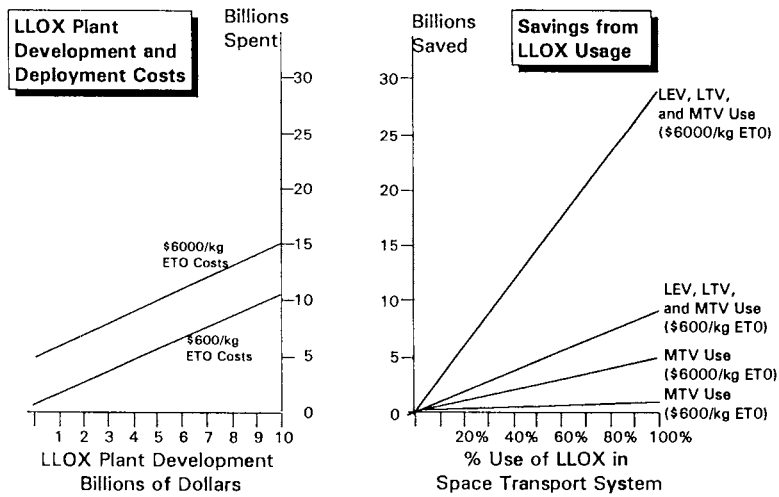
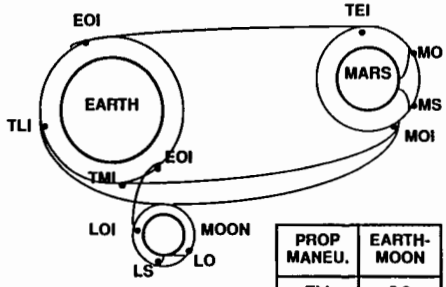


Figure 9. Diagram showing cost effectiveness of LLOX usage dependent upon LEO transport costs and LLOX development costs.

A. Mission Applications and Trades

Earth orbit to the Moon. Transferring cargos between the Earth's surface and the lunar surface is the primary mission driver for lunar support operations.

While different systems of operational scenarios can be examined to accomplish this (going directly from Earth orbit to the lunar surface, for example), and characteristics of the system elements can influence the results (I_{sp}, vehicle sizing, dry weights, etc.), a quick trade analysis can illustrate the effect of using lunar-derived propellants. Figure 10 shows a set of representative characteristics assumed in this analysis.



PROP MANEU.	EARTH-MOON
TLI	3.2
LOI	1.1
LS	1.9
LO	1.9
TEI	1.1
EOI	0.3
TOTAL	9.5

- Total trip time = 395 days
- 30 Day stay time
- Delta Vs in km/sec

PROP MANEU	EARTH-MARS		MOON-MARS	
	2007	2016	2007	2016
TLI	0	0	3.2	3.2
LOI	0	0	1.1	1.1
TMI	4.3	3.6	3.2	1.7
MOI	0	0	0	0
MS	1.2	1.2	1.2	1.2
MO	5.5	5.5	5.5	5.5
TEI	7.5	4.9	7.5	4.9
EOI	0.3	0.3	0.3	0.3
TOTAL	18.8	15.5	22.0	17.9

Figure 10. Representative mission parameters.

The key trade parameter here is the cost of transportation to LEO, provided by some form of Earth-to-orbit transportation. This will be compared against the relative costs of developing and operating a lunar oxygen production facility. Current transportation costs to LEO are in excess of $6000 per kg. With increased usage and operational improvements, these costs can be reduced by a factor of 2 to 3. Future systems, such as a new technology heavy-lift launch vehicle, can potentially provide an order of magnitude reduction in Earth-to-oirbt costs to $600 to $700 per kg.

Ascent/Descent Applications. Lunar propellants can be applied to supply the LEV vehicle used for transport between the lunar surface and lunar orbit. If lunar-derived liquid oxygen is used to offset the transport of LOX from the Earth's surface for this vehicle, a reusable LEV must be assumed, as it must be based upon the lunar surface, or left in LLO between uses.

A LLOX plant and support equipment must be added to the required missions to the lunar surface. This is estimated to add approximately 100 metric tons to the transportation requirements, including power supply, LOX storage system, and necessary surface equipment (loaders, transporters, etc.; Repic et al. 1990). Three additional cargo LEV flights are required to emplace this system on the Lunar surface, which is assumed to occur prior to 2004 in the scenario shown in Fig. 7. It should be noted this analysis still requires

LH_2 to be transported from LEO to lunar orbit for each round-trip LEV usage.

Figure 11 shows the total differential LEO mass requirements from this case. Using lunar-derived LOX (LLOX) to offset Earth-originated LOX reduces the need to transport 18.5 tons of LOX for each manned LEV usage. An additional 1096 kg of LH_2 is required to be transported to LLO to allow the transport of the LOX required for lunar descent into orbit. This net saving of 17.4 tons in LLO allows either more lunar surface cargo to be carried per mission (requiring a resizing of the LEV), or may be used to reduce the initial propellant required in LEO, because the total mission weight at LLO is reduced. Using the LLOX to reduce LEO mass, reduces the total LEO mass requirement by 1900 tons. This may be valued from 8.2×10^9 ($6000/kg Earth-to-orbit costs), to $\$2.7 \times 10^9$ ($2000/kg), or potentially as low as $820 million ($600/kg Earth-to-orbit costs). These benefits must then be compared to the development and operations cost of the LLOX facility.

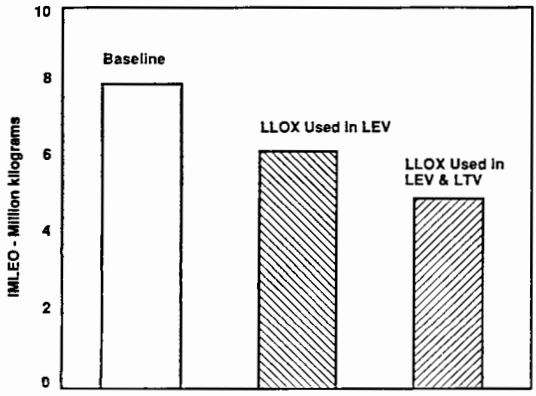


Figure 11. Diagram showing use of LLOX which reduces the total mass in LEO for lunar operations.

Orbit-to-Orbit Applications. Orbit-to-orbit applications of lunar-derived propellants is an area of great interest and potential significant payoff for lunar mission applications. Utilization of lunar propellants can directly and significantly influence the cost and effectiveness of space transportation for cis-lunar operations. Typical applications include: LEO-lunar orbits and LEO-lunar orbits-Mars orbits.

For most scenarios postulated for lunar operations, LEO-lunar orbit transportation requirements are the primary driver for propellant requirements. If lunar-derived LOX is used to supply the LTV operating from LEO to LLO, these propellant requirements may be substantially reduced. Again, a cost trade based upon differential mass in LEO can be performed.

Because all systems assumed in this analysis use LH_2/LOX engines, LH_2 must still be carried for each mission. However, LOX is assumed to be

supplied from the lunar surface for all LEO to LLO to LEO missions after the LLOX plant is installed. It is assumed that each LTV will return to LEO with sufficient LLOX on board to return to LLO. Tankage needed to contain the LLOX for transport to LEO and back to LLO is added to the analysis at 6% of the LLOX mass. Aerobraking is assumed to be performed at LEO at the end of each piloted LTV mission, and the aerobrake mass is estimated at 15% of the total mass decelerated by the aerobrake. As before, each cargo LTV is assumed expendable and they are not considered as candidates for LLOX usage. At LEO, LH_2 is loaded to accomplish the round trip from LEO to LLO and return, as well as sufficient LH_2 for the round-trip descent/ascent to the lunar surface.

For this analysis, using the rocket equation, a parametric relationship can be found between the inert mass, payload mass, and fuel mass for the LTV operating from LLO to LEO.

$$\mathrm{R1} = \mathrm{M_{o1}}/\mathrm{M_{v1}} = \exp[\Delta V/(G * I_{sp})]$$

ΔV = Delta Vee velocity change required, here being 1100 m s^{-1}.

$$\mathrm{G} = \text{Standard gravity} = 9.81 \mathrm{ms}^{-2}$$

I_{sp} = specific impulse of rocket system used, here being 475 s from the LH_2/LO_2 system used.

$$\mathrm{M_{o1}} = \mathrm{M_I}(1.15) + \mathrm{M_{PL1}}(1.06)(1.15) + \mathrm{M_{F1}}[1 + 0.06(1.15)] \qquad (1)$$

M_{I1} = inert mass, consisting of engines, crew module, and structure and will remain constant between the LEO to LLO and the LLO to LEO legs of the mission. This is 17050 kg. A 15% increase accounts for the aerobrake mass to decelerate it.

M_{PL1} = payload mass returned to LEO from LLO. This must equal the LLOX required for the return leg to LLO. A 6% tank fraction is added, and a 15% increase to account for aerobrake mass. It should be noted with the 6:1 LOX/LH_2 ratio rocket system assumed here, that this will be 6/7th of the fuel required for the LEO to LLO leg. M_{F1} = The mass of the fuel at the start of the propulsive burn sequence. Here, it is the burn from LLO to LEO. It is increased by 6% for tankage fraction. The 15% aerobrake increase is only applied to the tankage mass, because when the system reaches the Earth's atmosphere, the fuel will have been expended, and only the tank mass will remain.

Similarly, M_{v1}, the mass remaining at the end of the burn, is found to be

$$\mathrm{M_{v1}} = \mathrm{MI}(1.15) + \mathrm{MPL_1}(0.06)(1.15) + \mathrm{MF_1}[1 + 0.06(1.15)].$$

For the mission leg from LEO to LLO, a similar relationship can be found between the inert mass, payload mass and fuel mass.

$$\mathrm{R_2} = \mathrm{M_{o2}}/\mathrm{M_{v1}} = \exp[\Delta V_2/(g * I_{sp})]$$

ΔV_2 = Delta Vee velocity change required, here being 4300 m s^{-1}

g = Standard gravity = 9.81m s^{-2}

I_{sp} = 475 s.

$$M_{o2} = M_{I1}(1.15) + MPL_2(1.06)(1.15) + MF_2[1 + 0.06(1.15)] \quad (2)$$

M_{I1} = the inert mass stays constant over the whole mission.

M_{PL2} = payload mass sent to LLO from LEO. This must equal the mass of LH_2 needed to return from LLO to LEO, plus LH_2 to lift the LLOX returned to LEO from the lunar surface, plus any cargo for the surface (assumed to be 20 tons).

M_{F2} = the mass of the fuel at the start of the propulsive burn sequence. Here, it is the burn from LEO to LLO. Because we are assuming an engine mass ratio of 6:1 between the LOX and LH_2 used in the engine system, we can establish a relationship between the amount of payload to be carried on each leg of this journey. If we solve for M_{F1}, 6/7 M_{F1} is the amount of LLOX needed to be launched from the lunar surface into LLO to return the vehicle (plus LLOX cargo) to LEO. Looking at the capability of the LEV, it is observed that it required 0.279 kg of LH_2 to lift 1 kg of LLOX into LLO. Thus, $0.279 \times (6/7) \times M_{F1}$ = LH_2 necessary to launch the LLO-to-LEO fuel into LLO. Similarly, $M_{F1}/7$ = LH_2 necessary to propel the vehicle from LLO to LEO.

The total payload sent from LEO to LLO must also include the LH_2 to launch the return payload from the lunar surface to LLO. This is $0.279 \times M_{PL1}$. The total payload sent LEO-to-LLO is then, $M_{PL2} = 0.279(6/7)\ M_{F1} + M_{F1}/7 + 0.279 \times M_{PL1} + M_{SP}$, where M_{SP} is any surface payload mass.

This allows us to recursively calculate the necessary masses. From Eq. (1), a test value of M_{PL1} is assumed, and M_{F1} is calculated. Then, using the above relationship, M_{PL2} is calculated. From Eq. 2, M_{F2} is calculated. M_{PL1} is re-calculated from M_{F2} because the LLOX needed for the LEO-to-LLO flight, equals $(6/7) \times M_{F2}$, at the 6:1 mixture ratio assumed. Successive approximations are made a constant value of M_{PL1} is obtained.

This method assumes a LLOX return payload to LEO from LLO and figures out the mass of LH_2 needed to be sent to LLO to return this LLOX payload. This increased LH_2 needed in LLO increases the initial mass to be sent to LLO, which increases the amount of propellent needed in LEO, which increases the LLOX needed at LEO to start the mission cycle. The recursive calculations provide a closer and closer iteration to the required LLOX amount.

Solving this set of equations yields a total requirement for the system of 107.160 kg of LH_2 to be available in LEO at the start of each cycle. Again assuming a 6% tank fraction, this means that 113.589 kg must be launched into LEO to perform each mission cycle. Table IX summarizes this scenario for the mission model assumed.

TABLE IX
LLOX LEO-LLO Case

	Piloted Missions				Cargo Missions		
1	First Flight (ELOX)	@	272.16	1	First Flight (ELOX)	@	272.16
3	Flights (ELOX)	@	247.24	6	Flights (ELOX)	@	272.16
1	Flights (LLOX)	@	154.07	5	Flights (LLOX)	@	191.64
31	Flights (LLOX)	@	113.59				2863.32
			4689.24				
Total IMLEO = 7582.56 tons							

This application reduces the total initial-mass LEO mass requirement by 3792.44 tons. This may be valued from $\$22.8\times10^9$ ($6000/kg Earth-to-orbit costs), to $\$7.6\times10^9$ ($2000/kg), or potentially as low as $\$2.3\times10^9$ ($600/kg Earth-to-orbit costs). Again, this benefit must then be compared to the development and operations cost of the LLOX facility.

VI. APPLICATIONS OF LUNAR-DERIVED PROPELLANTS TO MARS MISSIONS

Further benefits are possible using lunar-derived LOX in a chemically powered MTV. As shown in Fig. 8, a large amount of the propellant needed in LEO is to support Mars missions. Two cases can be examined, to return additional LLOX to LEO (over that needed to return the LTV from LEO to LLO), or to transfer the partially filled MTV from LEO into LLO where the fuel tanks are topped off with LLOX.

In the first case, Eq. (1) can be used, if the M_{PL1} is increased by an amount of LLOX payload to be returned to LEO. However, it is found that it requires 159,358 kg of LH_2 to return 50,000 kg of additional cargo LLOX to LEO with the reference systems used here. Table VIII shows the impact upon initial-mass LEO for staging the MTV from LEO into LLO, and thence to Mars. Here, the MTV carries the LH_2 needed to launch the LLOX into LLO for the MTV. For the four Mars missions included in the scenario shown in Fig. 8, this results in a mass savings of 918 tons. This savings may be valued at $\$5.5\times10^9$ ($6000/kg Earth-to-orbit costs) to $\$1.8\times10^9$ ($2000/kg Earth-to-orbit costs), or as low as $\$0.55\times10^9$ ($600/kg).

Table X summarizes these results. These benefits are dependent upon the use of a chemically powered MTV. The long trip times and high-energy requirements for a Mars mission may, however, force the use of a nuclear-powered MTV. If a nuclear stage is used, the requirements for propellent in LEO are substantially lowered, and the applicability of lunar propellants is further reduced by the lack of hydrogen-rich lunar-derived propellants.

TABLE X
Summary of Transportation Options

Lunar Missions IMLEO (tons)	Mars Missions	Initial Mass in LEO (IMLEO)	Cost Savings (10^9*)
Earth LOX (ELOX)	LEO Launch of MTV, Earth LLOX	16151	Baseline
11375	4776		
LLOX for LEV Ascent/Descent	LEO Launch of MTV, Earth LOX	13583	$15.4 to $5.1
8807	4776		
LLOX for LEV and LTV	LEO Launch of MTV, Earth LOX	12359	$22.8 to $7.6
7583	4776		
LLOX for LEV and LTV	LEO Launch of MTV, LLOX returned to LEO for MTV	19116	($17.8) to ($5.9)
7583	11533		
LLOX for LEV and LTV	MTV Staged through LLO, LLOX Loaded at LLO	11441	$28.3 to $9.4
7583	3858		

* Note: Cost savings calculated at $6000/kg and $2000/kg or Earth-to-orbit cost savings.

A. Surface Transportation Applications

Several applications of lunar materials for surface transportation have been proposed to power lunar rovers or supply propellant for sub-orbital "hoppers." Such missions, if supplied from the Earth's surface, increase the required mass flow to the lunar surface. Using lunar materials reduces the mass flow from the surface of the Earth to the lunar surface allowing a direct cost trade from the differential transportation costs (Stern 1988). Propellant to support surface transportation is a minor portion of the overall total mission requirements to support an operational lunar base. The recurring mass requirements for surface transportation missions is fairly small in comparison to the mass-flow requirements to return crew to LEO and resupply the lunar base.

NASA/JSC (1989) provides parametric data on propulsive requirements for typical hopper missions. For trade comparison, a 3000 kg lunar hopper with a cryogenic propulsion system (I_{sp}=470 s) is estimated to perform a 1000 km range exploration mission every 90 days. Annual usage of propellants is only about 20 metric tons, which requires about 17 tons of LOX and 3 tons of LH_2 at a 6:1 mixture ratio. Supplying lunar-derived LOX instead of LOX

from the surface of the Earth provides a potential cost savings of \$419 to 420×10^6 annually, not including the cost of the LLOX facility.

VII. TRADE SENSITIVITIES

The trade data shown here assumes the use of a completely reusable LEV, and the use of a reusable LTV for LLOX usage for round-trip (LEO-LLO-LEO) missions. The conclusions stated above do not include the replacement of LEVs or LTVs as they wear out, nor additional mass requirements to provide replacement for life-limited items such as rocket engines.

As shown above, the primary driver for the cost effectiveness of the use of LLOX is the relative cost of launching payloads into LEO compared to the cost of developing and installing a production plant and supporting equipment on the lunar surface. If the LEO launch costs are significantly reduced over current costs, then the cost effectiveness of LLOX production will be questionable due to the signficant development costs projected for such a system.

The trade study performed only covers the time period of 1990–2020. If LLOX is found to be cost effective for this time period, then cost benefits may continue to accrue for a longer period of time. This has to be balanced off by the cost of money for developing and implementing the LLOX plant. Although not included in this top-level analysis, cost of money considerations allow the balancing of the investment/financing costs of implementing a LLOX facility development against the long-term cost benefits of LLOX production. To perform this calculation, the development cost of the LLOX plant, the relative development, production, and operations cost of the lunar transfer system, and the recurring operations cost of the LLOX production facility must be known as a function of time.

VIII. CONCLUSIONS

LLOX can provide significant benefits to lunar and Mars transfer systems if the supporting infrastructure is designed to use it. LLOX production on the lunar surface can provide significant cost savings over launching all propellants from the surface of the Earth at current launch costs. However, the cost benefit of using the LLOX is dependent upon the degree of utilization and the cost to develop and deliver the system. Figure 9 presents a summary of the cost effectiveness of use of LLOX, based upon the transportation demand model shown in Fig. 7. If LEO transportation costs drop by an order of magnitude as some have forecast, the cost benefit of using LLOX will be small unless the LLOX production system can be developed at low cost. Similarly, these conclusions are dependent upon the level of demand transportation to the Moon and Mars. If the demand for transportation greatly increases, the benefits derived from the use of LLOX will also increase.

REFERENCES

Bienhoff, D. G., and Skruch, G. J. 1990. Synergistic lunar lander for efficient operations. In *Engineering, Construction, and Operations in Space II: Proc. of Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 646–655.

NASA. 1989. *Report on the 90-Day Study of Human Exploration of the Moon and Mars* (Washington, D. C.: NASA).

NASA/JSC. 1989. Lunar hopper data. In *Lunar/Mars Common Vehicle Study* (Washington, D. C.: NASA/JSC Mission Support Directorate).

Repic, E. M., Horio, P., Biale, K., Bienhoff, D., Clemons, S., Duffy, J., Green, J., Hill, L., Jones, R., Jones, D., Kent, S., Kehrbaum, J., Marshall, M., McCullough, E., Skruch, G., Steller, D., Tapper, M., Waldron, R., Woo, H., Richter, P., and Drake, R. 1990. *To the Moon and Mars—Lunar/Mars Exploration.* Rockwell International Space Systems Div. Publ. 3547).

Schmidt, H. J. 1991. Chemical Piloted and Cargo Transfer and Excursion Vehicles for Scenario 4. Rockwell International Space Systems Div. Tech. Memo 385-110-91-044.

Stern, M. O. 1988. *Advanced Propulsion for LEO-Moon Transport*, California Space Institute, NASA Grant NAG-186.

Woodcock, G. R. 1988. Basing options for lunar oxygen for manned Mars missions. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 420–434.

PRODUCTION OF NON-VOLATILE MATERIALS ON THE MOON

R. D. WALDRON
Rockwell International

Potentially useful nonvolatile products can be produced from raw lunar rocks and soils or mass fractions thereof obtained by one or more physical or chemical processing operations. The feasibility and practicality of lunar manufacture of specific components or systems will be constrained by available materials properties, and complexity and capital requirements of processing and manufacturing facilities. Classifications of input materials and preprocessing and refining systems are presented and the differences in materials specifications required for surface and flight or export hardware applications are discussed. Rationale for selection(s) of refining processes are outlined. Generic descriptions of manufacturing operations applicable to metallic and nonmetallic hardware are reviewed. Specific examples of available candidate materials and applicability to provide structural, thermal, chemical, electromagnetic and optical properties needed for various applications classes are presented.

Regardless of foreseeable improvements in Earth-to-orbit transportation technology, the high cost of delivering useful cargo from Earth to the lunar surface will probably remain a primary economic driver toward developing an industrial capacity to manufacture or construct facilities and products primarily or solely from local lunar materials. The range of potential products and options for processing methods are extensive and the growth of industry will depend on both supply and demand considerations and constraints (base size, population, desired growth rates, support facilities, power, etc.).

The ultimate potentially accessible scope for products manufactured primarily from local resources will be governed by the materials properties of raw, semi-refined, or refined fractions of lunar soils or rocks with or without selected minor additions of import materials from Earth or elsewhere in space.

I. MATERIALS PROPERTIES REQUIREMENTS FOR LUNAR APPLICATIONS

There is a natural tendency to associate specific terrestrial materials with terrestrial consumer or industrial products because manufacturing practices and selections evolve in response to competitive factors on Earth. On the Moon, one must re-examine traditional terrestrial reasoning and search for the best materials for specific end applications, remaining cognizant of specific differences in lunar industries. In particular, one may note no mass market

economies of scale, lack of water, air, and organic chemical products, and importance of fire prevention or noncombustible products within habitats. While this process may lead to novel materials selections for specific products, it is important to utilize compositions that meet specifications for terrestrial alloys and nonmetallic materials where possible because these materials have well-documented properties which can provide required levels of product reliability.

The suitability of various materials for a variety of end applications can be evaluated in terms of intrinsic materials properties which can depend on purities, temperatures, processing histories and other factors. These properties may be classified into 5 major groups as shown in Table I.

TABLE I

Materials Properties Requirements for Lunar Applications

A.	Thermophysical Properties Melting point, density, vapor pressure, specific heat, thermal expansion, thermal conductivity, spectral and radiative properties vs temperature
B.	Mechanical Properties Tensile, shear, compressive, creep, and fatigue strengths, elastic moduli, hardness, elongation, reduction in area, ductility/toughness, abrasion resistance, etc.
C.	Chemical Properties Hot corrosion, cold corrosion, oxidation/reduction, response to reagent systems
D.	Dielectric/Magnetic Properties Resistivity, dielectric constant, loss tangent, magnetic susceptibility (permeability), coercivity, magnetization, hysteresis vs frequency and temperature
E.	Optical/Spectral, Response to Ionizing Radiation and Miscellaneous Properties

The first group (thermophysical properties) will broadly establish the temperature ranges of useful applications for respective soil fractions or derived elements or compounds. Class B properties will govern mechanical design for primarily structural members while classes C through E list properties of importance for specialty applications such as fluid containment, electromagnetic

components and optical or radiative transmission or shielding systems.

II. COMPOSITIONAL CLASSES ACCESSIBLE FROM LUNAR SOURCES

Materials processing in its restricted sense will generally involve a number of discrete steps from mining to finished feedstock which will be followed by manufacturing and assembly. The early processing steps excluding mining and materials handling may be designated as "preliminary processing" (PP) steps to distinguish them from chemical processing or certain other energy intensive operations. A mature materials industry on the Moon may evolve to an assembly of PP and subsequent steps as shown in Fig. 1. The PP steps may be further divided into preconditioning (PC), physical beneficiation (PB), and chemical beneficiation (CB) steps (Waldron 1980*a*). Specific examples are shown in Figs. 2, 3 and 4.

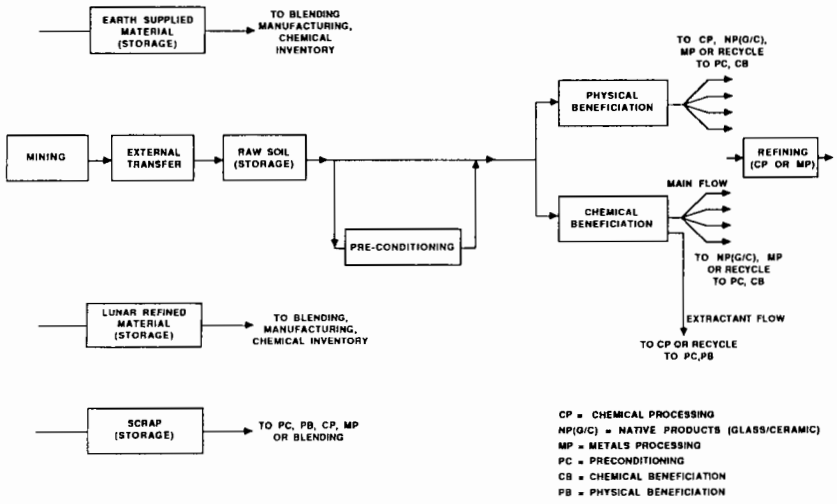


Figure 1. Flow diagram for integrated material processing.

One may classify potential lunar resource materials according to site location (or composition) and/or degree or extent of required PP prior to manufacturing. Because actual operations may likely involve blending of multiple source streams with varying levels of preliminary steps and partial or total recycling of step outputs to added PP steps, it is not productive to attempt a classification based on the generic nature of PP steps. Instead, we have chosen to classify lunar source materials into 6 groups according to increasing levels of PP complexity and specific energy requirements and for classes 4 through 6 on involvement of chemical refining or import element additions. This classification is shown in Table II.

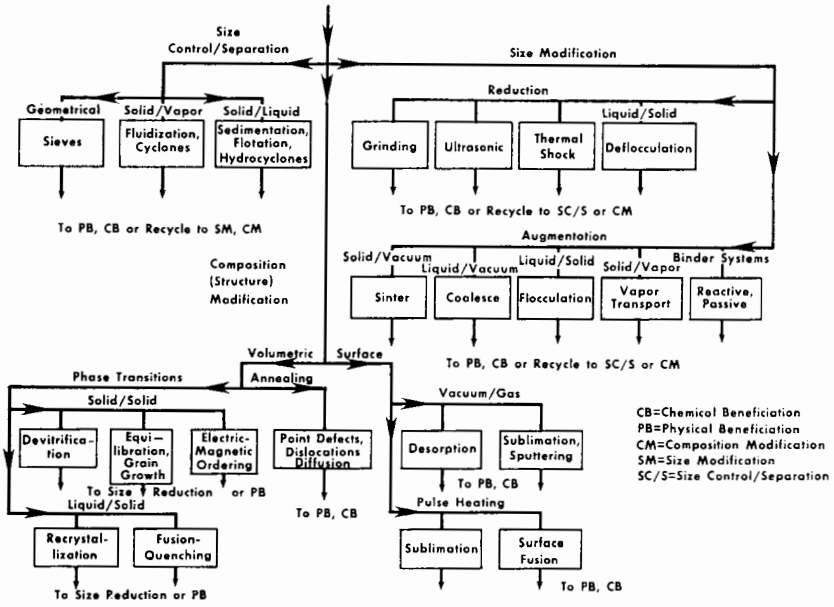


Figure 2. Classification of preconditioning (PC) methods.

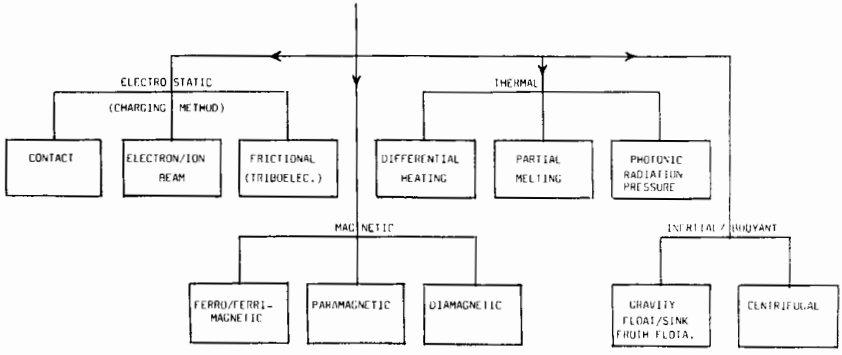


Figure 3. Classification of physical beneficiation (PB) processes.

The raw soils or rocks (Class 1) will have compositions characteristic of specific lunar sites—i.e., lunar mare locations have soil or rock compositions predominantly characteristic of lunar basalts, while highlands samples have mostly anorthositic components (Table III). Special locations of limited geographical extent such as dark mantled regions may have volcanic glasses or other special compositional or textural features.

Lunar soils possess varying levels of free metal particles, (usually >90 wt% iron with the balance mainly Ni and Co) averaging (for 79 soils) ~0.54%

TABLE II

Lunar Source Materials According to Increasing Preliminary Processing Complexity

1. Raw Soils or Rocks
2. Minimally Processed Soils and Rocks
   Cold or moderate temperature physical processing
3. Hot Processed Soils and Rocks
   No reagents or indigenous reagents processed at high temperature
4. Chemically Processed or Refined Components
5. Nonrefined Output Streams of Chemical Refining Operations
6. Modifications to Above Classes Using Import Material Additives

by weight (Morris 1980). For mature soils, typically 0.3 to 0.5% occurs in particles larger than 300 Å in diameter. Ambient temperature magnetic beneficiation may permit recovery of 0.2 to 0.25 wt% Fe from most mare soils (Oder 1991).

Class 2 or minimally processed soils or rocks will consist of raw samples which have been subjected to one or more PP steps involving ambient temperatures or moderately elevated temperatures (T_{max}<800°C) This temperature cutoff was somewhat arbitrarily chosen to include vacuum desorption of light elements H, C, N, and noble gases and exclude operations that would permit significant grain sintering (NASA 1988)

Class 3 or feedstocks subjected to hot processing operations above 800°C can be regarded as a source material class or alternatively as a manufacturing operation or general fabrication step to form sintered or cast ceramic parts. The high-temperature operation(s) may optionally include contacting the solid or liquid mineral phase(s) with solid, liquid, or gaseous reactants or extractants—primarily those derivable from lunar resources (O_2, Si, Fe, sulfides, or phosphides). These operations, called chemical beneficiation can preferentially extract various minor and trace element constituents from raw soils (Waldron 1980*a*) (Fig. 4).

Class 4, chemically processed or refined materials can use any of the preceding source classes as inputs. The relative production rates of specific refined outputs of major elements, oxides or silicides will be limited by the input compositions employed and especially by the refining process or processes selected. To obtain access to significant quantities of relatively pure feedstocks of all 7 major elements (Al, Ca, Fe, Mg, Si, Ti and O) or their binary oxides or silicides, one must either use a "whole soil" refining system such as fluoroacid (HF) leach process or carbochlorination, or alternatively use a simpler initial step process and develop multiple secondary refining steps to achieve desired separations. Either option will invariably require imported reagent fractions and quasi-closed cycle recovery and regeneration of reagents and/or solvents containing appreciable Earth import element fractions.

Limited production quantities of certain lunar minor or trace elements or their compounds can also be obtained from either whole-soil refining process

TABLE III

Compositions of Lunar Soils, Rocks, and Mineral Fractions

Average Compositions of Apollo and Luna Soils*								High-Titanium Basalts+				
								Modal abund.	Pyroxene	Olivine	Plagioclase	Opaques (mostly ilmenite)
R A-11	A-12	A-14	A-15	A-16	A-17	L16	L20	Vol% ⟶	42.60%	0–10%	15–33%	10–34%
								Component (Wt%)				
42.47	46.17	48.08	46.20	45.09	39.87	43.96	44.95	SiO_2	44.1–53.8	29.2–38.6	46.9–53.3	<1.0
13.78	13.71	17.41	10.32	27.18	10.97	15.51	23.07	Al_2O_2	0.6–6.0	—	28.9–34.5	0–2.0
7.67	3.07	1.70	2.16	.56	9.42	3.53	.49	TiO_2	0.7–6.0	—	—	52.1–74.0
0.30	0.35	0.22	0.53	.11	.46	.29	.15	Cr_2O_3	0–0.7	0.1–0.2	—	0.4–2.2
15.76	15.41	10.36	19.75	5.18	17.53	16.41	7.35	FeO	8.1–45.8	25.4–28.8	0.3–1.4	14.9–45.7
0.21	0.22	0.14	0.25	.07	.24	.21	.11	MnO	0–0.7	0.2–0.3	—	<1.0
8.17	9.91	9.47	11.29	5.84	9.62	8.79	9.26	MgO	1.7–22.8	33.5–36.5	0–0.3	0.7–8.6
12.12	10.55	10.79	9.74	15.79	10.62	12.07	14.07	CaO	3.7–20.7	—	14.3–18.6	<1.0
0.44	0.48	0.70	0.31	.47	.35	.36	.35	Na_2O	0–0.2	—	0.7–2.7	—
0.15	0.27	0.58	0.10	.11	.08	.10	.08	K_2O	—	—	0–0.4	—
0.12	0.31	0.50	0.11	.12	.07	.14	.11	P_2O_5	—	—	—	—
0.12	0.10	0.09	0.06	0.06	.13	.21	.08	S	—	—	—	—
51.0	45.0	79.6	63.6	56.0	59.6			H (PPM)				
60	10	8	8	6	36			He (PPM)				
135	104	130	95	106.5	82			C (PPM)				
119	84	92	80	89	60	134	107	N (PPM)				
206	189	321	146	345	131	174	208	Ni (PPM)				
32	43	35.8	54.4	25.3	35	37	40.5	Co (PPM)				

	Highlands Rocks+				Low-Titanium Basalts+				
Modal abundance	Pyroxene	Olivine	Plagioclase	Opaques (mostly ilmenite)	Modal abund.	Pyroxene	Olivine	Plagioclase	Opaques
(Vol%) ⟶	5–35%	0–35%	45–95%	0–5%	(Vol%) ⟶	42–60%	0–36%	17–33%	1–11%
Component (Wt%)					Comp (Wt%)				
SiO_2	51.10–55.4	37.70–39.9	44.00–48.0	0–0.1	SiO_2	41.2–54.0	33.5–38.1	44.4–48.2	<1.0
Al_2O_3	1.00–2.5	0–0.1	32.00–36.0	0.80–65.0	Al_2O_3	0.6–11.9	—	32.0–35.2	0.1–1.2
TiO_2	0.45–1.3	0–0.1	0.02–0.3	0.40–53.0	TiO_2	0.2–3.0	—	—	50.7–53.9
Cr_2O_3	0.30–0.7	0–0.1	0–0.02	0.40–4.0	Cr_2O_3	0–1.5	0.3–0.7	—	0.2–0.8
FeO	8.20–24.0	13.40–27.3	0.18–0.34	11.60–36.0	FeO	13.1–44.5	21.1–47.2	0.4–2.6	44.1–46.8
MgO	16.70–30.9	33.40–27.3	0–0.18	7.70–20.0	MnO	0–0.6	0.1–0.4	—	0.3–0.5
CaO	1.90–16.7	0.20–0.3	19.00–20.0	0–0.6	MgO	0.3–26.3	18.5–39.2	0.1–1.2	0.1–2.3
Na_2O	—	—	0.20–0.6	—	CaO	2.0–16.9	0–0.3	16.9–19.2	<1.0
K_2O	—	—	0.03–0.15	—	Na_2O	0–0.1	—	0.4–1.3	—
					K_2O	—	—	0–0.3	—

* Database Compilation.
+ Williams and Jadwick 1980.

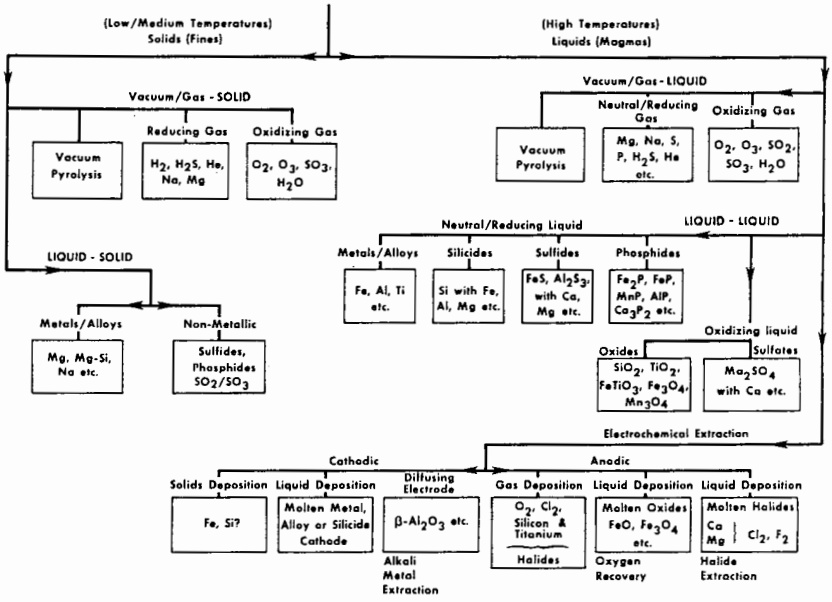


Figure 4. Classification of chemical beneficiation (CB) processes.

fractions or beneficiated feedstocks.

Class 5 represents potential source materials available in large amounts from chemical refining steps designed primarily to produce propellants or other purified outputs. These source materials will generally contain a mixture of other compounds or elements. For example, the H_2 reduction of ilmenite should produce a by-product output stream containing free iron, TiO_2, and varying levels of silicate minerals depending on beneficiation efficiency. Similarly, magma electrolysis should produce, in addition to O_2, 2 other output streams—(1) a low iron silicate slag, and (2) a metallic stream consisting primarily of ferrosilicon.

Class 6 represents modifications to prior class materials using Earth import additions. For example, additions of minor amounts of lunar deficient elements to alloys can significantly improve the properties and uniformity of metallic materials and products. When such additions are used, they will generally be made at the latest step possible to prevent excess attrition of these costly elements.

III. APPLICATIONS CLASSES

Nonvolatile materials applications may be divided into 3 classes based on end use location: (1) surface systems; (2) flight hardware; and (3) export materials/hardware. The implementation of specific applications will generally

evolve in the order listed. Table IV gives a general tabulation of potential applications for the above classes.

A. Surface Systems

In contrast to almost all previous space hardware design requirements, minimum mass, high technology materials and sophisticated design are relatively unimportant for surface systems designed for use at or near the point of manufacture. Product reliability standards may also be relaxed (where not life threatening) because local repair or replacement may also be assumed to be feasible in most cases.

For the ambient temperature applications, we may expect that the vehicular and pressure vessel applications would be met primarily using ductile metallic materials, while most of the remaining categories could employ brittle (non-metallic) materials with lower refining and power requirements.

The multi-temperature applications which would be desirable for base expansion would generally require control of critical parameters and normally dictate use of refined materials. A notable exception would be basalt or other lunar rock for heat storage purposes.

B. Flight Hardware

Flight hardware in contrast to surface systems would usually be mass sensitive and require premium performance materials of high specific strength and high reliability. For ambient temperature applications, premium light metal (Al or Mg) or titanium alloys would be useful as well as fiber reinforced composites. One can extract composition fractions from lunar soils to produce fused silica, "S" glass, and high modulus fibers or whiskers such as Al_2O_3 to reinforce metallic or nonmetallic matrix materials. Flexible fibrous or fiber-bonded cables can also be used for tether applications.

For higher-temperature applications such as aerobrake materials and some insulation needs, very high melting temperature materials such as MgO (mp 2852°C) or CaO (mp 2614°C) may useful. A somewhat higher melting temperature can be obtained by combining 20% Earth supplied carbon with 80% lunar titanium (TiC: mp 3140°C).

C. Export Materials/Hardware

Export of fluids, especially propellants, may be of great economic benefit in lunar industrial development. This may prompt development of lunar manufacture of propellant tanks and insulation for liquid oxygen (LOX) and various fuels. Lunar orbital propellant depots may prove useful to permit use of lunar LOX for powered descent to the lunar surface and avoid transporting fuels to the lunar surface which are needed only for trans-Earth injections for return to Earth orbit.

Lunar raw materials or manufactured parts may eventually be exported for additional processing or use elsewhere in space or even to the Earth's surface. Development of an extensive export trade will probably require

TABLE IV
Application Classes

1. Surface Systems

a. Ambient Temperature (Primarily)
Construction Materials
(Wires, cables, beams, strips, panels, piping, tubing)
Interior furnishings, floors, walls, ceilings, exterior paneling, solar shading, solar panel substrates, mounting and deployment hardware
Radiation Shielding
Pressure Vessels
(Fluids, habitat, work space)
Vehicle/Robots
Construction & Mining Equipment
(Rovers, trailers, mobile cranes, excavators)
Abrasives, Tooling
(tool fabrication, machine tools, welders, girders, etc.)

b. Multitemperature Applications
Processing Equipment
Heat Transfer Equipment
Power/Energy Storage Systems
Thermal Insulation/Refractories/Reservoirs
Electric/Magnetic Systems
(Conductors, electric motor, permanent magnets, solenoids, transformers, antenna systems)
Optical Systems

2. Flight Hardware

Pressure Vessels
(Fluids, cargo, personnel)
Insulation
Aerobrakes
Tethers, etc.
Reaction Mass

3. Export Materials/Hardware

Fine Particles
(Raw or refined)
Manufactured Items
(Propellant tankage, orbital hardware, feedstock for zero-G orbital manufacture, etc.)

implementation of transport methods not dependent on chemical propellants to reach lunar orbit and for transport to Earth orbit. These will probably involve external propulsion technology such as electromagnetic accelerators and/or laser or beam power systems.

Very high efficiency transport of discontinuous streams of raw or beneficiated lunar soil fractions to lunar or Earth orbit can be achieved by hypervelocity impact and collection of low altitude, low kinetic energy lunar particles by vehicles in cyclic Earth-Moon trajectory orbits (Waldron 1988*a*). Development of such technologies could establish the most efficient means of supplying lunar LOX to low Earth orbit by manufacturing LOX in Earth orbit from raw lunar ores delivered from the Moon.

IV. CLASSIFICATION OF PROCESSING OPERATIONS

A. Preliminary Processing

The classification of PP steps has already been discussed in Sec. II. Such steps conducted at low or elevated temperatures in which process feeds or reactants have negligible vapor pressure may be conducted either in lunar vacuum or in a pressurized facility under a suitable oxidizing, neutral, or reducing atmosphere, Unfortunately, no abundant source of reduction gas composed entirely of lunar accessible elements is currently known. Very low pressure hydrogen may be useful for some lower temperature operations with relatively little import mass penalty, and sodium vapor for steps with $T > 400°C$ (vp>0.4 Torr) may be a viable option. Abundant supplies of Mg vapor are potentially available with a vapor pressure >0.1 Torr at $T > 510°C$. Physical beneficiation steps, particularly magnetic and electrostatic beneficiation, and the preconditioning step of thermal desorption can be expected to provide useful concentrates of various minor trace and major soil or rock constituents. Table V shows some anticipated concentrates available from raw soil or pretreated soils or rocks.

B. Chemical and Electrochemical Refining Processes

Chemical refining processes may be defined as a series of processing steps designed to treat specific input fractions of various input classes of lunar materials (Table II) and provide two or more output stream flows. The processing system will require an inventory of reagent/solvent and/or other specialty items subject to wear or attrition. In most cases these inventories will involve some substance containing significant mass fractions of lunar deficient elements (LDEs), i.e., elements not readily recoverable in significant amounts from lunar source materials.

Depending on the nature of such inventories, we may classify chemical refining processes into 3 groups as follows:

These classes may also be used to describe the initial main step of a processing system. Most processing systems proposed to date belong to Class

Class I	Nonreagent processes;
Class II	Processes containing one or more steps employing reagents derivable solely from lunar accessible elements;
Class III	Processes containing one or more steps employing reagents containing LDEs.

III, for which it is necessary to provide quasi-closed cycle reagent/product recovery and regeneration to avoid excessive mass import requirements.

Prior attempts to classify entire processing systems according to some generic combinations of types or methods of chemical and/or physical interactions have shown that the vast number of possible options for processing steps precludes a systematic or exhaustive analysis or comparison of potential systems (Waldron 1978,1988*b*).

A compromise classification method based on chemical or operational similarities of first main step was adopted for a previous lunar process survey paper (Waldron 1988*b*) that listed virtually all methods previously proposed in the literature. An updated version of this classification is shown in Table VI.

A generic description of the first stage products of the initial process steps appears in the table. These products generally have the recoverable oxygen transferred to fluid or volatile compounds (or elemental O_2) and two or more by-product elements or compounds which may have varying degrees of purity and separability. In addition to the processes shown, secondary processes may be developed to process nonvolatile output streams of the primary processes or beneficiation steps previously listed.

Considerable versatility in metals production can be achieved by purification of iron and separation and recovery of other minor and trace element transition metals (Cr, Mn, Ni, Co) present in chemical refining step outputs or beneficiated lunar free metals. These impure metals can be treated by electrorefining processes using auxiliary chemical or ion exchange separations (Waldron 1980*c*), or by production and separation of metal carbonyls (Lewis 1983).

V. METHODS OF COMPARATIVE EVALUATION OF ALTERNATIVE PROCESSES

The rationale for process selection may depend on several factors including relative economic values of propellant manufacture vs nonvolatile product fabrication and primary and secondary objectives of lunar-base operations and growth. It may prove advantageous to adopt a simple, robust process primarily for early oxygen propellant production even though it restricts the scope of nonvolatile products which can be produced initially. In such cases, evolutionary development may lead to secondary refining processes that use secondary output streams of the initial process as input streams for subsequent

TABLE V

Some Anticipated Concentrates Available from Raw Soil or Pretreated Soils/Rocks

Magnetic Beneficiation		
Fractions	Raw Soil	Pretreated Soils
Ferro-ferrimagnetic	Native Fe agglutinates	Magnetic ferrospinels
Paramagnetic	Ilmenite, Ti Cr spinels olivine, Pyroxene, most glasses	Fe, Ti oxides
Diamagnetic	Feldspar, silica, iron-poor glasses	Feldspar, silica, Fe-poor pyroxene
Electrostatic Beneficiation		
Fractions	Raw Soils	Pretreated Soils
Highly or partially conductive	Native Fe, high-Fe minerals (depending on temperature)	Fe_3O_4 ferrospinels (elev. temperature)
Dielectric (negative charge)	Pyroxene, anorthite (weakly −)	Silica
Dielectric (positive charge)	Olivine	
Vacuum Desorption		
Fractions	Raw Soil	Pretreated Soils
Intermediate temperature (T <800°C)	H compounds C compounds N_2, noble gases	
High temperature	S, P, coumpounds (depending on O_2 chemical potential Na, K)	

Note: H compounds are those that contain hydrogen; C compounds are those that contain carbon.

TABLE VI

Process Classification Based on Initial Major Step

Generic Class	Process/Class		Reagents			First Stage Products	Tech. Perf	Reliability/Availability/ Longevity						Import Mass		References
	No.	Description	Class	Cyclic	Lunar		P	T	HC	CC	RL	RR	MH	EE	MBR	
A		THERMAL VAPORIZATION														
	A1	Sublimation	1		LS	MO, SiO_2	X	S						L		1,2
	A2	Pyrolysis	1		LS	MO, M', Si, O_2	X	S						L		3
	A31	Chem. Vapor Deposition	3	X_2,HX	LS	MX_z, O_2, H_2O	X		X		X	?		L-M		4
	A32	Plasma	3	O_2?X_2, H_2	LS	MO_z, MX_z, MH_y	X	X	X		?	?		L		5
B		PYROCHEMICAL REDUCTION														
	B3	Hydrogen Redn.	3	H_2	MQO_3	MO, H_2O, M', QO_2	X				X		X	L-M?	S	6,7
	B4	Carbothermal Redn.	3	C	LS	M, Si, CO	X	S			X	X	X	M		8,9
	B5	Hydrocarbon Redn.	3	CH_4	LS	Si, M', MO, CO, H_2O	X				X	X				10
	B6	Silicothermal Redn.	3	Si	LS	M', MO, SiO_2	X									8
	B7	Sodium Redn.	3	Na(K)	LS	Na_2QO_3, M', MO, Si	X		X							11
	B8	Magnesium Redn.	2	Mg(Ca)	LS	M, Si, MgO(CaO)	X		X							11
C		PYROCHEMICAL OXIDATION														
	C9	Fluorination	3	F_2	LS	MF_2, SiF_4, O_2			X	X	X	X		L		12
	C10	Fuoro-oxidation	3	$KBrF_4$	LS	$MF_2, SiF_4, KF, KBr, O_2$			X	X	X	X		L		13
	C11	Magma Partial Oxidation	2	O_2	LS	$MSiO_3, Fe_3O_4, SiO_2$			X					M		14
D		ELECTROCHEMICAL REDOX														
	D12	Magma Electrolysis	1		LS	M', O_2, SiO_2, Si	X		S		EL*			?		15,16

E		AQUEOUS											
	E13	Fluoroacid	3	HF/H_2QF_6	LS	MF_2, MQF_6, QF_4, H_2O			X	X		M		17,18,19
	E14	Sulfuric/Sulfurous	3	$H_2SO_4(SO_2)$	LS	SiO_2, $MSO_4(MSO_3)$, H_2O	X	X				H		18,20
	E15	Hydrochloric	3	HCl	LS	MCl_2, SiO_2, H_2O			X	X			18,20	
	E16	Fluoboric	3	HBF_4	LS	$M(BF_4)_2$, SiO_2, $HBOF_x$,H_2O			X	X				18
	E17	Nitric	3	HNO_3	LS	$M(NO_3)_2$, SiO_2, H_2O			X	X	X			18,20
	E18	Alkaline	3	NaOH	LS	Na_2QO_3, $NaAl(OH)_4$, H_2O, MO	X			X				18
F		NON-AQUEOUS, ELEV. TEMP.												
	F19	Fused Alkali	3	$NaOH(Na_2CO_3)$	MQO_3	Na_2QO_3, $NaM''O_z$, MO, $H_2O(CO_2)$	X	X						18
	F20	Na_2O	2	Na_2O	MQO_3	Same as F19 (exc, no $H_2O(CO_2)$)	X	X						18
	F21	Anhydrous Acids	3	H_2SO_4/H_3PO_4	LS	SiO_2, MSO_4 or $MHPO_4$, H_2O	X	X		X				21
	F22	Fused Acid Salts	3	NH_4FHF, NH_4HSO_4	LS	Same as E13 or 14 (plus NH_3)	?	X		X	X			17,18
	F23	Carbochlorination	3	C, Cl_2	LS	MCl_2, $SiCl_4$, CO		X	X	X	X			23
	F24	Fused Fluorides	3	MF_2	LS	SiF_4, MF_2, Mo_xFy	?	X		X	X			22
	F25	Lime Sinter	3	CaO	LS	$CaSiO_3$, MO	X	X				X		20
G		MISCELLANEOUS												
	G26	Supercritical Steam	3	H_2O	LS	H_2O, $Si(OH)_4$, MO	X	X						18
	G27	Other												

LS=Lunar Soil ($MSiO_3$, M_2SiO_4); M=Fe, Mg, Ca (2/3 A1); M′=Fe; M″=Al, Fe; Q=Si, Ti; X=Halogen; S=Severe: X=Nominal; T=Thermal resistance; HC=Hot corrosion; CC=Cold corrosion; P=Output stream purity; RL=Reagent loss; RR=Reageant regeneration; EL*=Electrode loss; MH=Materials handling; EE=Energy efficiency L, M, H (low, med, high); MBR=Mining, beneficiation, requirements

References: 1. Agosto 1983; 2. Grimley and Lipschutz 1984; 3. Steurer 1985; 4. Lynch 1989; 5. Lynch et al. 1989,1990; 6. Dalton 1972; 7. Downs 1971; 8. Phinney et al. 1965; 9. Cutler and Krag 1985; 10. Rosenberg et al. 1965; 11. Waldron 1983; 12. O'Donnell 1972; 13. Roberts and Roberts 1966; 14. Waldron 1989; 15. Oppenheim 1970; 16. Colson and Haskin 1991; 17. Waldron 1980*a*, *b*, *c*, *d*; 18. Waldron 1978; 19. Waldron 1985; 20. Peters and Johnson 1974; 21. Waldron 1984–89; 22. Jarrett et al. 1980; 23. Rao et al. 1979.

refinement. Alternatively, second generation processing systems of "whole soil" refinement type may be introduced in parallel with the original plant to provide propellant capacity expansion as well as desired nonvolatile product capability.

A third class of processes may evolve to permit useful salvage of product source materials from either imported or lunar indigenous scrap material which would otherwise accumulate near lunar bases.

Comparative evaluation of alternative processes or approaches for lunar materials industry operations cannot be usefully conducted without a philosophical basis or generalized scenario for lunar-base development. There are two limiting cases which may constrain lunar-base development: (1) base evolution governed by a prescribed activity schedule where activities are selected by a benefit: cost evaluation subject to technology driven cost reductions based on *in-situ* manufacture; or (2) base evolution governed by an overall budget or affordability determination or schedule with an activity scope subject to technology driven expansion based on import mass leverage due to ISMU. The first case might be characterized as fixed benefit, cost reduction approach, while the second as an expanding benefit, fixed cost approach. The second option is more likely to evolve into a true expansion mode with substantial private enterprise involvement. Substantial industrial diversity appears to be necessary to produce economic synergism of various facets of the materials economy. Regardless of the philosophy chosen, the actual development of materials industry may be expected to evolve through phases limited by capabilities of support services as shown in Fig. 5.

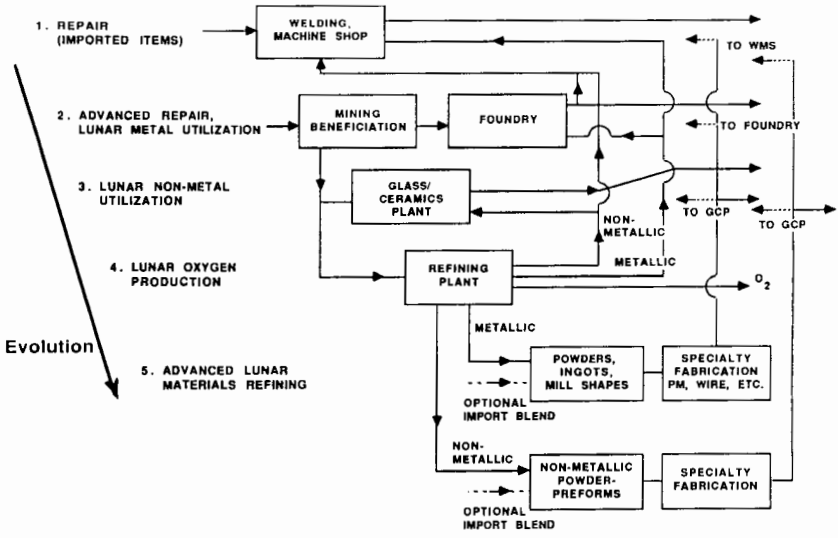


Figure 5. Time sequence of imported and lunar materials derived hardware and processes.

A rational method for comparative evaluation of alternative processes would be based on benefit: cost considerations for all major products including fluids/volatiles (propellants) as well as nonvolatile products. The criteria for process selection should be based on such factors as size of the mass market of imported equivalents, system specific throughput (payback time), ease of manufacture, reliability, maintainability, repairability, energy intensitivity, automation/manpower adaptability, etc. Unfortunately, quantitative measures of these specific performace parameters are not available for most proposed processes (or their individual steps). Reliability can only be confidently predicted for process steps adapted with at most minor variations from commercial terrestrial practice or by expensive and time consuming statistical testing of new process subsystems. A subjective estimate of 6 factors bearing on process reliability of the various candidate processes is included in Table VI.

VI. MANUFACTURING METHODS AND REQUIREMENTS

A generic classification of manufacturing methods according to the nature of kinematic, shaping, joining or assembly operations is given in Table VII. Each of the general or specific manufacturing operations requires some particular forms of input material adapted to the specific procedures and these can impact the preceding or processing steps. It is beyond the scope of this chapter to exhaustively analyze the myriad examples of possible manufacturing operations, but we shall attempt to survey certain methods of general utility of providing useful products for diverse requirements at low production capacities. Some of the generic processing classes of Table VII, particularly most hot-forming operations, are applicable to either brittle or ductile materials while most cold deformation methods can only be used on ductile materials.

A. Metals

Metallic components or hardware can be fabricated from various preforms depending on the fabrication method selected. These may include various processes shown in Table VIII.

B. Nonmetals

This group includes ceramics, glasses and, in the terrestrial case, polymers and elastomers. In the lunar case, the members of this group are generally brittle at ambient temperatures which restricts the use of most cold-forming methods. However, at sufficiently high temperatures ($T > 75$ to 80% of the absolute melting point), most ceramics will yield under suitable conditions and permit extrusion, compression forming, etc. Glasses have a unique suite of plastic forming operations such as spinning, blow molding, many types of hot deformation, float forming of sheet, etc.

TABLE VII
Generic Classification of Manufacturing Processes

1. Kinematic
   K1 Fiber operations (felting, paper, textiles)
   K2 Particulate operations (mixing, conveying, custering, dispersing)
   K3 Fluid operations (mixing, pumping, heating, cooling, phase separations)
   K4 Discrete operations (small scale assembly)

2. Forming
   F1 Hot forming (casting, forging, powder metallurgy, sintering, hot rolling, bending)
   F2 Cold forming (extruding, bending, punching, drawing, machining, cold molding)
   F3 Unconventional forming (electromagnetic)

3. Surface Treatment
   S1 Removal (washing, etching, electropolishing)
   S2 Addition (coating, plating, anodizing)
   S3 Modification (bonding, hardening, shot peening)

4. Internal treatment
   I1 Internal heating (resistance, induction, dielectric)
   I2 Irradiation (X-ray, gamma ray, electron beam, UV)
   I3 Miscellaneous (magnetic poling, ultrasonic)

5. Bonding
   B1 Hot processes (welding, brazing, soldering)
   B2 Cold process (adhesives, chemical bonding, ultrasonic welding)
   B3 Unconventional processes (explosive, electromagnetic)

6. Large-scale assembly
   L1 Reduced gravity (lunar surface)
   L2 Microgravity (orbital assembly)

C. Textural Assemblies/Composites

Special geometric forms of various materials including fibers, whiskers, flakes, and foams can be useful to provide important properties for various applications including composite materials structures. Metal wires and glass or ceramic fibers would be highly useful in any diverse materials industrial economy and should receive a high priority. While composite materials (mainly metal or ceramic/glass matrix systems) would be more important for mass sensitive flight hardware, wires and fibers would facilitate manufacture of electrical hardware, furnace, habitat, and cryogenic insulation, fabrics and

TABLE VIII
Fabrication Methods Classification

Class	Description	Preforms	Other Requirements	Applicable for Metals	Applicable for Nonmetals
F1	Hot Forming				
	Casting (gravity or centrifugal) sand, investment, die	Ingot/shot	patterns, molds	X	X
	Sintering (ceramics, powder metal) cold pressed, hot pressed directional or isostatic	Powders	Molds, foils (binders)	X	X
F1, 2	Pressure Deformation (Hot or Cold)				
	Bending, compression molding, forging, extrusion, swaging, wire drawing, etc.	Ingot/strip, billet/rod plate	Die(s), lubricants	X	HT
S2	Surface Deposition				
	Vacuum evaporation, sputtering, CVD	Strip/shot wire	Carrier gas (CVD)	X	X
	Electroforming, electroplating	Anode strip shot	substrate, mandrel electrolyte	X	
S2	Surface Removal				
	Machining, grinding, chemical milling, EDM	Near net part	Tooling, reagent (CM) fluid	X	X
F3	Miscellaneous				
B3	Impulse (explosive, electromagnetic)	Sheet/plate, etc.	Chem. explosive/pulse forming network	X	HT

upholstery, air and water filters, etc.

VII. SPECIFIC EXAMPLES

This section will describe a series of applications classes of nonvolatile products which can be produced primarily from lunar materials with varying degrees of processing complexity.

A. Products Derivable from Lunar Metals

The source materials may include natural lunar free metal (Fe with <10% combined Ni + Co), refined lunar free metals, metals refined from lunar silicate or oxide minerals (Al, Ca, Fe, Mg, Ti and in lesser quantities, Cr or Mn) and alloys derived from the above sources plus optional minor fractions of Earth-imported elements. Terrestrial alloys are generally classified by the element of greatest mass fraction—on that basis, large-scale production of alloys based on the 5 major lunar metallic elements (Fe, Al, Mg, Ti and Ca) and limited production of accessible minor or trace element based alloys (Cr, Mn, Ni and Co) appears feasible if adequate recovery and purification processes are available. No significant terrestrial commercial alloys have been develped based on 3 of the preceeding 9 elements, namely, Ca, Cr and Mn. (Chromium is useful for electroplating, however.) Except for Ca, all of the above elements are useful for minor alloy constituents and elemental calcium may be useful for electrical applications.

Table IX lists commercial alloys derivable from these sources while Tables X through XII list compositions and properties of alloys of Al, Mg, Ti and Fe (steels) which can be made predominantly or solely from lunar materials. The numbers at the bottom of the alloy section of Table IX give the number of alloys producible for the specific columns indicated and the total number of terrestrial alloys of the respective element base. It is worth noting that a number of commercial alloys currently obsolete on Earth may be particularly useful on the Moon, especially where no imported element additions are required.

As previously noted, the most favorable properties are not usually required for surface applications so conservative designs based on alloys derivable solely from lunar materials (except possibly C and Mo in steels) are usually preferred. Also fabrications from lunar free metal sources are less energy intensive and simpler to produce than systems using reactive metals derived from oxide or silicate materials.

Iron-base alloy systems (steels) can undoubtedly provide the most versatile range of materials properties of any element by suitable alloying, mechanical and heat treatments.

The ductility or notch toughness of many ferritic (body centered cubic) steels are inferior at cryogenic temperatures which might be reached during the lunar night for certain exterior structures. In such cases, compositional

TABLE IX

Potential Output Materials Derivable Solely or Predominantly from Lunar Sources

Lunar Elements Only[a]									Material Class	Lunar Elements Plus 5% or Less Earth Imports[*]						
High Capacity					Limited Capacity				Structural Elements	High Capacity				Limited Capacity		
Al		Mg	Fe	Ti	Cr	Mn	Ni	Co	Metals	Al	Mg	Fe	Ti	Cr	Ni	Co
Wrgt	Cast	AM100	1020	99.2	High Cr	Aust. Mn Steels	Z-Ni	H 150	Alloys	7075		4042	6–4	SS	Incoloy 807	Stellite 6B 6%
										7178	AZ80A	4340	Al/V		Incoloy 825	
EC	413	M1A	1095	99	St. Steel		Permalloy					440C				
1060	B443	A3A	1340	Ti-	410–430			Permendur		7090		8640	5–2.5 Al/Sn	446	Inconel 706, X750	2V Permendur
1100	514	AM60	5140	8Mn	nichrome					7091		07				
3003	520	AS21	A242	4–4			200,201			W049	ZK60		7–4 Al/Mo			
5005	356	AS41		Al/Mn			211,212				6%					
5050	360		9260				D301			7–9%	LDE		6-2–4-2 Al/Sn			
5052			501				Inconel			LDE						
5056	535						600,702						Zr/Mo			
5083							721,722						2.5–8% LDE			
5086																
5154																
5457																
6063																
6101												286				
6151																
	166	6	105	4			8	2	number of alloys b,b′	137	18	575	14		34	2
	384	39	794	45			105	37	total number of alloys	384	39	794	45		105	37

TABLE IX (cont.)

Al_2O_3 in Al, Mg, Fe, glass in Mg, Ti_5Si_3 in Ti	Al_2O_3 in Ni, SiO_2 in Ni	Reinforced Metals		
Cast basalt, dark glass, foamed glass		Structural non-metals		
Al_2O_3, CaO, MgO, TiO_2, SiO_2, spinels, mixed	Cr_2O_3, K_2TiO_3	Thermal materials		
ceramics, "S" fiber, Ti_5Si_3		refractories, insulation, fibers		
		Electric/Magnetic materials	d	d
Fe, Al, Mg, Ca		(a) conductors		
Kanthal A-1	Ni-Cr	(b) resistance alloys		
Si	AlP, FeS_2, NiO, CoO	(c) semiconductors		
same as thermal (except Ti_5Si_3) + titanates		(d) dielectrics, insulation		
Fe, Si-steels (M15, M5–8)	Permalloy,	(e) magnetics		SiC(30%), TiC(20%)
Fe_3O_4, $MeFe_2O_4$, sendust	Permendur, CrO_2			
Fe_3O_4, TiO		(f) electrodes		
same as retractories (except CaO) + garnets		Abrasives		SiC (30%), TiC (20%)
O_2, O_3	SO_2, SO_3	Fluids/Volatiles	H_2O(11%),H_2O_2(6%)	H_2S(6%)H_3P(9%), NaOH
	CrO_3	Cryogenic, ambient mp <500°C	H_2SO_4, H_2SO_3, H_3PO_4	
CaO, CaO_2, MgO_2, P_2O_5, MnO_2	Ca, Mg, Al, Fe, sulfates, phosphates	Chemicals/Reagents	Li, Be, B chemicals	NaH
	chromates, Na/K	volatiles, mp>500°C		

* Except where noted.
[a] Except C(<1%) for Fe base alloys.
[b] Number of alloys producible from lunar elements only.
[b1] Number of alloys with <5% LDE.
[c] <1% LDE.
[d] Thousands of metallic and non-metallic materials possible.

TABLE X

Properties of Carbon and Low Alloy Steels

	Composition					Tensile Strength	Yield Strength	Elongation	Bhn
AISI No.	C	Mn	Si	Cr	Other	MPa	MPa	(%)	Hardness
1020 HR	0.2	0.3				415	240	30	
CR						550	415	18	
1095 HT	1.0					750	670	10	
1340 HT	0.4	1.75	0.3			1560	1420		448
4140 HT	0.4	0.9	0.3	1.0	0.2 Mo	1550	1430	10	426
5140 HT	0.4	0.8	0.3	0.8		1600	1450	11	448
HSLA									
A 242	0.15	1			0.2 Cu min.	480	345	18 min.	
A 588 D	0.15	1.2	0.7	0.7		485	345	18	
A 656	.18	1.65	0.6		V, N	620	550	12	
970 X	0.26 max.	1.65 max.			Nb, V, N	590	485	14 min.	

TABLE X (cont.)

Properties of Medium Alloy Steels

	Composition					Tensile Strength	Yield Strength	Elongation	Bhn	Conductivity
AISI No.	C	Mn	Si	Cr	Other	MPa	MPa	(%)	Hardness	(%IACS)
501 stainless	0.1 min.	1.0 max.	1.0 max.	5.0		480–790	205–620	20–28	160–240	
4042	0.4	0.8	0.3		0.25 Mo	1590	1450	12	448	
4340	0.4	0.7	0.3	0.8	1.8 Ni 0.25 Mo	1720	1590	9	484	
8640	0.4	0.9	0.3	0.6	0.5 Ni 0.2 Mo	1650	1520	10	472	
9260	0.6	0.9	2.0	0.3		910–2450	460–2275	14–15	27–64 RC	
Tool steels										
O7	1.2	1 max.	0.6 max.	0.8	0.4 V,max 0.3 max 1–2 W				58–64 RC	
L2	.45–1	.5	0.5 max.	1	0.2 V	710–2000	510–1790	5–25	45–62 RC	
L6	.7	0.5	0.5 max.	0.75	0.25 V (opt) 1.5 Ni	655–2000	380–1790	4–25	45–62 RC	
S2	0.5	0.4	1		0.5 V,max 0.4 Mo, .3Ni max.				50–60 RC	
S6	0.45	1.4	2–2.5	1.3	0.4 Mo,.3 V				50–56 RC	
Electric Steels										
M15 (Tr 58)	0.07		4			434	345	4	65 RB	2.5–3.8
M5–8 (Or 60–80)	0.01		3			415			65 RB	3.3–3.8

Properties of High Alloy Steels

AISI No.	Composition C	Mn	Si	Cr	Other	Tensile Strength MPa	Yield Strength MPa	Elong-ation (%)	Bhn Hardness	Conductivity (%IACS)
Stainless										
410	0.15	1.0 max.	1.0 max.	12.5	0.5M Ni	515–795	275–550	23–30	150–212	
430	0.12 max.	1.0 max.	1.0 max.	16.0		515–630	275	15–35	160–190	
440 C	1.1	1.0	1.0	17.0	0.75 Mo	760–2000	450–1720	2–12	19–57 RC	
446	0.35 max.	1.0 max.	1.0 max.	25.0	.025 Mo	590–790	380–720	25	180–200	
T8	0.8	0.3	.3	4.0	14 W, 2 V, 5 Co, 0.75 Mo				63–65 RC	
Kanthal A-1 (Heat ox. resisting alloy				22.0	5.5 Al, 0.5 Co	635–1035	440–635	12–20	200–260	1.2
Sendust			9–10		5.0 Al					2.2

TABLE XI
Properties of Light Metal Alloys

Alloy	Composition Al (+%)	Mn	Mg %	Si	Cr	Tensile strength MPa	Yield strength MPa	Elongation (%)	Bhn[a] hardness	Shear strength MPa	Fatigue limit MPa	Electrical conductivity (%IACS)
						Wrought Aluminum						
EC	99.45					83–186	28–165		na	55–103		62
1060	99.60m			0.25M		69–131	28–124	6–43	19–35	48–76	21–45	62
1100	99.0m					75–150	26–115	5–45	23–44	62–90	34–62	57–59
3003		1.2		0.6M	(0.12Cu)	110–200	42–185	4–40	28–55	76–110	48–69	40–50
5005			0.8	0.3M		124–200	41–186	4–30	28–51	76–110		52
5050			1.4	0.4M		145–220	55–200	6–24	36–63	105–138	83–97	50
5052			2.5	0.25M	0.25	195–290	90–255	7–30	47–77	125–165	110–140	35
5056		0.1	5.2		0.1	290–414	152–345	10–35	65–105	179–221	138–152	27–29
5083		0.7	4.5	0.55	0.15	290–345	145–283	16–22	na			29
5086		0.5	4.0	0.4M	0.15	260–325	115–255	10–22	65–86	115–130		31
5154			3.5	0.45M	0.25	240–330	117–269	10–27	58–80	152–193	117–145	32
5457		0.3	1.0	0.08M		130–205	50–185	20		85–125		
6063		0.1	0.7	0.4		90–290	48–269	9–22	25–95	69–186	55–69	50–58
6101			0.5	0.5	0.03M	221	193	15	71	138		56
6151		0.2M	0.6	1.0	0.25M	330	298	17	100	221	76	42–54

High Strength Commercial Alloys

	Zn	Cu	Mg		Cr							
7075	5.5	1.5	2.5		0.3	118–572	103–503	11	150	317–331	159	30
7178	6.8	2.0	2.7		0.3	605	505–540	10–11		295–305	200–290	31
Weldalite 049		5.4	0.4	1.3Li, 0.4 Ag 0.14Zr		484–713	331–692	5.3				

PM Alloys

7090	8.0	1.0	2.5		1.62 Co	675	640	9–10				
7091	6.5	1.6	2.42		0.37 Co	640	600	11–13				

Cast Aluminum

		Mn	Mg	Si	Ti							
413		0.35M		12.0		296		3.5	na	172	130	31
B443				5.0		130–230	62	8–10	40–45	110	55	37–42
514			3.8			145–170	84–95	9	50	138	48	35
520			10.0		0.25M	330	180	16	75	235	55	21
356		0.35 M	0.3	7.0	0.2 M	262	186	2–6	60–80	207	90	39–43
360		0.35 M	0.5	9.5		324	172	3	na	200	124	37
535		0.2	7.0		0.2	275	140	9–13		207	62	

TABLE XI (cont.)
Properties of Light Metal Alloys

Magnesium Alloys												
		Composition				Tensile strength	Yield strength	Elongation (%)	Bhn[a] hardness	Shear strength	Electrical conductivity (%IACS)
Alloy	Type	Al	Mn	Ca	Th/Zn						
AM100A	Cast	10.0	0.1			150–275	83–150	1–10	52–69	125–145	10–12
MIA	Wrought		1.2	0.09		230–255	125–180	7–17	42–54	115–125	34.5
A3A	Wrought	3.0				248–290	152–230	15–21	46–73	130–160	
HM21A	Wrought		0.5		2 Th	235	170	10		125	33–34.5
HM31A	Wrought		1.2		3 Th	290	230	10		150	26
AZ81A	Cast	7.5	0.15		0.7 Zn	275	83	15	55	125	12
AZ91C	Cast	9.0	0.2		0.7 Zn	275	145	5	70	145	10–11.5
"High Strength" Commercial Alloys											
ZK60	Wrought			(0.45M Zr)	5.5% Zn	300–365	215–300	11–16	65–82	165–180	29–31
AZ80A	Wrought	8.5	.5		0.5 Zn	330–380	230–275	6–11	67–80	150–165	10.6

Note: m=minimum; M=maximum; na=data not available.
[a] Dependent on temper; +500-kg load, 10-mm ball.

TABLE XII
Mechanical Properties of Titanium Alloys

Alloy		Alloy Type	Tensile Strength MPa	Yield Strength MPa	Elongation (%)	Electrical Conductivity (%IACS)
99.2% Ti	Grade 2	α	400	276		—
99.0% Ti	Grade 4	α	620–758	517–620	19–23	3.1–3.5
Ti–8Mn		$\alpha\beta$	965–1100	895–1070	14–22	1.9
Ti–4Al–4Mn		$\alpha\beta$	1000–1035	860–1035	na	1.2
Ti-5Al-1.5Fe-1.4C-1.2Mo		$\alpha\beta$	1060–1350	1000–1275	15	1.0
High-Strength Commercial Alloys						
Ti-8Al-1Mo-1V		nrα	900–1000	830–951		
Ti-4Al-3Mo-1V		$\alpha\beta$	1380	1205	4–7	
Ti-13V-11Cr-3Al		β	1310–1650	930–1515	2–15	

modifications or design or operational solutions can eliminate serious problems.

Steel properties can benefit substantially from small additions of C, Mo, and various other Earth-imported elements. For example, the yield strength of heat treated AISI 4140 alloy is about 6 times larger than for AISI 1020 at room temperature but requires about 0.6% combined C+Mo (imported mass).

For applications where specific electrical, magnetic or thermal properties are paramount and mechanical properties are less important, one may use a variety of metallic as well as nonmetallic materials for *in-situ* manufacture. Thus, for general purpose electrical wiring, Al, Mg, or Ca are satisfactory for exterior AC conductors. While calcium is seldom if ever used terrestrially for electrical purposes due to moisture sensitive corrosion, it has about 30 to 40% better conducitivity than Al or Mg based on mass with thermal and mechanical properties comparable to Al (Table XIII). Steels are generally unsatisfactory for AC conductors, but may provide the best option for DC transmission lines and busbars.

Magnetic materials applications can be divided into soft (high permeability) and hard (high coercivity) property requirements. The soft magnetic materials can be further classified according to useful frequency ranges of AC operation. Metals are most frequently employed in low-frequency systems such as transformer and motor cores or laminations and for permanent magnet systems where minimum volumes are needed. Ceramic alternatives to the metal products are available for most applications. The most useful soft magnetic materials are silicon steels such as M8, M5 and sendust (containing Al) and for premium permanent magnets, the Alnico series (Table XIII).

Electrical materials for high-temperature resistance heating can be met in most cases using high Cr steels, particularly Kanthal A-1 (Kanthal Corp.) which can operate in vacuum or air to temperatures of 1150° or 1300°C, respectively.

Other high alloy stainless steels are useful for corrosion resistance for storage or processing liquids, water and waste holding tanks, etc. Titanium alloys are also useful in this field. Specialty tool steels are useful for cutting tools for light metals and for abrasion resistant surfaces.

Light metals (Al and Mg) will be of primary interest for AC conductors and flight hardware. For Al alloys containing 7 to 10% imported elements (such as 7075, 7178, 7090 PM, Weldalite 049) one can realize up to about a 75% increase in specific tensile or yield strength from the best alloys made solely from lunar elements (such as 5056). For Mg alloys, as little as 0.5% Zn in AZ80A can result in a 25 to 30% elevation in specific tensile or yield strengths.

The mechanical properties of light metal alloys deteriorate rapidly with increasing temperature due to low melting points (640–660°C). Considerable improvements in elevated temperature properties can be obtained using refractory fiber reinforced composite materials (Lynch 1975).

TABLE XIII

Mechanical Properties of Pure Metals

Metal	Tensile Strength MPa	Yield Strength MPa	Modulus of Elasticity GPa	Elongation (%)	BHN Hardness*
Aluminum (99.9996%)	40–140	15–120	62	55–49	17–27
Iron (99.9+)	245–280	70–140	208.2	30–60	82–100
Magnesium (99.98%)	160–220	90–140	40	2–15	35–47
Titanium (99.9%)	235	140	105	54	70–74
Nickel	317	59	207	30	60–65
Cobalt	760–860	310–345	211	15–20	163
Chromium	413	362	248	44	125
Calcium	48	13.7	19.6	19.6	16–18

* 500 kg load, 10 mm ball.

Electrical Conductivity of Pure Metals

	Resistivity	Conductivity		Density	Mass Conductivity	
Element	nΩm	10^{-8}(Mho/m)	(%IACS)	10^{-3}(Kg/m^3)	10^{-3}(mho m^2/kg)	(%IACS)
Cu	16.730	0.5977	103.0	8.96	6.671	103.0
Ag	15.9	0.6289	108.4	10.492	5.994	92.6
Al	26.548	0.3767	64.9	2.699	13.96	215.6
Mg	44.5	0.2247	38.7	1.738	12.93	199.7
Na	49.0	0.2041	35.2	0.971	21.02	324.7
Fe	97.1	0.1030	17.8	7.874	1.308	20.2
Cu (IACS)	17.241	0.5800	100.0	8.96	6.473	100.0
Ca	34.76	0.2877	49.6	1.55	18.56	286.7
Ti	420	0.0238	4.1	4.507	0.528	8.16

TABLE XIII (cont.)
Properties of Magnetic Materials

	Induction (Tesla)		Field Oersted	Field (AMP/M)	Core Loss (W/lb) 60 Hz 0.014″	Permeability		Curie Temp.	Resistivity
Type	B_{max}	B_{rem}	Hc	@1T	@1.15T	Initial	Max.	(°C)	(Ω cm)
M15	1.95	0.5	0.25	(19.9)	1.46	1, 300	2–11,000	730	$45–69\times10^{-6}$
M8	2.01	0.66–0.86	0.05–.1	(4–8)	0.8	2, 500	55–60,000	740	$45–52\times10^{-6}$
M5					0.5				
Sendust	1.0		.035			30, 000	120,000	500	81×10^{-6}
$(Mg,Mn)Fe_2O_4$	1.0								
S1, S3, R1	0.16–0.2		0.6–2.5	(47.7–199)		40–50 (1 MHz)	350–1800	265	
4373A			0.3	(23.9)		1, 000		170	
Fe_3O_4	0.6					70		585	10^{-2}
Fe_2O_3	0.52							675	
$MnO\text{-}xAl_2O_3$ $(1-x)Fe_2O_3$						Useful @ 10^9 Hz			
SA,B1–2	0.1–0.145							140–180	$10^5–10^7$

Hard Materials

Type	Induction (Tesla)			Field		Gauss-Oersted		Curie Temp.	Resistivity
	B_{max}	B_{rem}	B_d	H_c Oersted	H_d (Amp/m)	$(BH)_{max}$ x	(Tesla-Amp/m)	(°C)	(Ω cm)
$MO{\cdot}6Fe_2O_3$ oriented	0.48	0.385	0.1922	2,200 (175,000)	1,820 (144,800)	3.5×10^6	(27,850)	450	10^4
M=Ba,Sr (Ca-Ba-Sr)									
$MO{\cdot}2FeO{\cdot}8Fe_2O_3$	0.48							450	
Alnico 5 DG 8 Al, 14Ni, 24Co, 3 Cu, bal. Fe		1.29	1.05	650 (51,700)	580 (46,200)	6.1×10^6	(48,500)	890	$47{\cdot}10^{-6}$

B. Products Derivable from Nonmetallic Materials

This class of materials can be divided into crystalline, glassy, and mixed (glass-ceramic) materials. All of these categories can be obtained from either raw, semi-refined or refined compositions derivable from lunar resources.

Crystalline Materials. The simplest representatives of this class are cast basalts or highland soils or rocks cooled sufficiently slowly to provide nearly complete devitrification. Quarried basalts might also be used, but these may be subject to random internal flaws due to meteoritic or other environmental events. The mechanical properties of cast terrestrial basalts are given in Table XIV. Their values are substantially superior to typical unreinforced concretes.

Refined crystalline nonmetals such as Al_2O_3, MgO, TiO_2, can be used for refractories, abrasives, insulation, dielectrics, etc., (see Table IX).

Electrically conductive crystalline materials such as Fe_3O_4, TiO, or Ti_5Si_3 can be used for high temperature electrodes and Fe_3O_4 or other spinels, and magnetoplumbites of general composition $Mo{\cdot}6Fe_2O_3$ can be used for soft and hard magnetic materials according to their specific magnetic properties.

A considerable literature has appeared advocating use of hydraulic concretes on the Moon. Use of such materials requires production of various Portland or high alumina cements which requires fractionation or partial refinement of lunar soils to extract CaO. Hydrogen must also be imported to form water and curing in vacuum must be done with considerable precautions to avoid significant water loss through evaporation. These substantial liabilities would generally outweigh general benefits in comparison with anhydrous metal or basalt structures. Portland cements might prove useful, however, as an ambient temperature bonding agent for joining basalt panels or beams in large assemblies.

Glassy Materials. Raw basalts as well as SiO_2 and silicates can be drawn and quenched to form glass fibers. Addition of several metal oxides (particularly alkali or boron oxides) can substantially lower the melting and working temperatures of typical silicate glasses whose original compositions contained little or none of these constituents. Such modifications can facilitate fabrication of glass matrix composites containing more refractory fibers. Alkalic glasses possess limited water solubilities which restricts certain exterior terrestrial applications, but could be beneficial for lunar manufacturing steps.

Bulk glass products from light or dark melts can be obtained by adjusting the iron and residual Mn, Cr, and Ti contents of the melts. Water white glasses may require chemically refined binary oxides as starting materials. Proper dopants have been developed to provide substantial resistance to radiation darkening as for cover glass for solar cells in space. Light, dark, or foamed glasses can be used as structural members, and glasses of any color may be used for reflecting mirror substrates. Powders, flakes, sheets or films of clear glass can provide spectrally selective cool surfaces with low solar spectral absorptivity and high thermal emissivity (Waldron 1980*d*). (This is also true for most white crystalline oxides.)

TABLE XIV
Compositions and Properties of Nonmetals

Code No.	Application	Composition of Commercial Glasses — Composition (wt %)								Properties				
										Working/ (melting) temp. °C	Density g/cm^3	Compressive strength MPa	Tensile strength (fiber or whisker) GPa	Elastic Moduli GPa
		SiO_2	Al_2O_3	CaO	MgO	Na_2O	K_2O	B_2O_3	Other					
1720	electrical	62.0	17.0	8.0	7.0	1.0		5.0		1190	2.52			87.6
1723	electrical	57.0	15.0	10.0	7.0			5.0		1175	2.64			86.2
6720	general/ lighting	60.0	10.0	5.0		9.0	2.0	1.0		1023	2.58			70.3
6810	general/ lighting	56.0	10.0	4.0		7.0	1.0	1.0	3.0 PbO	1010	2.65			
7900	high temperature	96.0	0.3					3.0		>1500	2.18			69
7913	high temperature	96.5	0.5					3.0		1530	2.18			72.4
7940	high temperature	99.9								1580	2.20		5.86	73
9606	radome	56.0	20.0		15.0				9.0 TiO_2	1350	2.61	138^+		119.3
G-20	lab ware	75.7	5.1	1.3		6.2	1.2	6.9	3.6 BaO	1190	2.89			72
E-fiber	electric	54.0	14.0	17.5	4.5			10.0		—	2.54		3.45	82.8
C-fiber	chemical	65.0	4.0	14.0	3.0	8.0	0.5	5.5			2.49		3.10	69
S-fiber	high strength	65.0	25.0		10.0					—	2.5		4.48	89.7

TABLE XIV (cont.)
Compositions and Properties of Nonmetals

Crystalline Nonmetals	(melting) temp. °C	Density g/cm^3	Compressive strength MPa	Tensile strength (fiber/whisker) GPa	Elastic Moduli GPa
Al_2O_3	(2070)	3.97	2400	6.90(W)	483
MgO	(2852)	3.5	1380		345
TiO_2	(1857)	4.23	138$^+$		340
Ti_5Si_3	(2120)	4.32			125–150 est.

	Terrestrial Basalt	Lunar Basalt
Sample Number	RC-11*	12041**
Location	Rocky Canyon, ID	Lunar Mare
Analysis (wt%)	100%	100%
SiO_2	49.41	48.10
Al_2O_3	17.92	17.60
TiO_2	2.41	1.80
Fe_2O_3	2.00	—
FeO	9.66	10.50
MnO	0.17	0.10
CaO	9.06	11.10
MgO	5.64	9.30
K_2O	0.81	0.50
Na_2O	2.57	0.70
P_2O_5	0.35	—
B_2O_3	—	—
Cr_2O_3	—	0.30

Physical Properties of Cast Basalt vs Concrete

MPa	Cast Basalt[a]	Concrete[b]
Compressive Strength	162.5 to 203.5	7.25 to 35
Tensile Strength	10 to 14.25	.8 to 5
Shear Strength	16.25 to 18.3	1.5 to 4

$^+$ Shear strength
* Freitas and Gilbreath (1982)
** Simonds (1988); see Capps and Wise (1990)
[a]Dalton and Hohmann (1972)
[b]McCormac (1986)

Amorphous thin film photovoltaic solar cells using silicon or other active layers will probably not be an early application for ISMU, but could eventually assume monumental importance if large-scale power beaming from the Moon occurs (Criswell and Waldron 1990).

Glass-Ceramics. Subjecting quenched glass melts to dual stage heat treatments can first nucleate submicron-sized crystallites and then permit ultrafine grained partial devitrification that may be arrested at the desired level of conversion. This produces dispersion strengthened products with substantial property improvements compared to the original glasses. Experiments on terrestrial basalt melts have demonstrated formation of useful glass-ceramic structures (Beall and Rittler 1976).

VIII. SUMMARY

Nonvolatile products can be derived primarily or solely from lunar resources using processing methods of varying degrees of complexity and specific energy requirements. Both metallic and nonmetallic feedstocks can be derived *without* employing chemical refining operations where the metallic systems are limited to free metal fractions of lunar soils (predominately Fe with minor fractions of Ni and Co). Chemical refining is required for propellant production, specifically for oxygen, which will co-produce metals such as Fe, Al, Mg, Ti, Ca and Si (semi-metal) or mixtures depending on the refining systems selected. Woodcock (1986) has determined that the economic value of nonvolatile products will probably exceed that of propellants for lunar industries. The selection and optimization of processing methods and scope of lunar industrial activities will require more detailed analyses of base scenarios and better defined engineering parameters for both refining and manufacturing operations.

Acknowledgments. The author wishes to acknowledge useful suggestions received from W. Agosto and A. Cutler which greatly improved this chapter.

REFERENCES

Agosto, W. N. 1983. Solar furnace extraction of volatiles, metals and ceramics from nonterrestrial materials. In *Space Manufacturing 1983: Proc. of the Sixth Princeton/SSI Conf.*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 275–276.

Beall, G. H., and Rittler, H. R. 1976. Basalt glass-ceramics. *Amer. Ceram. Soc. Bull.* 55:579.

Capps, S., and Wise, T. 1990. Lunar basalt construction materials. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), p. 123.

Colson, R. A., and Haskin, L. A. 1991. Magma electrolysis—an update. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 7.

Criswell, D. R., and Waldron, R. D. 1990. Lunar system to supply solar electric power to Earth. In *Proc. of the 25th Intersoc. Energy Conversion Engineering Conf.: IECEC-90*, vol. 1, eds. P. A. Nelson, W. W. Schertz and R. H. Till, p. 61.

Cutler, A. H., and Krag, P. 1985. A carbothermal scheme for lunar oxygen production. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 559–569.

Dalton, C., and Hohmann, E. 1972. *Conceptual Design of a Lunar Colony*, NASA CR-129164.

Database Compilation. Lunar Sample Curator (Houston: NASA Johnson Space Center); quoted by Phinney et al. 1977

Downs, W. R. 1971. *Oxygen and Water from Lunar-Surface Material*, NASA TM X-58061.

Eagle Engineering. 1988. Conceptual Design of a Lunar Oxygen Pilot Plant. EI 88–182, NASA Contract No. NAS 9-17878.

Freitas, R., and Gilbreath, W., eds. 1982. *Advanced Automation for Space Missions*, NASA CP-2255.

Grimley, R. T., and Lipschutz, M. E. 1984. Processing of extraterrestrial materials by high temperature vacuum vaporization. In *Space Manufacturing 1983: Proc. of the Sixth Princeton/SSI Conf.*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 341–349.

Jarrett, N., Das, S. K., and Haupin, W. E. 1980. Extraction of oxygen from lunar ores. In *Extraterrestrial Materials Processing and Construction II*. LPI Final Report (NASA NSR 09-51-001, Mod. No. 24), Ch. 8).

Lewis, J. S., and Nozette, S. 1983. Extraction and purification of iron group and precious metals from asteroidal feedstocks. *Space Manufacturing 1983: Proc. of the Sixth Princeton/SSI Conf.*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 351–353.

Lynch, C. T., ed. 1975. *Handbook of Materials Science* (Cleveland: CRC).

Lynch, D. C. 1989. Processing of metal by-products from lunar oxygen facilities. Space Mining and Manufacturing: Proc. of Annual Invitational Symp. of the UA/NASA SERC, Oct. 24–26, Tucson, Ariz., pp. III-24–III-67.

Lynch, D. C., Bullard, D., and Ortega, R. 1990. Cold plasma processing of local planetary ores for oxygen and metallurgically important metals. In *UA/NASA SERC Annual Progress Report 1989–90*, pp. I-4–I-19.

McCormac, J. C. 1986. *Design of Reinforced Concrete* (New York: Harper & Row).

Morris, R. V. 1980. Origins and size distribution of metallic iron particles in the lunar regolith. *Proc. Lunar Planet. Sci. Conf.* 11:1697–1712.

Oder, R. R. 1991. Magnetic separation of lunar soils. *IEEE Trans. Magnetics* 27:5367.

O'Donnell, P. M. 1972. *Reactivity of Simulated Lunar Material with Fluorine*, NASA TM X-2533.

Oppenheim, M. J. 1970. On the electrolysis of basalt, II: Experiments in an inert atmosphere. *Mineralogical Mag.* 37:568–577.

Peters, F. A., and Johnson, P. W. 1974. Revised and Updated Cost Estimates for Producing Alumina from Domestic Raw Materials. U. S. Dept. Interior, Bureau of Mines IC 8648.

Phinney, W. C., Criswell, D. R., Drexler, E., and Garmirian, J. 1977. Lunar resources and their utilization. *Prog. Astron. Aeron.* 57, Ch. III.

Rao, D. B., Choudary, U. V., Erstfeld, T. E., Williams, R. J., and Chang, Y. A. 1979. *Extraction Processes for the Production of Aluminum, Titanium, Iron, Magnesium and Oxygen from Nonterrestrial Sources*, NASA SP-428.

Roberts, G. J., and Roberts, J. P. 1966. An oxygen tracer investigation of the diffusion of "water" in silica glass. by O tracer. *Physics Chem. of Glasses* 7(3):82–88.

Rosenberg, S. D., Guter, G. A., and Miller, F. E. 1965. The utilization of lunar resources for propellant manufacture. *Adv. Astron. Sci.* 20:665.

Simonds, C. 1988. Hot pressing of lunar soils and qualification for manned applications. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 90–101.

Steurer, W. H. 1985. Lunar oxygen production by vapor phase pyrolysis. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSI Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Mayniak (New York: AIAA), pp. 123–131.

Waldron, R. D. 1978. In *Extraterrestrial Materials Processing and Construction I.* LPI Final Rept. (NASA NSR 09-51-001, Mod. No. 24), Ch. 2).

Waldron, R. D. 1980*a*. In *Extraterrestrial Materials Processing and Construction II.* LPI Final Rept. (NASA NSR 09-51-001, Mod. No. 24), Ch. 3).

Waldron, R. D. 1980*b*. In *Extraterrestrial Materials Processing and Construction II.* LPI Final Rept. (NASA NSR 09-51-001, Mod. No. 24), Ch. 4).

Waldron, R. D. 1980*c*. In *Extraterrestrial Materials Processing and Construction II.* LPI Final Rept. (NASA NSR 09-51-001, Mod. No. 24), Ch. 5).

Waldron, R. D. 1980*d*. In *Extraterrestrial Materials Processing and Construction II.* LPI Final Rept. (NASA NSR 09-51-001, Mod. No. 24), Ch. 7).

Waldron, R. D. 1983. Non-electrolytic route to oxygen and metallic elements from lunar soil. In *Space Manufacturing 1983: Proc. of the Sixth Princeton/SSI Conf.*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 297–313.

Waldron, R. D. 1984–1989. Unpublished studies.

Waldron, R. D. 1985. Total separation and refinement of lunar soils by the HF acid leach process. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSI Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Mayniak (New York: AIAA), pp. 132–149.

Waldron, R. D. 1988*a*. Efficient transportation systems for periodic Earth-Moon rendezvous to facilitate repetitive mass transport. In *Papers Presented to the Symp. on Lunar Bases and Space Activities in the 21st Century*, April 5–7, Houston, Tex., LPI Contrib. 652, p. 247 (abstract).

Waldron, R. D. 1988*b*. Lunar manufacturing: A survey of products and processes. *Acta Astronautica* 17:691.

Waldron, R. D. 1989. Magma partial oxidation: A new method of oxygen recovery from lunar soil. In *Space Manufacturing 7, Proc. of the Ninth Princeton/AIAA/SSI Conf.: Space Resources to Improve Life on Earth*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 69–77.

Williams, R. J., and Jadwick, J. J. 1980. *Handbook of Lunar Materials*, NASA RP-1057.

Woodcock, G. R. 1986. Economic Potentials for Extraterrestrial Resources Utilization. IAA 86-451.

DEVELOPMENT AND MECHANICAL PROPERTIES OF STRUCTURAL MATERIALS FROM LUNAR SIMULANTS

CHANDRA S. DESAI, HAMID SAADATMANESH
and KIRSTEN GIRDNER
University of Arizona

Development of structural materials from locally derived space materials and study of their mechanical properties for construction and other uses are important toward establishment of outputs and habitats on the moon and other planets. This chapter provides a brief review of various methods such as sintering and casting for the manufacture of construction materials from lunar simulants. Then two alternative and new methods developed are described. They involve (1) thermal liquefaction in which the mixture of a lunar simulant and various metallic fibers is liquified at a temperature of about 1100°C so as to form a matrix of everlapping zones of melted simulants and fibers, and (2) compaction of dry simulants under different vacuum levels using a newly developed vacuum triaxial device. The stress-strain-strength properties of the resulting materials are determined by using beam bending, uniaxial compression and triaxial compression tests. It is found that the thermal liquefaction leads to materials with significantly enhanced flexural and ductility properties compared to the sintered materials that are found to be brittle. The future research will involve such factors as use of glass fibers, techniques for optimization of mechanical properties and additional testing for tensile, fracture, and multiaxial properties of the resulting materials. It is believed that such materials may be produced on the moon by using solar energy without the use of water, and can lead to applications not only for construction, but also for machines and other structures for use in the colonization of space.

It is now expected that the human race will establish, in the near future, outposts and habitation on the Moon, Mars and other planets. Construction of facilities like habitats, factories and roadways would require structural materials, most of which will need to be manufactured locally. The expertise and methodologies that engineers and scientists develop now for manufacturing in space can determine the successful colonization of space and may also determine the direction that space colonization would take.

Although the research and methodologies proposed and discussed in this chapter can be applicable on various planets, at this time most attention is directed toward lunar development which will require structural materials made from lunar soil or regolith. For engineering applications, it is not sufficient to merely develop materials. Testing and determination of constitutive models based on stress-strain-strength behavior of materials under the expected loads and environmental conditions are of vital importance. The determination of

properties such as deformation moduli, strength parameters, elastic, plastic and creep response, fracture toughness, ductility, flexural response, dynamic and impact response is a prerequisite for practical use. With this knowledge, the applicability of materials for construction, machinery and space-flight structures can be established.

It is possible to manufacture structural materials on the Moon, Mars and other planets from locally available materials mixed with stocks that can be obtained locally or can be carried (initially) in small quantities from the Earth. In the case of the Moon, the lunar soil or regolith can be utilized with admixtures such as metallic and nonmetallic fibers and powders, sulfur (small quantities), or other materials from the regolith. It is important that the technology developed eliminates the use of water, as it will be expensive to carry water from the Earth in large quantities.

Among the possible methods to develop structural and construction materials are compaction, and heat treatment of regolith, with or without various admixtures.

The scope of this chapter includes a brief review of previous work; development of an intermediate ceramic composite (ICC) by thermal liquefaction at about 1100°C temperature of a lunar simulant with various admixtures, and its bending and compression characteristics; development of compacted lunar simulant, and its triaxial characteristics; comments on concrete-type material at lower temperatures with mixtures of the simulant, fibers, sulfur and powders; and finally comments on future research.Brief reviews of some of the previous research results are presented below; additional reviews are available in Desai et al. (1992*a*, *b*) and Desai and Girdner (1992).

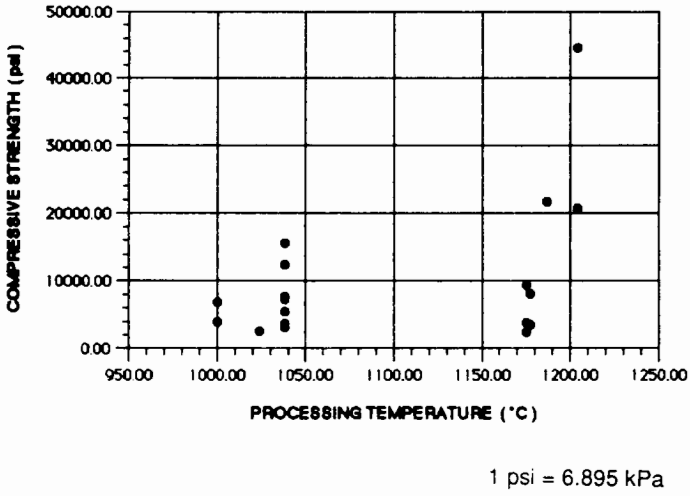


Figure 1. Compressive strengths of sintered simulants (figure from Meek et al. 1988).

I. PREVIOUS WORK

Previously proposed methods for developing construction materials on the Moon from the lunar regolith include sintering of the soil into a solid matrix. Sintering is a process in which a granular material is heated and held at temperatures below its melting point. Viscous flow on the grain surfaces form necks between particles, bonding them together. In general, the resulting material is porous and brittle. The porosity and brittleness can be reduced by using higher temperatures, longer sintering times, smaller material grain sizes and the application of external pressure during sintering.

Early experimentation of sintering as applied to lunar soils is described by Simonds (1973), in an attempt to explain the formation of lunar breccias under low stress conditions and heat produced by meteorite impacts. The range of temperatures at which a cohesive solid was produced under no applied stress was 800°C to 1000°C. Under hot pressing, a densely welded material can be produced at 600°C, when held at this temperature for 7 days. These experiments were conducted on a powdered glass with composition approximating that of the returned Apollo 14 breccias. In a subsequent paper (Simonds 1988), the hot pressing of lunar regolith for construction materials and the resulting useful by-products is further investigated. He proposed that a small solar powered test plant producing sintered products can also provide 31 kg of hydrogen (H_2) and 187 kg of oxygen (O_2) per month. It was conceded that structural design from such sintered materials, which are brittle, is limited by their available tensile and flexural strengths and ductility.

Research by Meek et al. (1988) focused on the production and testing of simulants sintered by using microwave radiation. Samples of three different lunar simulants, approximating the compositions of returned soils from Apollos 11, 15 and 16, were processed and tested. The cylindrical pellets 0.9 cm long, were cold pressed to 50,000 psi (350 MPa), then sintered at 2.45 GHz of microwave energy frequency. Sintering times ranged from 5 min. to reach a temperature of 1000°C, to 45 min. to reach 1204°C. The mechanical and physical properties determined by testing the samples included compressive strength, Young's modulus, percent strain to failure, thermal shock, hardness and density. Test results for compressive strength from this research are shown in Fig. 1. The variation of mechanical properties, although showing a general trend of increasing strength and density with higher sintering temperatures, shows a large variation of strength within the temperature range considered.

Uhlmann et al. (1975) analyzed the process of elastic breccia formation by viscous sintering in a stress-free environment. The crystallization and sintering were treated as concurrent competing processes. The kinetic analysis of sintering was based on a modification of the Frenkel treatment of viscous sintering appropriate for conditions of continuous cooling. The approach was applied to material composition similar to that of lunar soil. It was shown that the kinetic analysis can be used to estimate both the rate at which a given breccia cooled and the minimum temperature at which the matrix particles

came in contact. The results obtained seemed physically reasonable and suggest that the sample used cooled on the surface of the Moon in its present form rather than buried in an ejecta blanket from which it was later excavated.

It is generally acknowledged that with greater energy expenditure, the lunar regolith or basaltic rock can be melted and cast in order to form a glass-ceramic with greater mechanical properties than those of sintered materials. The process of basalt casting is described by Kopecky and Voldan (1965) where, in their native Czechoslovakia, manufacture of basaltic pipes, tiles and other industrial products began in 1950. A maximum temperature of 1350°C was required to fully melt the crushed basaltic rock, which was poured at 1200°C. After casting into sand or metallic molds, the pieces solidified at 900°C to 1000°C and were cooled from 800°C to nearly room temperature over a period of 24 hr. The long cool down period was necessary to promote crystallization and release of internal stresses, which may often cause bursting, if no preventive action is taken. The following range of values for mechanical properties of cast basalt were reported by the authors:

Compressive strength	57,000 to 71,500 psi (399 to 501 MPa)
Tensile strength	3550 to 5000 psi (25 to 35 MPa)
Bending strength	5500 to 6500 psi (39 to 46 MPa)
Thermal expansion	$78 \times 10^{-7}/°C$.

In terms of thermal resistance, the less complex pieces such as tiles or pipes fared well, as an instantaneous change in temperature of up to 100°C caused no damage. Other advantages of these products include high resistance to abrasion, moisture, weathering and chemical agents.

In order to increase the properties such as flexural and tensile strengths and ductility of materials made from the regolith, it may be advantageous to add fiber reinforcements. These fibers can be made of metal or glass, and can be produced either from resources on the Moon or imported from Earth, the latter in the initial stages of colonization. The production and use of fibers drawn from basalt is discussed by Subramanian et al. (1977). In this study the fibers were tested as reinforcements in polymer composites. Fibers can be drawn from basalt within the temperature range of 1175°C to 1375°C. The tensile strength of the fibers drawn between 1250°C and 1325°C ranged from 330,000 psi (2.3 GPa) to 432,000 psi (2.9 GPa). The authors tested several basaltic rocks of slightly varying chemical composition, so the above range is not strictly a function of processing temperature. The authors determine that under optimum processing and testing conditions, the strength of the basaltic fibers can approach that of commercially available E-glass.

Ho and Sobon (1979) have presented a conceptual design of a system to produce a fiberglass made of lunar materials. For every kilogram of lunar soil,

590 g of fiberglass can be produced in this design. In the proposed processing system, the mass initially transported from Earth would be approximately 110×10^3 kg (120 ton). In one year, it is estimated the plant could process raw materials of 90 times its own mass, producing 5.84×10^6 kg of fiberglass. The plant uses solar energy as the source of power to melt and process the soil.

Goldsworthy (1985) presented a comprehensive study of the production, uses and economics of fiber composites for structural materials to be produced on the Moon. He maintained that the cost of producing metals and their alloys on the Moon would not be competitive with regard to the cost of transportation from Earth. He proposed that for a structural material produced in space the most efficient would be a glass/glass composite with both matrix and fibers originating in the lunar soil.

The use of metal fibers has been mentioned in several studies; however, no test results for mechanical stress-strain-strength properties or detailed discussions have been presented. Metals have an advantage over glass fibers in that their inclusion can exhibit increased ductile behavior. One of the objectives of the research presented herein is to develop ceramic composites from lunar simulants with various metal fibers and to study their mechanical properties.

II. LOADING CONDITIONS

The gravity level on the Moon is 1/6 that of the Earth. As a result, the forces exerted by equivalent masses on the Moon will be proportionally reduced. However, the structures will be subjected to relevant mechanical loads as well as additional loads due to such environmental effects as extreme temperature changes and impact of meteorites.

The Moon's virtual lack of atmosphere results in some advantages as well as disadvantages in the design of structures. For example, the lack of lateral forces exerted by wind pressure results in smaller member sizes. However, the lack of atmosphere will result in direct exposure of the lunar surface to harmful solar radiation. Such harmful radiations can prematurely disintegrate certain types of construction materials such as fiber composites. It has been calculated that 2.4 m of lunar regolith could protect human habitats, reducing radiation to safe and acceptable long-term levels (Guerra 1988).

The lack of atmosphere requires any structure in which people are to work and live without specialized protective equipment to be pressurized to nearly one atmosphere or 14.7 psi (100 kpa). This will result normally in high tensile stresses in the structures used to shield people and equipment. Lack of atmosphere also is a major cause of extreme ranges of temperature on the lunar surface. Temperatures measured on the Moon range between 111°C and −171°C, Taylor (1982). Such a wide temperature range will cause severe restrictions on the type of structural materials that can be used for lunar habitation.

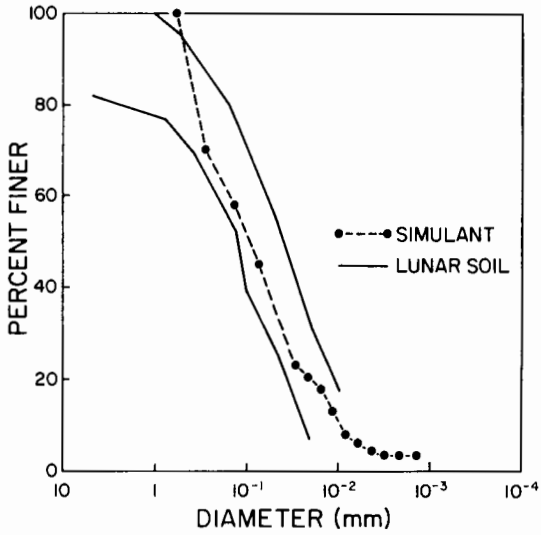


Figure 2. Grain size distribution of ALS compared to lunar soils (lunar soil data from Carrier et al. 1973).

Due to the harsh lunar environment, most of the human labor must be done in low-risk Intra Vehicular Activity (IVA), where crew members are not exposed to life-threatening risks (Toups 1990). Any construction to be done as Extra Vehicular Activity (EVA) must be minimized to prolong life expectancy of those living on the lunar surface. Materials and procedures developed for lunar construction must be sensitive to these requirements.

Overall, it is necessary to develop materials that can provide required structural properties to handle the mechanical loads as well as to provide protection against the environmental and other loads. The present research can allow development of a methodology to manufacture materials for these considerations.

III. LUNAR SIMULANTS

Because large quantities of lunar regolith are not available, it is necessary to develop and utilize lunar simulants. Several lunar simulants have been developed. Two of these are the Arizona Lunar Simulant (ALS) (Desai et al. 1992*a*) and the Minnesota Lunar Simulant (MLS) (Weiblen et al. 1990). The ALS has been developed specifically for the work in this chapter. However, future use of the MLS is planned for the purpose of comparison.

A. Arizona Lunar Simulant

The ALS is produced from a basaltic rock, the Pomona flow, from the

Columbia Plateau near Hanford, Washington; it is described by Fuenkajorn (1985). The grain size distribution curve for the simulant in comparison with that for the regolith samples returned during the Apollo missions is shown in Fig. 2 in comparison to the distributions for lunar soils reported by Carrier et al. (1973). The specific gravity of the simulant is 2.86. The chemical composition of the simulant is similar in some respects to low-Ti mare soils; Table I shows the chemical composition of the simulant in comparison with those of the MLS and mare and highlands lunar soils.

TABLE I
Chemical Composition (% by weight)

	Simulants		Lunar Soils (Taylor 1975)	
Compound	ALS (Fuenkajorn 1985)[a]	MLS (Weiblen et al. 1990)	Mare	Highland
SiO_2	48.0–50.0	43.86	45.4	45.5
Al_2O_3	13.5–16.0	13.68	14.9	24.0
TiO_2	1.6–3.2	6.32	3.9	0.6
FeO	7.0–12.5	13.4	14.1	5.9
MnO	0.2–0.25	0.198	—	—
MgO	4.3–6.5	6.68	9.2	7.5
CaO	8.3–10.3	10.13	11.8	15.9
Na_2O	2.7–3.0	2.12	0.6	0.6
K_2O	0.5–1.5	0.281	—	—
Fe_2O_3	1.9–4.6	2.6	—	—
P_2O_5	—	0.2	—	—
CO_2	—	0.0015	—	—

[a] The range is repeated here as given in the reference.

B. Comparison of Simulant and Lunar Soils

Table I presents the chemical compositions of the simulant used and those of typical lunar soils. As can be seen from this table, the simulant used herein closely approximates most of properties of the lunar soils. It was, however, not feasible to develop agglutinates, the small spherical glassy fragments produced by meteor impacts on the lunar surface. It is possible, in later research, to consider simulants mixed with such agglutinates produced on the Earth (Weiblen et al. 1990). For our objective, however, the simulant composition is sufficiently close to that of the lunar soils for investigating the feasibility of developing construction materials and study of their stress-strain-strength characteristics by using the proposed procedure of thermal liquefaction and compaction.

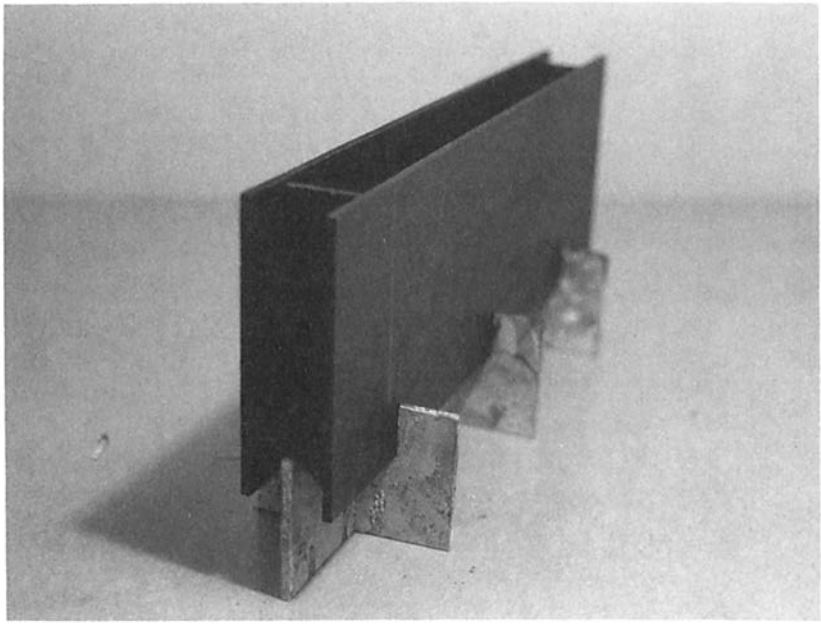


Figure 3. Mold for specimens.

IV. MATERIAL PROCESSING

Among the possible methods of preparing structural materials, two that are currently considered are (1) thermal liquefaction of a mixture of simulants with fibers and powders at high temperatures (~1100°C), and (2) thermal binding of the mixture (with fibers, sulfur, powders, etc.) at lower temperatures, ~150°C to 500°C. *Thermal liquefaction* is defined as the process of the creation of an intermediate ceramic composite ice material from the mixture at a given temperature or through cyclic application of temperatures under which the simulant may melt before fibers (powders) or not at all, or the fibers (powders) may melt before the simulant or not at all, but the resulting material is a "solid" matrix. The solid matrix is composed of melted (or unmelted) fibers with melted or unmelted overlapping influence zones of the simulant. It is felt that this material will have enhanced engineering properties such as ductility and fracture toughness compared to the sintered material (Simonds 1988). *Thermal Binding* is a process involving lower temperatures under which materials such as sulfur and iron or their combination would provide binding effects in the simulant. The resulting material is expected to be more like "mortar" and "concrete" with deformation moduli and strengths comparable to those of conventional concrete.

TABLE II
Beam Bending Test and Sample Data

Sample	Fiber—% by Weight[a]	Bending Strength psi[b]	Matrix Density g cc^{-1}	% Fiber By Volume
1A	None	1300.4	2.38	0
1B	None	1518.2	2.48	0
2	None	2620.7	2.43	0
4A	AL-7.5	2249.8	2.58	7.19
4B	AL-7.5	2872	2.61	7.27
5B	CS-15	1075.9	2.18	4.65
7A	AL-15	2553.3	2.39	13.5
7B	AL-15	2661.5	2.36	13.3
8B	SS-15	4278.8	2.2	4.54
8C	SS-15	3207.1	2.46	5.04
9	None	1272.8	2.44	0
10	AL-10	594.4	1.82	6.95
11	SS-7.5	1953.2	2.39	2.32
12A	CS-7.5	1599	2.39	2.39
12B	CS-7.5	1958.1	2.41	2.42
18A	None	1672.1	2.44	0
18B	None	1616.1	2.53	0

[a] Al = aluminum; SS = stainless steel; CS = carbon steel.
[b] 1 psi = 6.895 kPa.

This chapter gives details and initial results for the thermal liquefaction studies. Further details including a review of the subject are available elsewhere (see Desai and Girdner 1992). Research on thermal binding is ongoing and results are not yet available.

A. Details of Thermal Liquefaction

For beam bending and compression tests, samples were prepared in molds (Fig. 3) and heated to ~1100°C in a furnace with capacity up to 1700°C, kept at that temperature for about two hours and cooled. The variations are programed and the temperatures obtained are shown in Fig. 4. Samples with approximate dimensions of 25×2.5×6.25 cm (10×1×2.5 inches) were cut and ground for beam-bending tests. Samples of various sizes were then cut from the beam-bending specimens for compression testing.

B. Testing

The beam-bending tests were performed using a test device (Fig. 5) constructed to ASTM specifications (ASTM C78-84 1990). This device measured applied load, displacement, and strains at four locations (Fig. 6). The load was gradually increased to failure. The loading program included one or more unloading-reloading cycles. The test specimens, at this time, include

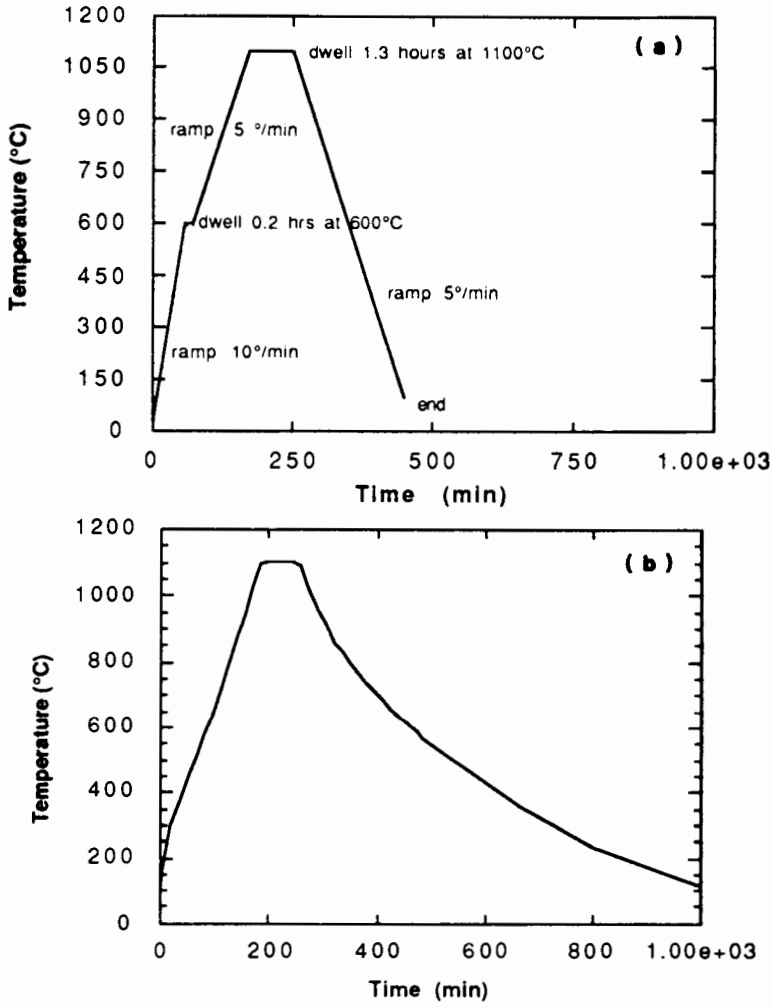


Figure 4. Temperature vs time cycle for sample preparation. (a) Programmed temperautre cycle; (b) actual temperature cycle.

those of simulant without admixtures and those of simulants with aluminum, steel and carbon-steel fibers.

Compression tests on specimens cut from the beams were also performed until failure. For these tests an MTS frame was used, and applied loads and axial deformations (strains) were measured.

C. Results

Figure 7 shows a typical load displacement curve at midspan for specimen No. 11 (Table II) with 7.5% stainless steel fibers. Figure 8 shows correspond-

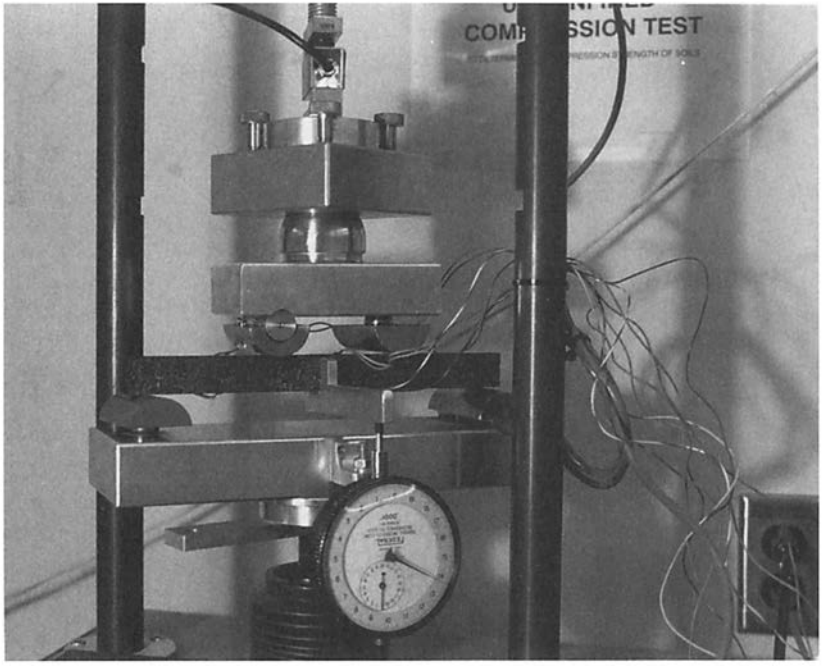


Figure 5. Beam bending test device.

sponding results for compressive and tensile stresses vs measured strains. In Fig. 7 the peak and residual loads indicate ductile and softening behavior. Because the strain gauges were damaged at the peak load, the stress-strain results (Fig. 8) do not show post-peak behavior. Table II gives details of samples tested together, along with the densities, fiber contents, and bending strength (f'_b) for each specimen. Bending strength was computed by using the following equation:

$$f'_b = \frac{p_{\max}\ell}{bh^2} \tag{1}$$

where $p_{\max}$ = maximum or failure load, ℓ = span of beam, b = width and h = height of the beam.

Figure 9 compares the bending strengths of various specimens. It can be seen that the specimens with aluminum (AL) and stainless steel (SS) fibers show significant increase in bending strength compared to specimens without admixtures. The values f'_b for carbon steel (CS) specimens are essentially similar to those for specimens without fibers.

The AL and SS fibers both enhance bending strength, due to the formation of overlapping zones, as discussed above. This effect is due to the diffusion of melted fibers into heated (or melted) simulants. The resulting matrix increases the flexural strength of the material. Three AL-fiber composites (samples 4B,

TABLE III

Moduli From Beam Bending Tests

Sample #	Fiber — %	Initial Modulus		1st Unloading Modulus		2nd Unloading Modulus	
		Compression psi in 10^6	Tension psi in 10^6	Compression psi in 10^6	Tension psi in 10^6	Compression psi in 10^6	Tension psi in 10^6
18A	None	3.50	3.08	3.48	3.21	—	—
18B	None	4.63	3.27	4.69	3.33	—	—
4A	AL-7.5	4.60	4.25	4.65	4.41	—	—
7A	AL-15	7.20	6.10	8.44	5.80	7.78	6.58
12B	CS-7.5	4.26	3.71	4.26	4.10	4.08	3.90
5A	CS-15	3.46	2.65	3.37	2.74	3.10	2.71
8B	SS-15	5.99	5.99	5.55	6.10	5.42	5.78
11	SS-7.5	4.33	3.81	4.01	4.38	—	—

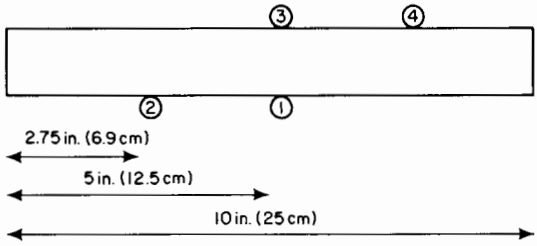


Figure 6. Strain gauge locations.

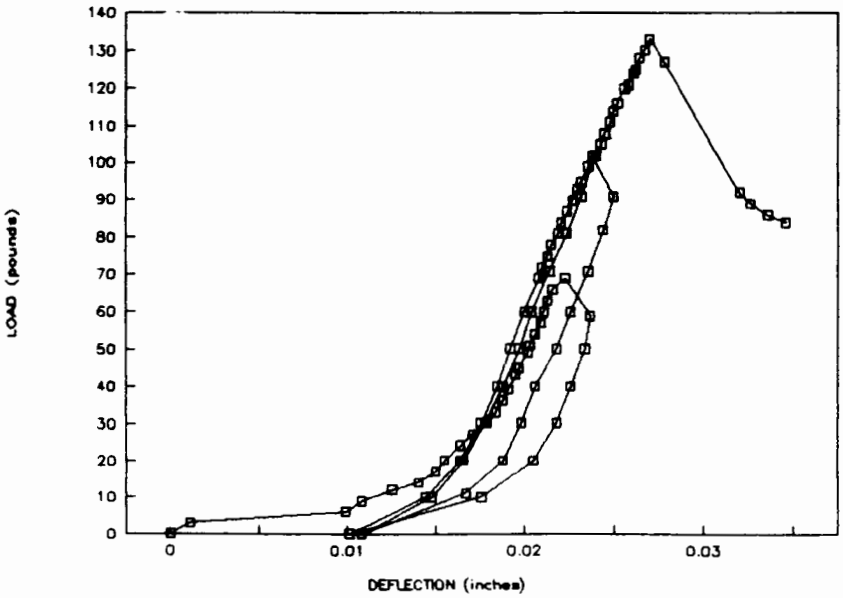
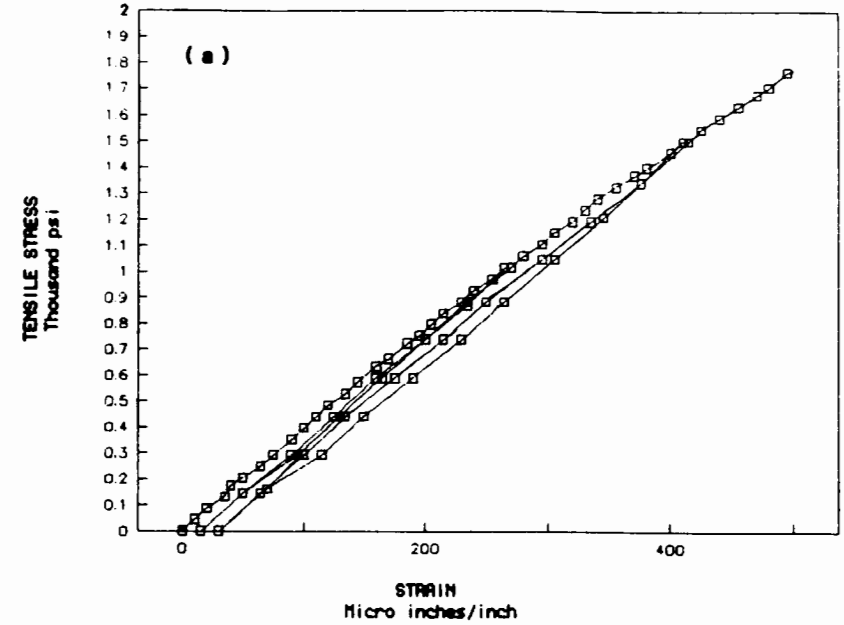


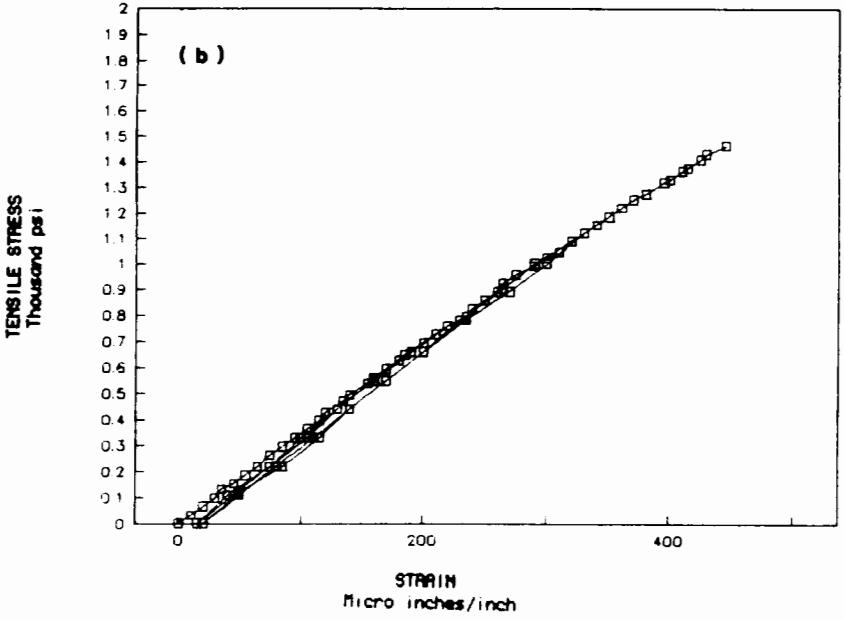
Figure 7. Bending test results for sample 11.

Table II) are shown in Fig. 10a. Although the CS-fiber composites appear to have less bending strength than the AL- and SS-fiber composites, this strength difference may be because (1) the CS fibers were longer, and (2) they produced hollows and flaws in the samples (Fig. 10b), perhaps through chemical reaction.

The elastic (Young's) modulus in compression and tension was found by two methods: (1) by measuring the average middle slope of the bending stress-strain curves, and (2) by examining the unloading slope of the curves. Table III shows the moduli for typical specimens. These results are in the same range as those for a basalt reported in the literature (Goodman 1989), and they are generally higher than those for the conventional concrete.

Table IV shows maximum strain and maximum compressive strengths


(a)
TENSILE STRESS
Thousand psi
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
STRAIN
Micro inches/inch


(b)
TENSILE STRESS
Thousand psi
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
STRAIN
Micro inches/inch

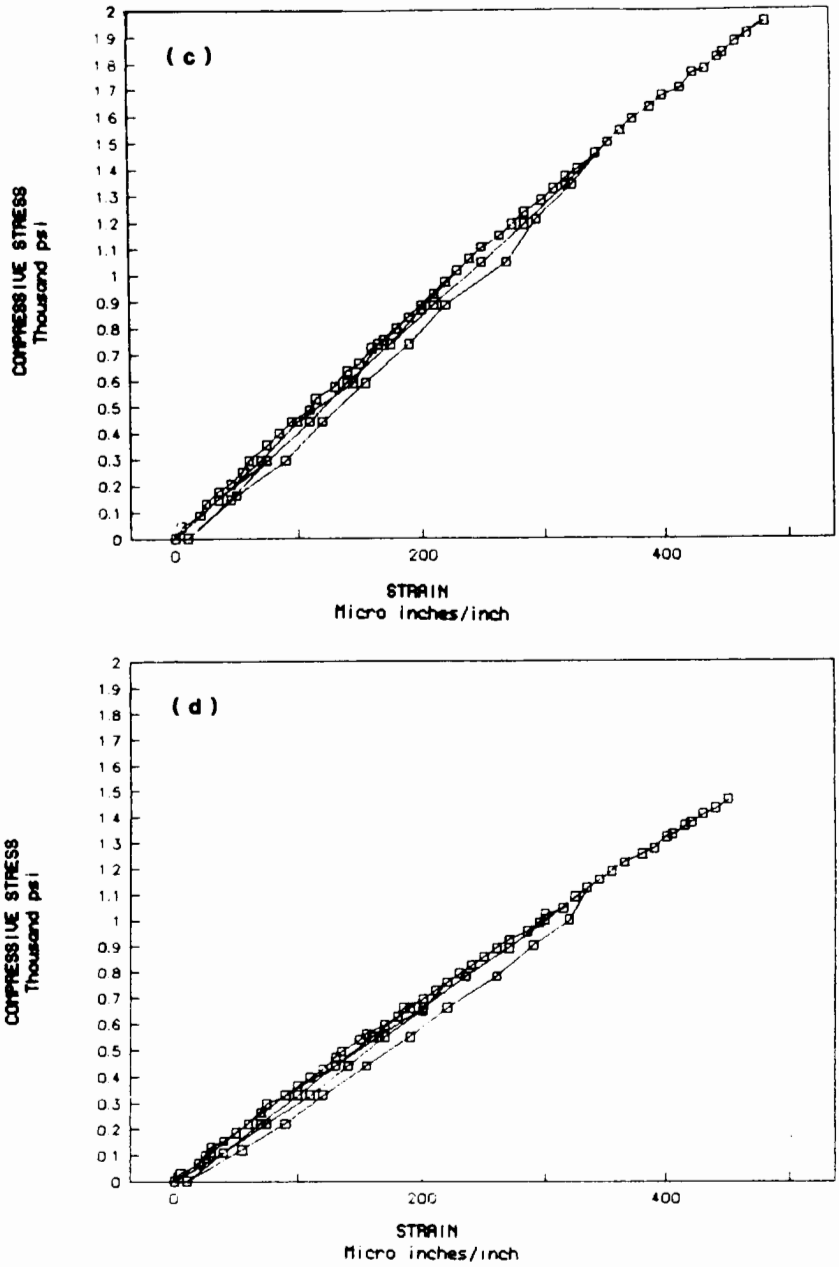


Figure 8. Bending test results for sample 11: stress-strain curve. (a) location 1; (b) location 2; (c) location 3; (d) location 4.

TABLE IV

Compression Test and Sample Data[a]

Sample	Fiber[b] %	Max Strain	Max Strain psi[c]	L/r	L in	y in	x in
2	None	0.0125	17871.55	13.64	2.22	0.56	0.99
2	None	0.0162	19371.15	9.13	1.48	0.56	0.76
18B	None	0.0181	26306.71	8.88	1.16	0.45	0.61
18B	None	0.0183	24109.39	8.67	1.25	0.50	0.74
4A	7.5 AL	0.0166	23249.67	8.75	1.14	0.45	0.68
4A	7.5 AL	0.0240	29175.22	8.77	1.15	0.45	0.68
4B	7.5 AL	0.0125	18473.10	8.23	1.12	0.47	0.50
4B	7.5 AL	0.0164	23789.90	8.48	1.22	0.50	0.65
4B	7.5 AL	0.0147	18698.70	9.40	1.36	0.50	0.71
7B	15 AL	0.0183	23343.11	7.92	1.26	0.55	0.62
7B	15 AL	0.0159	25283.97	7.66	1.22	0.55	0.67
5A	15 CS	0.0135	10336.60	7.81	1.18	0.52	0.55
5A	15 CS	0.0109	8701.18	8.70	1.33	0.53	0.66
11	7.5 SS	0.0167	20316.04	7.80	1.29	0.57	0.77
8C	15 SS	0.0129	11792.18	8.66	1.13	0.45	0.51
8C	15 SS	0.0182	25723.90	8.45	1.24	0.51	0.64
8C	15 SS	0.0137	18557.60	7.46	1.10	0.51	0.56

[a] AL = aluminum; SS = stainless steel; CS = carbon steel.

[b] L = Height of specimen; x, y = cross-sectional dimensions; r = radius of gyration about x-axis.

[c] 1 psi = 6.895 kPa.

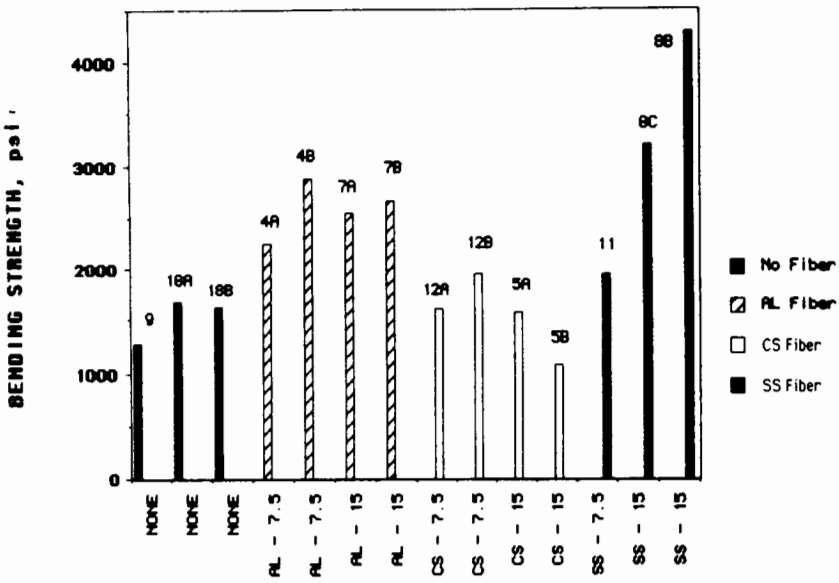


Figure 9. Beam bending strengths; 1psi = 6.895 kPa.

from compression tests for various specimens. A typical stress-strain curve is shown in Fig. 11. Typical elastic moduli from bending and compressive tests are compared in Fig. 12 for CS fibers. It can be seen that the moduli from the bending tests are much higher than those from the compression tests. This may be due to various reasons such as the very small size of compression specimens (Table IV), the existence of friction between specimen faces and loading-end platens, and possible localized stress concentrations leading to nonuniform stresses in the compression specimens.

D. Comments

The above results indicate that it is possible to manufacture a ceramic-composite type of material on the Moon by thermal liquefaction without the use of water. Solar energy may be sufficient as a heat source. The bending strengths and elastic moduli of the materials produced with fibers (except for the CS fibers) are higher than those of the materials produced without such admixtures. Also, the liquefaction process is capable of producing materials that have increased levels of ductility and hysteretic or damping properties; this is not possible with sintered materials which are significantly more brittle.

E. Future Research

Future research will involve thermal liquefaction and thermal binding with various admixtures. The possibility of mechanical (or electromagnetic) methods for optimization of fiber orientation, or thermal cycling to induce "prestressing" will be evaluated. Also considered will be the possibility of lower-

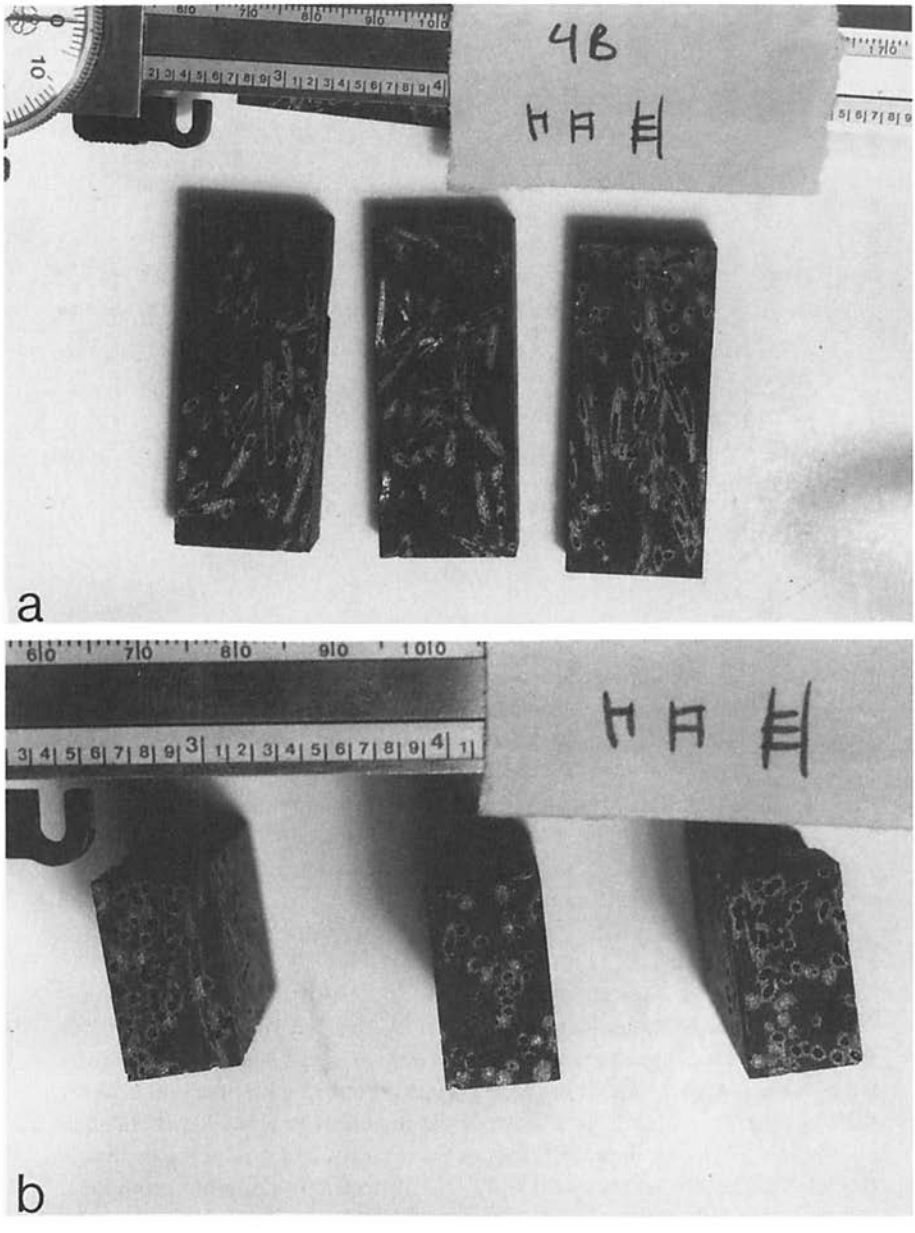


Figure 10. Material matrix with fiber. (a) Al fibers; (b) CS fibers.

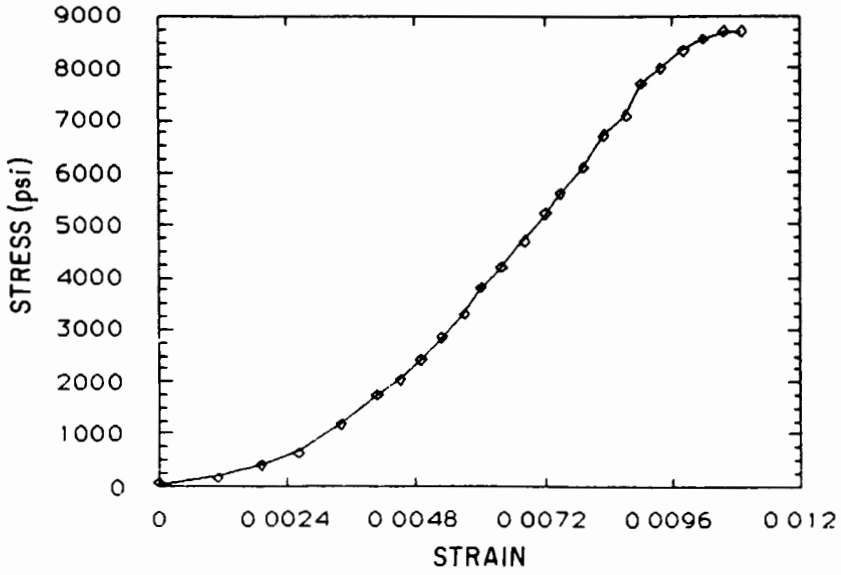


Figure 11. Stress vs strain for sample 5A-2.1 psi = 6.895 kPa.

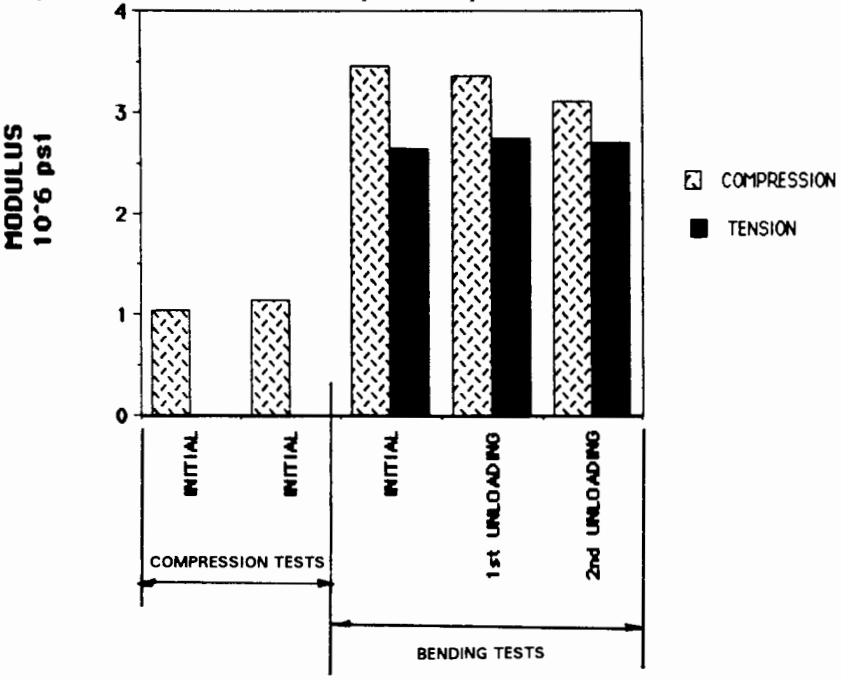


Figure 12. Comparison of moduli from compression and bending tests for beam 5A with 15% carbon steel fibers. 1 psi = 6.895 kPa

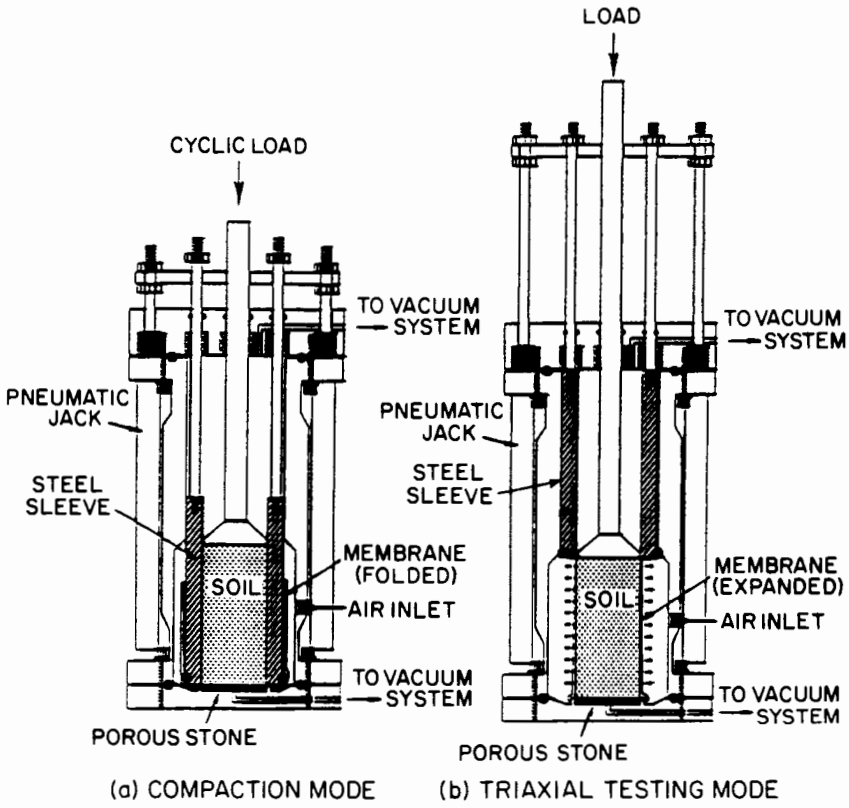


Figure 13. Vacuum triaxial device. (a) Compaction mode; (b) triaxial mode.

ing the temperature of liquefaction by addition of appropriate (alkali) chemicals. Other significant engineering properties such as fracture and damage formation and strength, toughness, ductility, tensile strength and multidimensional properties will be studied by casting and testing of appropriate plate, cylindrical and cubical specimens.

V. COMPACTION OF LUNAR SIMULANTS USING A NEW VACUUM TRIAXIAL DEVICE

One of the objectives of our work is to design and produce a new vacuum triaxial test device that can compact lunar simulants under cyclic loading, with different levels of initial vacuum. Triaxial testing is performed in this device itself without removing the compacted specimen; testing is performed by a special mechanism within the device. Preliminary constrained compression and triaxial shear tests were performed to identify effects of initial confinements and vacuums; results were used to define deformation and strength parameters. A more detailed description of the new test device and test-

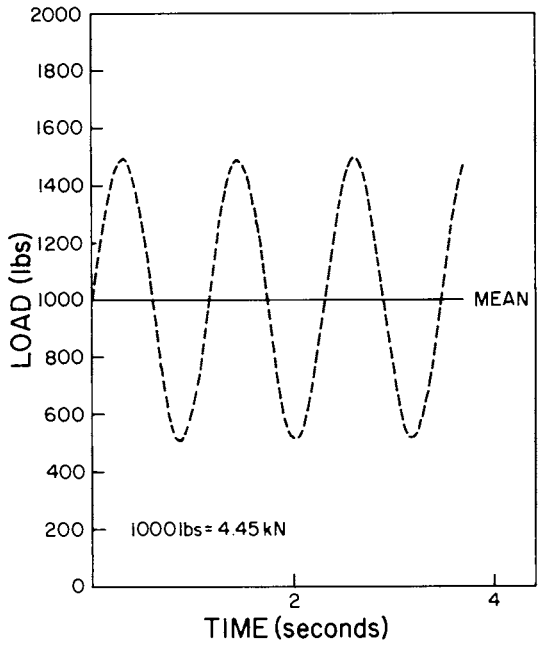


Figure 14. Cyclic load pattern: vacuum triaxial tests.

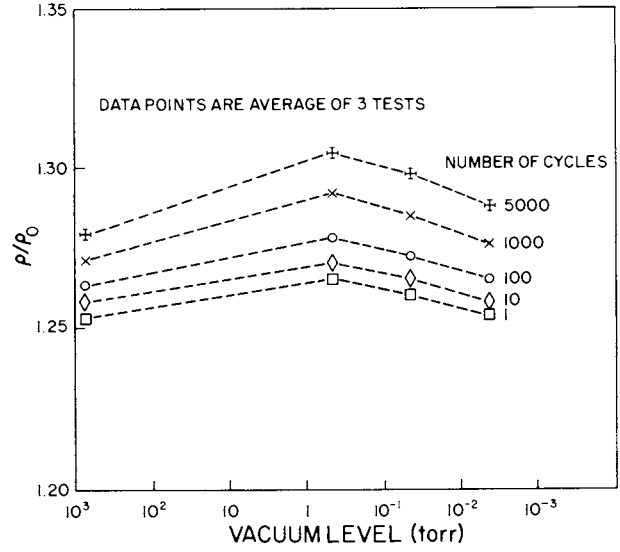


Figure 15. Density ratio vs vacuum level for selected numbers of cycles of loading.

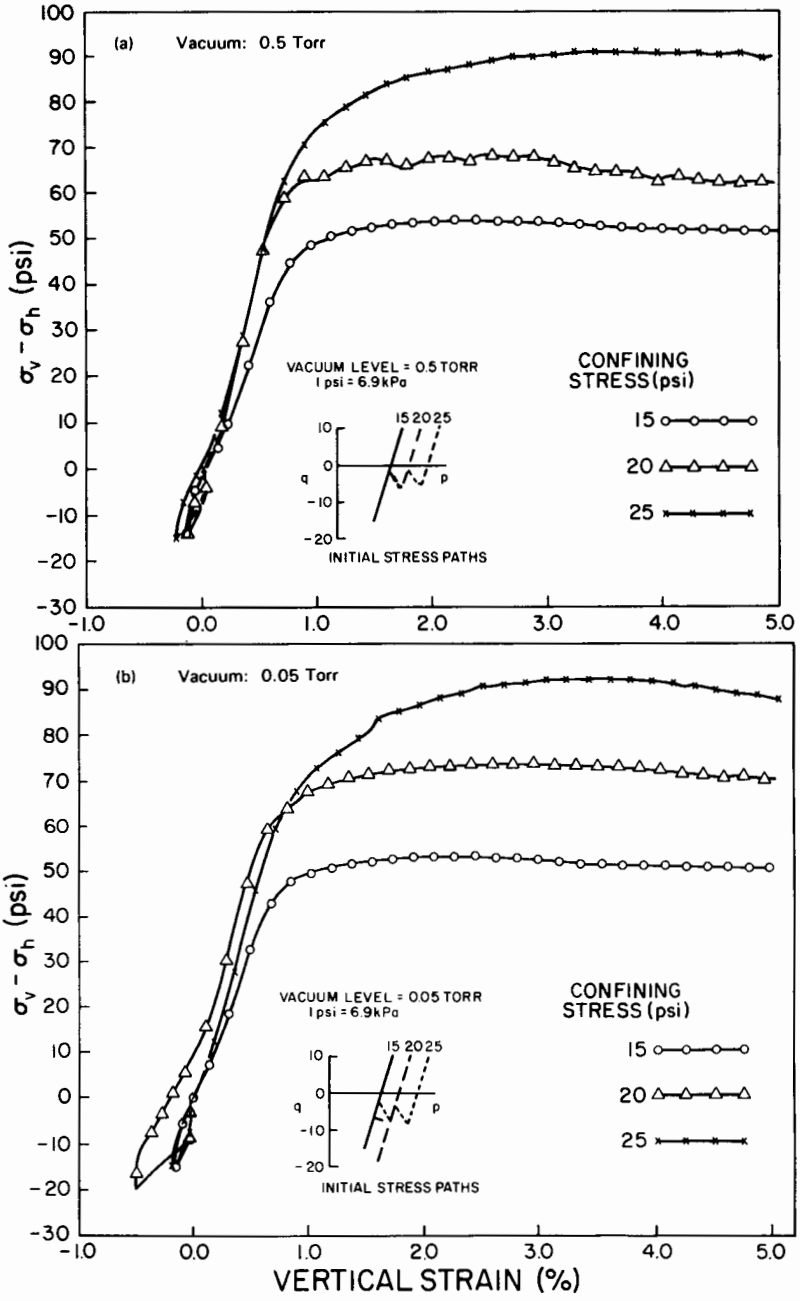
TABLE V
Simulant Densities for Selected Number of Cycles of Loading: Vacuum Triaxial Tests

Vacuum Level (torr)	Density g cm^{-3}					
	No. of Cycles					
	0	1	10	100	1000	5000
	(2)	(3)	(4)	(5)	(6)	(7)
760	1.552	1.947	1.955	1.962	1.974	1.985
	1.552	1.940	1.947	1.955	1.970	1.983
	1.552	1.940	1.947	1.957	1.967	1.978
5×10^{-1}	1.552	1.958	1.967	1.979	2.000	2.020
	1.552	1.964	1.971	1.985	2.007	2.029
	1.552	1.968	1.975	1.987	2.008	2.028
5×10^{-2}	1.552	1.957	1.963	1.974	1.992	2.010
	1.552	1.946	1.955	1.965	1.985	2.007
	1.552	1.965	1.971	1.984	2.005	2.025
5×10^{-3}	1.552	1.947	1.954	1.962	1.978	1.993
	1.552	1.939	1.947	1.958	1.976	1.993
	1.552	1.951	1.958	1.969	1.989	2.009

ing results are available in Desai et al. (1992*a*, *b*); the principal details are summarized here.

Devices used by previous investigators to study lunar simulants do not allow compaction and stress-strain-strength response to be studied at different vacuum levels. It is desirable that a test device be designed and constructed to allow maintenance of vacuum during testing.

Standard or conventional devices allow testing of cylindrical soil specimens under various confinements and loading or stress paths. For example, tests may be performed under hydrostatic compression ($\sigma_1 = \sigma_2 = \sigma_3$; where σ_1, σ_2 and σ_3 are principal stresses), under conventional triaxial compression in which the confining stress $\sigma_2 = \sigma_3$ is held constant while the axial stress, σ_1 is increased, and under triaxial extension in which σ_1 is held constant, while $\sigma_2 = \sigma_3$ is decreased. The commonly conducted tests involve applying an initial confining stress $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_0$ to simulate the *in-situ* geostatic stresses, and then increasing (or decreasing) σ_1 to induce shear stress in the specimen. This device thus allows testing for stress-strain-strength response and evaluation of deformation parameters such as Young's modulus (E) and Poisson's ratio (υ) and strength parameters such as cohesion (c) and angle of friction (ϕ).


(a) Vacuum: 0.5 Torr
VACUUM LEVEL = 0.5 TORR
1 psi = 6.9 kPa
CONFINING STRESS (psi)
15
20
25
INITIAL STRESS PATHS
(b) Vacuum: 0.05 Torr
VACUUM LEVEL = 0.05 TORR
1 psi = 6.9 kPa
CONFINING STRESS (psi)
15
20
25
INITIAL STRESS PATHS
σv - σh (psi)
VERTICAL STRAIN (%)

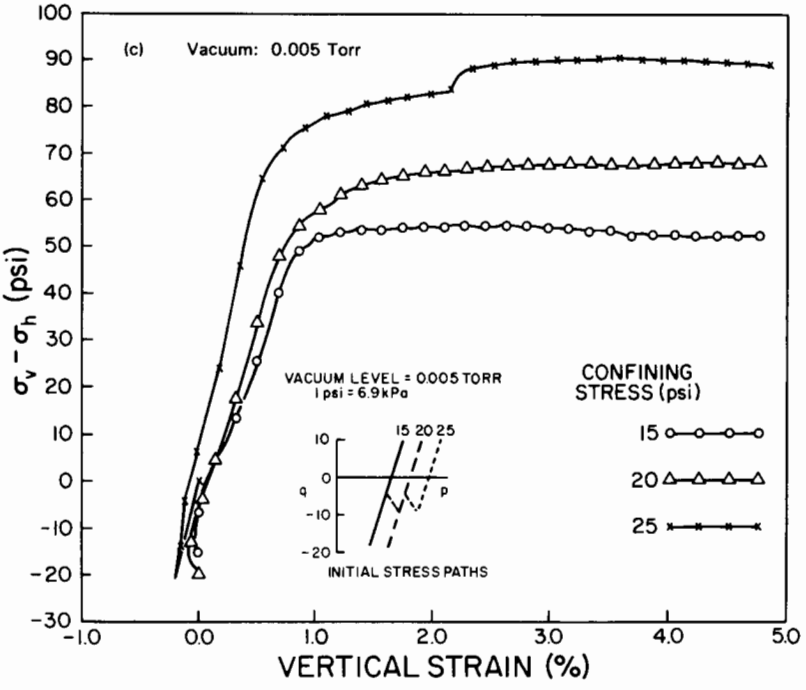


Figure 16. Stress vs strain for vacuum triaxial tests. (a) vacuum = 0.5 torr; (b) vacuum = 0.05 torr; (c) vacuum = 0.005 torr.

TABLE VI

Initial Moduli (E_i) at $\sigma_v - \sigma_h = 344.5$ k_{pa}[a]

Condition (1)	E_i(MPa) (2)
Atmospheric (no vacuum)	21.03
0.5 torr	68.94
0.05 torr	68.94
0.005 torr	96.52

[a] 1 kPa = 0.145 psi.

A. Vacuum Triaxial Device

This instrument consists of a rigid cylindrical steel sleeve inside a vacuum chamber within which a loose sample could be compacted (Fig. 13). A unique feature of this device is that it allows preparation and testing of samples without exposure to atmosphere.

A polished steel sleeve, hardened to Rockwell 65 hardness number, provides the cell for the sample during compaction. The length of the sleeve is 20.3 cm (8.0 in.), the inside diameter is 7.62 cm (3.0 in.) and the outside

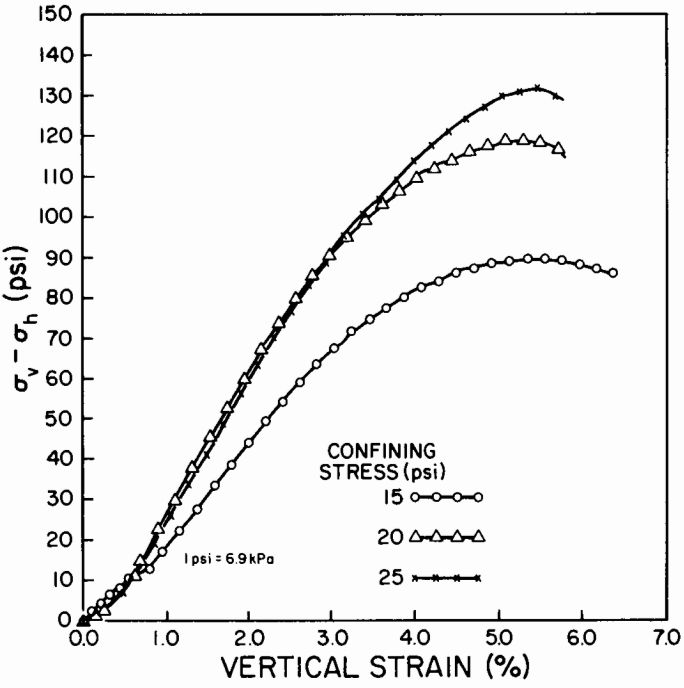


Figure 17. Stress vs strain for conventional triaxial tests.

diameter is slightly less than 11.4 cm (4.5 in.). The sleeve rests on the bottom flange during sample compaction and is raised until it contacts the top flange during triaxial testing. When the steel sleeve is fully raised, the large O-ring holding the membrane to its base contacts the tapered portion of the vessel wall. This effectively isolates the upper region of the vacuum chamber and the evacuated sample so that vacuum is maintained while the confining stress is applied. A small port in the lower region of the chamber is now opened to allow the application of air pressure as the confining stress. Vertical stress, controlled by the MTS testing system, is then applied.

B. Compaction Tests Under Vacuum

The initial density of the ALS simulant used in the tests was 1.55 g cm^{-3} (97 lb ft^{-3}). This density corresponded to the loosest state of the simulant. The simulant was heated to 130°C for a minimum of 36 hr before it was placed in the test chamber, so as to remove (adsorbed) moisture.

A triaxial compaction test was conducted in the vacuum triaxial chamber with thecyclic load amplitude varying from 2.22 kN (500 lbs) to 6.67 kN (1500 lbs). It was found that after 5000 cycles the sample density was comparable to densities achieved under much higher amplitudes after 4000 cycles in a pilot study. Therefore, the 5000-cycle loading program was

selected for compaction studies with the vacuum triaxial instrument. Figure 14 shows the loading pattern employed for the compaction phase in vacuum triaxial tests.

C. Compaction Tests Under Vacuum

Table V shows initial and final densities resulting from 1 to 5000 loading cycles at four vacuum levels. Three tests were conducted at each vacuum level. These results show that the densities increased with the number of cycles and there appeared to be only a small increase in density after about 5000 cycles. The highest density was reached for the vacuum of 0.50 torr and the lowest for vacuum of 760 torr. For vacuums higher than 0.50 torr, the final density reached was lower as vacuum increased. This is shown on Fig. 15 where the ratio of compacted density ρ to the initial density ρ_o is greatest at a vacuum of 0.50 torr compared to atmospheric and to higher-vacuum levels. One of the reasons for maximum compaction at 0.50 torr may be that with higher vacuums, the soil is not able to achieve the level of compaction possible in the presence of certain amounts of air that may act as a lubricant. These results are consistent with those obtained by Vey and Nelson (1965) who reported that the porosity of an early simulant was lowest around 10^{-3} torr and higher at atmosphere or at higher vacuums.

Stress-strain curves for the vacuum triaxial tests are shown in Fig. 16a,b and c. Figure 17 shows the results for conventional triaxial tests under atmospheric conditions. Various terms in these figures are: σ_v = vertical stress, σ_h = confining stress, $p = (\sigma_v + 2\sigma_h)/3$ = mean pressure and, $q = (\sigma_v - \sigma_h)$ = shear stress. In the standard triaxial test, the confining pressure σ_h is constant and a shear stress q is applied. Such a stress path will follow the p-axis first and then a straight-line shear path (see inserts in Fig. 16). In the case of the tests with vacuum, it was not possible to apply uniform initial confining stresses (i.e., $\sigma_c = \sigma_h$). Hence the initial path involved an anisotropic confining stress condition, which is indicated in the inserts in Fig. 16 (a) to (c).

Based on these results, the following conclusions can be made:

1. The peak shear stress $(\sigma_v - \sigma_h)$ increases with the confining stress.
2. The peak shear stresses at atmospheric pressure (Fig. 17) are higher than those for tests under vacuum.
3. Although the initial slopes (i.e., the initial moduli E_i) are different, and generally increase with vacuum, the peak shear stresses for different vacuums are not significantly different. Table VI shows typical values of E_i at $\sigma_v - \sigma_h = 345$ kPA (50 psi) with $\sigma_h = 172$ kPa (25 psi); these represent an arbitrary level of $\sigma_v - \sigma_h$. It can be seen that the material exhibits a significantly stiffer deformation response (higher E_i) under vacuum than under atmospheric conditions.
4. Plots of the Mohr-Coulomb diagram showed that there was not much variation in the angle of friction ϕ for the tests under vacuum. The

average value of ϕ was about 39 deg. The cohesion value was found to be small—about 14 kPa (2 psi) for the vacuum tests and about 55 kPa (8 psi) at atmospheric pressure—similar to the results of Vey and Nelson (1965) for powdered olivine.

D. Limitations and Future Research

In order to develop lunar construction materials of sufficient strength using the techniques described in this chapter, additional studies must be undertaken at higher vacuum. Vacuum levels may be increased by design of the membrane to reduce vacuum leaks. The very high vacuum of the lunar surface (up to 10^{-14} torr) may greatly affect the compressibility and strength of structures made from lunar materials; at this time the capacity of the available pump is limited to the vacuum of 10^{-4} torr. With the future availability of pumps that can permit higher vacuum levels, it will be possible to perform tests by using the device developed, so as to simulate vacuum levels that occur on the Moon for both dry as well as ceramic materials. Techniques such as thermal liquefaction of simulants with admixtures that can be manufactured on the Moon (powders and fibers) might be used to develop lunar construction materials. The mechanical properties of the resulting materials can then be obtained by using the device described herein, as well as by using devices for tensile, fracture and multiaxial properties.

Acknowledgments. A part of the research reported herein was performed under research grants from SERC-NASA Center at the University of Arizona, Tucson with T. Triffet, Director; and K. Ramohalli and J. Lewis, Principal Investigators. Useful comments by D. Vaniman are gratefully acknowledged.

REFERENCES

ASTM C78-84. 1990. Standard test method for flexural strength of concrete. Concrete and aggregates. In *1990 Annual Book of ASTM Standards*, vol. 04.02 (Philadelphia: ASTM), pp. 32–34.

Carrier, D. W., Cromwell, L. G., and Martin, R. T. 1973. Behavior of returned lunar soil in vacuum. *J. Soil Mech. Found. Div.* 99:979–996.

Desai, C. S., and Girdner, K. 1992. Structural materials from lunar simulants through thermal liquefaction. In *Engineering, Construction, and Operations in Space III: Proc. Space 92*, vol. 1, ed. W. Z. Sadeh, S. Sture and R. J. Miller (New York: American Soc. of Civil Engineers), pp. 528–536.

Desai, C. S., Saadatmanesh, H., and Allen, T. 1992*a*. Behavior of compacted lunar simulants using new vacuum triaxial device. *J. Aerospace Eng.* 5:425–441.

Desai, C. S., Saadatmanesh, H., and Allen, T. 1992*b*. Mechanical properties of compacted lunar simulant using new vacuum triaxial equipment. In *Engineering, Construction, and Operations in Space III: Proc. Space 92*, vol. 2, ed. W. Z.

Sadeh, S. Sture and R. J. Miller (New York: American Soc. of Civil Engineers), pp. 1240–1249.

Fuenkajorn, K. 1985. Experimental Assessment Of Borehole Wall Drilling Damage In Basaltic Rocks. M. S. Thesis, Univ. of Arizona.

Goldsworthy, W. B. 1985. Composites, fibers and matrices from lunar regolith. In *Space Manufacturing 5, Proc. of the Seventh Princeton/AIAA/SSI Conf.: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 165–171.

Goodman, R. E. 1989. *Introduction to Rock Mechanics* (New York: J. Wiley and Sons).

Guerra, L. 1988. A commonality assessment of lunar surface habitation. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 274–286.

Ho, D., and Sobon, L. 1979. Extraterrestrial fiberglass production using solar energy. In *Space Resources and Space Settlements*, NASA SP-428, pp. 274–286.

Kopecky, L., and Voldan, J. 1965. *The Cast Basalt Industry*, Annals of the New York Academy of Sciences, vol. 123, pp. 1086–1105.

Meek, T. T., Fayerweather, L. A., Godbole, M. J., Vaniman, T., and Honnell, R. 1988. Sintering lunar simulants using 2.45 GHz radiation. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 102–110.

Simonds, C. H. 1973. Sintering and hot pressing of fra mauro composition glass and the formation of lunar breccias. *American J. Sci.* 273:428–439.

Simonds, C. A. 1988. Hot pressing of lunar soil and qualification for manned applications. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 90–101.

Subramanian, R. V., Wang, T. J. Y., and Austin, M. F. 1977. Reinforcement of polymers by basalt fibers. *SAMPE Quart.*, pp. 1–10.

Taylor, S. R. 1975. *Lunar Science: A Post Apollo View* (Elmsford: Pergamon Press).

Taylor, S. R. 1982. *Planetary Science: A Lunar Perspective* (Houston: Lunar and Planetary Inst.).

Toups, L. 1990. A survey of lunar construction techniques. In *Engineering, Construction, and Operation in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 399–408.

Uhlmann, D. R., Klein, L., Onorato, P. I. K., and Hopper, R. W. 1975. The formation of lunar breccias: Sintering and crystallization kinetics. *Proc. Lunar Sci. Conf.* 6:693–705.

Vey, E., and Nelson, J. D. 1965. Engineering properties of simulated lunar soils. *J. Soil Mech. Found. Div.* 91:25–52.

Weiblen, P. W., Murawa, M. J., and Reid, K. J. 1990. Preparation of simulant for lunar surface materials. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 98–106.

PROCESSING OF LUNAR BASALT MATERIALS

B. J. PLETKA
Michigan Technological University

The development of an infrastructure for a manned lunar base requires the fabrication of components such as launch and landing pads, foundations, unpressurized structures and possibly roads. The fabrication of these structural elements using the lunar regolith is explored in this chapter. Methods appropriate for producing blocks or slabs of lunar basalt were selected based on the characteristics of the lunar regolith and the lunar surface conditions. Liquid-phase sintering appears to be the most promising technique for producing large structural components, and the key processing issues which have to be addressed in order to construct a regolith sintering facility are discussed. Melting and solidification of lunar basalt to yield glassy structures are also examined as the anhydrous lunar environment may lead to significant improvements in the mechanical properties of glasses; data in support of this hypothesis are presented.

The utilization of indigenous raw materials in the construction of a base on another planetary body can take many forms. For example, the extraction of metals or the production of advanced ceramics, such as semiconductors or solar cells, from extraterrestrial resources would certainly reduce the dependency of long range space activities on Earth support. History has shown, however, that the natural resources utilized first have to satisfy several criteria: the resources are required for basic survival, are readily available, and need the least processing to form a useful product. As a result, the focus of this chapter will be to examine processing techniques that take the layer of debris that covers the lunar surface, the regolith, and use it as a feedstock for producing simple shapes such as slabs or blocks that could be used in building an infrastructure for a manned lunar base. This approach allows useful structural materials to be produced with minimal beneficiation and processing, consistent with the desired criteria for resource utilization, and yields an economic benefit as it would be very costly to transport construction materials from Earth.

Materials processing on the lunar surface was examined because information exists as to the resources available at various sites as a result of the Apollo and Luna return missions. Such information, which is currently not available in the same detail for other planetary bodies, forms the subject matter of Sec. I. Once we understand what raw materials are available on the lunar surface, the selection of appropriate processing methods follows directly from this information.

I. RAW MATERIALS

The lunar surface is composed primarily of a layer of loose, fine- grained material known as the lunar regolith. Formation of the regolith has resulted from the breakup of surface rock by mostly small meteoroid impacts. This "weathering" of the lunar surface means that the chemical composition and mineralogy of the regolith reflect that of the underlying bedrock, which has been pulverized over the history of the Moon. In addition, because of the unique genesis of the Moon, the lunar surface can be differentiated into two distinct geologic units: the more ancient highlands which developed mostly on anorthositic bedrock, and the maria which have formed from basaltic lava flows. Excellent reviews of the chemistry, mineralogy and petrology of the lunar regolith can be found elsewhere (see, e.g., Papike et al. 1982; Taylor 1990; Heiken et al. 1991) and in this book. For the purposes of this chapter, it is sufficient to note that the mineralogy of these regions has been found to consist primarily of plagioclase feldspar [predominantly anorthite ($CaAl_2Si_2O_8$)], pyroxenes [$CaSiO_3$ and $(Mg,Fe)SiO_3$] and olivine [$(Mg,Fe)_2SiO_4$]; in addition, the important oxide mineral, ilmenite ($FeTiO_3$), is present in the maria regions. The relative abundance of these minerals in the mare basalts and anorthositic rocks is shown in Table I. As ilmenite is considered to be the primary raw material in several oxygen extraction processes (see, e.g., Cutler and Krag 1985; Williams 1985), we will concentrate on evaluating fabrication methods used to form components based upon basaltic materials.

TABLE I
Relative Abundance (vol.%) of Major Minerals in Lunar Materials[a]

	Mare Basalts	Anorthositic Rocks
Plagioclase Feldspar	15–35	40–98
Pyroxene	40–65	0–40
Olivine	0–35	0–40
Ilmenite	0–25	trace

[a] Williams and Jadwick 1980.

In addition to lithic and mineral debris, the regolith also consists of isolated glassy particles formed by impact melting or from volcanic eruptions. Glass is also present as the matrix in so-called agglutinates, which are lithic, mineral and glass fragments embedded in a glassy matrix as a result of micrometeoroid impacts. Thus, the raw basaltic materials for any processing method consist of crystalline silicate minerals plus the oxide ilmenite, interspersed with glassy material. The typical amount of glass in the lunar regolith ranges from about 10 to 20%. For example, an analysis of Apollo 11 and 12 sites, which are dominated by mare materials, indicates that approximately

17 and 22%, respectively, of the components in the regolith are glass (Heiken et al. 1991).

Although there are similarities in the composition and mineral phase distribution of terrestrial and lunar rocks, there are several aspects of lunar materials which make them unique with respect to their terrestrial counterparts and which will have an impact on the processing. First, lunar materials contain native iron (Fe^0) as a stable phase while Fe-bearing minerals do not possess any ferric iron (Fe^{3+}). Second, there is a complete absence of water in the lunar materials. Finally, lunar materials have been subjected to shock damage from meteorite impact, and radiation damage from the solar wind as well as solar and galactic cosmic rays. Implantation of certain elements (e.g., H, C and N) has also occurred in the surface of lunar materials by solar wind and galactic cosmic rays.

The cumulative grain size distribution for a large number of lunar soils is shown in Fig. 1. Most returned regolith samples have been found to possess mean grain sizes of between 45 to 100 μm, although larger mean grain sizes up to $\sim$ 800 μm have been observed.

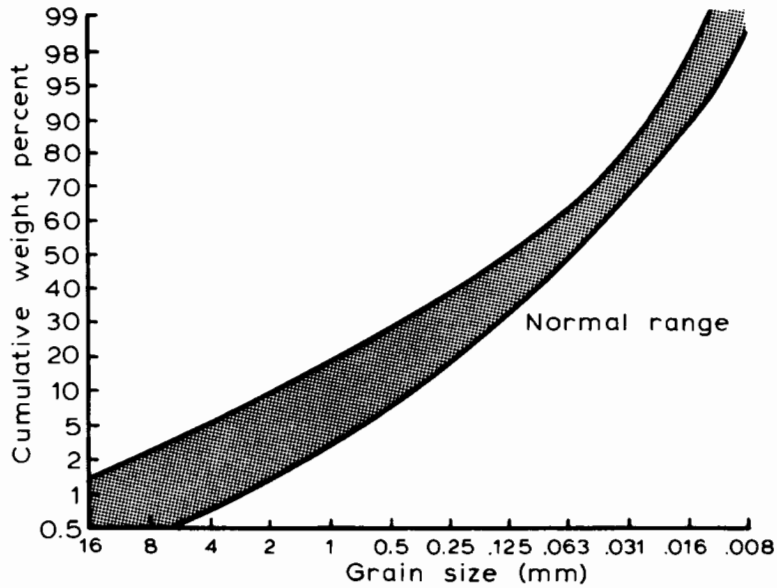


Figure 1. The cumulative grain-size distribution obtained from a number of lunar soils. (Reprinted by permission of the publisher from D. S. McKay and D. W. Ming, *Lunar Base Agriculture: Soils for Plant Growth*, D. W. Ming and D. L. Henninger, eds. Madison, Wis.: American Soc. for Agronomy, 1989.)

II. PROCESSING METHODS

A number of potential processing methods are available to fabricate components from lunar basalt materials. A major characteristic of terrestrial basalts, as they fall into the broad class of materials known as ceramics, is that they exhibit little or no plastic deformation when stress is applied to them and often fail in a catastrophic manner. As a result, basalt materials cannot be formed into a given shape using deformation processes such as rolling or extrusion commonly used for metals. Two basic processes, therefore, have evolved over the past several thousand years for forming ceramics into specific shapes with adequate mechanical strength. One process involves consolidation of a fine powder using some type of pressure to form the desired near-net shape, followed by a sintering step in which a bond is obtained between the powder particles so that a sufficiently high density (usually measured as a percentage of the theoretical density) is attained. The second basic method involves melting and casting the lunar basalt into a desired shape.

The first processing method has been used for many years to fabricate terrestrial ceramic components such as pottery and sanitary ware. The fine ceramic powder may be mixed with sufficient liquid so that appropriate rheological (plastic flow) properties are imparted to the mixture to permit shaping; the consolidation step is performed using techniques such as slip casting, extrusion, or plastic pressing operations (Reed 1988). This general method is not suitable for the lunar surface because liquid additions to the powders are required. Another type of consolidation step involves pressing the powder into the desired shape using a die or mold (Reed 1988). Organic additives are typically added to the powder to promote adhesion between the ceramic particles in the so-called green body (binders) and to reduce the friction that develops between powder particles and the mold or die wall (lubricants). This general approach can be adapted to the lunar surface, as discussed in some detail in the Battelle report, "A Preliminary Design Concept for a Lunar Sintered Regolith Production Facility" (Shirley et al. 1989). A modification of this second technique involves the simultaneous application of pressure and temperature so that the consolidation and densification are achieved in a single processing step.

Casting is a common practice for producing near-net shape metallic materials and is widely used in forming terrestrial glasses. So-called fusion casting of refractories has been used since the 1920's in the United States to produce corrosion resistant refractory linings for glass melting tanks (Bardhan and McNally 1980). Furthermore, a stable cast basalt industry has existed in Europe since the early 1920's (Kopecky and Voldan 1965). Therefore, the technology for producing terrestrial cast products is currently available. The production of lunar glass or glass ceramics using casting techniques has an additional advantage (that will be discussed in Sec. IV.B) in that the anhydrous lunar environment will not allow hydrolytic weakening of the glass to occur (Blacic 1985).

In the next two sections, we shall explore compaction/sintering and solidification in more detail as the two processing techniques for producing lunar basalt components. The application of these processes to lunar materials has been described in other sources (see, e.g., Dalton and Hohmann 1972; Steurer 1982; Simonds 1973). Our approach will be to develop a fundamental understanding of each method before considering the unique aspects of processing lunar basalt. We will also review the experiments that have been conducted to assess the applicability of these processes to simulant lunar materials; however, we shall see that much additional work is necessary before we can even begin to carry out pilot plant studies.

III. SINTERING OF LUNAR BASALT

A. Compaction Step

The compaction of the basalt in a mold to form a desired shape is an important initial processing step. The compaction process involves both the rearrangement and the deformation and/or fracture of the particles, leading to the development of interparticle contacts and a reduction in the porosity present in the ceramic compact. We want to reduce the porosity in the pressed (compacted) green body to a minimum because, during the subsequent sintering operation, this results in (1) less shrinkage; (2) a reduction in sintering time; and (3) a better chance to reach full (theoretical) density. The microstructure of the consolidated green compact also has a very strong effect on the properties of the final sintered product. In particular, factors that affect the homogeneity of the green compact have to be controlled.

Before any compaction takes place, some raw materials preparation is necessary. This would consist of a screening operation to eliminate coarse particle sizes. This preliminary step is necessary for two reasons. First, the rate at which sintering takes place is inversely proportional to particle size. Therefore, to make the overall process energy and time efficient, the consolidated powder should be composed of as fine a particle size as possible. Second, the mechanical strength is also inversely proportional to the scale of the sintered microstructure (i.e., grain or mineral size). Heat treatment (as in sintering) inevitably leads to a coarser microstructure and inferior properties. In addition, as the microstructural scale increases, the distribution of defects within the body resulting from the processing treatment increases, leading to a greater variation in fracture strength. This latter fact makes it very difficult to use brittle, ceramic materials as structural components. Therefore, it is desirable to control the initial particle size distribution of the regolith so that only finer-sized particles are utilized in the pressing operation. The simplest technique would be to use screens of varying mesh size to eliminate the coarser particle sizes. The minimum particle sizes that could be obtained would have to be established, based upon how easy it is to perform the necessary screening operations in one-sixth gravity and upon the characteristics of the

lunar regolith material; that is, how much of the regolith is finer than a certain desired minimum particle size.

The actual consolidation step is accomplished by pressing the sized powder in a rigid die or flexible mold to achieve the simultaneous compaction and shaping of the powder. The two basic techniques used in pressing are: (1) unidirectional pressing, in which compaction is achieved by means of punches in hardened steel dies (which are necessary because of the abrasive nature of ceramic materials), and (2) isostatic pressing, in which powder is pressed in flexible rubber molds using hydrostatic pressure (either a liquid or gas). The latter technique would be difficult to implement in a process which must be relatively simple, reliable, and easily automated. Thus, only unidirectional pressing into a mold or die will be considered.

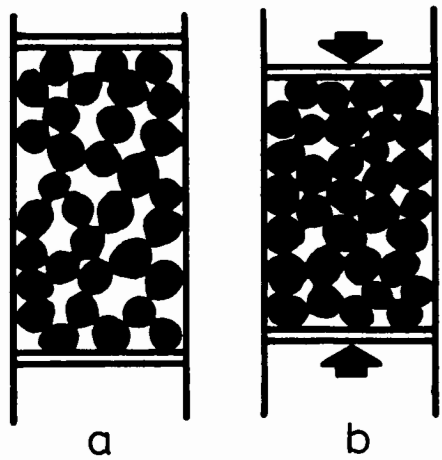


Figure 2. A simplified view of powder compaction. Note the increase in particle contacts and the reduction in pore space when pressure is applied to the filled die (a) and densification occurs (b).

Let us consider now what happens when pressure is applied to a loose array of particles in a mold. On filling the mold cavity, the loose powder has an excess of void space, no tensile or shear strength, and a low number of interparticle contacts (low coordination number), as shown schematically in Fig. 2a. As pressure is applied, rearrangement and sliding of the particles occur, leading to an increase in the number of particle contacts and a decrease in the porosity present in the compact (Fig. 2b). This leads to a rapid increase in density, as shown in Fig. 3 for some common ceramic materials. Further increases in pressure above 5 to 10 MPa typically lead to relatively little densification for ceramic powders because they can only undergo fracture; particles that are ductile, such as metals or KBr (Fig. 3), can deform plastically, and densities near 100% of theoretical can be achieved. The net result is that ceramic compacts after consolidation will typically contain 25 to 60 volume

percent porosity. No data similar to that shown in Fig. 3 is available for lunar simulant materials.

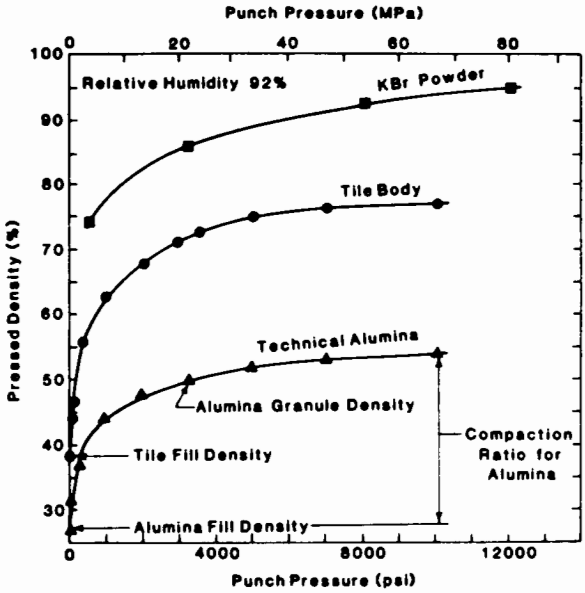


Figure 3. Pressed density as a function of punch pressure for several ceramic materials. (Reprinted by permission of the publisher from J. S. Reed, *Introduction to the Principles of Ceramic Processing*, copyright by John Wiley & Sons, New York, 1988.)

An additional goal during the compaction process is to minimize or (ideally) to have no variations in density throughout the body. Minimizing density gradients within the body is very important because density gradients lead to nonuniform densification during sintering. This can produce residual stresses, warping of the component, or the development of cracks within the body.

Organic materials (binders) are often added to the powder in terrestrial compaction operations to provide a temporary bond between ceramic particles, giving the green body enough strength so that it can be handled prior to sintering. Whether sufficient strength can be achieved after pressing lunar basalt without the addition of a binder is an open question. Tests of the mechanical properties of lunar regolith samples have shown that the regolith can be very cohesive (Mitchell et al. 1972). Thus, it may be possible to form green bodies with sufficient strength, although the size of the component is also an important determining factor. Compaction experiments conducted on lunar simulants and returned regolith samples are required.

Lubricants are also typically added to ceramic powders in pressing operations to (a) reduce die wear and ejection pressures by minimizing die wall

friction; (b) aid in redistribution of particles during pressing to obtain maximum packing; and (c) improve flow of the powder into the die. The first factor is particularly important when pressing components that have large length to diameter (L/D) ratios. Friction is inherent in die compaction and places significant limitations on the sizes and shapes that can be fabricated. The die wall friction occurs because stress applied by the punches is transmitted to the die walls. Thus, frictional stresses oppose the applied stress, leading to stress gradients and, ultimately, green density gradients. The severity of the friction between the powder and the wall depends upon how extensively the powder becomes imbedded in the mold wall or fits with the microscopic roughness of the wall. The wall friction decreases if the mold wall is hard and if its surface is finished to a high degree of smoothness. Because of the abrasive nature of ceramic powders, any metal mold material will be gradually scratched and friction between the powder and mold wall will slowly increase. This is an inherent problem when pressing ceramic powders, including any lunar material. Reductions in the density gradients can be achieved without the addition of lubricants simply by using low L/D ratios. This should be possible when pressing components such as slabs or bricks, and the addition of lubricant to lunar materials will probably not be necessary.

B. Sintering Step

Sintering is a term applied to an elevated temperature heat treatment of a porous material in which the particles form strong bonds with a concurrent reduction in the volume of pore space. Therefore, in order for sintering to occur, a mechanism for material transport must be present. The two dominant transport mechanisms that occur in ceramic powders are diffusion and viscous flow, although other mechanisms may also be present (Kingery et al. 1976); in particular, vapor transport may be relatively important in a vacuum sintering condition (which would exist on the lunar surface). On the basis of the transport mechanisms, sintering can be categorized as either solid-state sintering or liquid-phase (viscous) sintering. We will examine each sintering mechanism in turn and show that liquid-phase or viscous sintering is the technique that would be applicable to lunar basalt materials.

Solid-state sintering involves material transport by diffusion. Rapid diffusion rates are achieved by heating the compacted powder to a sufficiently high temperature (typically greater than half of the melting temperature) so that the initial bridges or necks which exist between particles grow as a result of mass transport to these regions. The mass transport can take place by the movement of atoms or vacancies along a surface or grain boundary or through the bulk of the material. The driving force for this process results from the fact that the powder particles are in contact with one another in the compacted body but a significant amount of pore space exists between the particles (typically 25 to 60 volume percent porosity). At high temperatures, when sufficient atom mobility is present, atoms diffuse to the initial particle contacts or bridges to eliminate the solid/vapor interfaces that exist within the

compact, and replace these interfaces with lower energy solid/solid interfaces (see Fig. 4). As atom diffusion (or other transport mechanisms) permit the particles to be bonded together, the pore space originally present shrinks and may be eliminated if sintering is carried out for sufficiently long times.

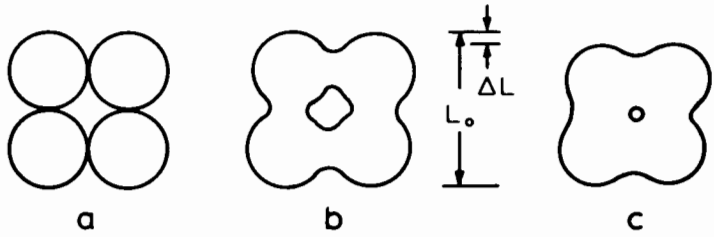


Figure 4. Schematic illustration demonstrating the microstructural changes that occur during sintering. (a) Powder particles after pressing. (b) Particle coalescence results as sintering proceeds with a change in pore shape but no shrinkage. (c) Pores change size and shape as shrinkage takes place.

Although a variety of transport mechanisms exist, only atom motion from the original particle contact to the particle/pore surface succeeds in causing pore shape changes and decreases in the center-to-center distance of powder particles (Fig. 4; Kingery et al. 1976). However, atom mobility requires diffusion, which is a strong function of temperature. Hence, the rate of sintering depends strongly on temperature and the original size of the particles [because the surface area to volume ratio of a sphere (particle) increases significantly as the sphere diameter decreases, which increases the driving force for sintering]. It would be difficult to predict theoretically the rate of sintering in a material such as basalt because of its multi-element and multi-component nature. It would be expected, however, that because of the strong directional covalent bonds existing within the various silicate minerals, solid-state diffusion would not be rapid even at temperatures approaching the melting points of some of the mineral constituents. Little densification should be achieved, therefore, in these materials.

Experiments to confirm this hypothesis have been conducted in our laboratory. Model experiments were carried out in which a lunar simulant material was densified; the simulant selected was the Minnesota lunar simulant (MLS-1) because its bulk chemistry and mineralogy resemble a high titanium mare basalt (Weiblen and Gordon 1988; Weiblen et al. 1990). The as-received rock material was crushed and ground to produce a powder size distribution typical of the lunar regolith. It is important to recognize, however, that no glass was present in the ground material, only crystalline phases. The powder was hydrostatically pressed into rods measuring approximately 100 mm in length and 10 mm in diameter using a pressure of 345 MPa. The rods were then sintered under an argon (inert) atmosphere. Sintering times ranged from 8 to 72 hours and the temperatures were varied from 1000 to 1110°C

(see Table II). The microstructures of the as-received and sintered samples were examined using optical microscopy after petrographic thin sections were prepared. Compression tests were also performed on selected specimens. Parallelepiped specimens were cut from the original rock material and selected sintered specimens, and mechanical tests were conducted using a hydraulic testing machine at a crosshead velocity of 1.5 mm/minute.

TABLE II
Mechanical and Microstructural Properties of MLS-1

Sintering Temperature (°C)	Sintering Time (hr)	Compressive Strength (MPa)	Porosity (Volume %)
Fine-grained as-received rock		354[a]	
Coarse-grained as-received rock		86[b]	
1064	72	90	50
1070	24	59	30
1080	8	188	26
	24	132	18
	72	223	18
1090	8	212	16
	64	172	19
1100	8	194	23
	24	119	31
1110	8	—	40

[a] Average of 5 specimens.
[b] Average of 3 specimens.

Two distinct microstructures were identified in the as-received rock specimens: a fine-grain and a coarse-grain structure. Representative micrographs of these structures are shown in Fig. 5. The difference in constituent size had a significant effect on the compressive strength; an average compressive strength of 354 MPa was obtained for the fine-grain basalt, but a much lower

compressive strength of 86 MPa was determined for the coarse-grain basalt. These results are not surprising because a reduction in the microstructural scale produces strengthening in most materials. These results do indicate, however, that a variability in strength exists in the original simulant rock.

No appreciable sintering (reduction in porosity) occurred at temperatures $<1064°C$. For this reason, compression testing was performed only on samples sintered at or above 1064°C. Figure 6 shows a representative microstructure of the sintered MLS-1. A reduction in the constituents' size is apparent (compare with Fig. 5). Extensive porosity was present even in the specimens sintered at 1064 to 1110°C and ranged from 16% to 50% (see Table II). Sintering experiments at temperatures greater than 1110°C resulted in the melting of one or more constituents of the simulant material.

The compressive strengths of the MLS-1 sintered between 1064° and 1110°C fell between the values measured for the fine and coarse grain as-received rock specimens. This result, coupled with the extensive porosity existing in the sintered MLS-1, led us to conclude that extensive bonding had not occurred between the powder particles. Therefore, the fabrication of lunar basalt components via solid-state sintering does not appear feasible.

Liquid-phase or viscous sintering is a densification process which occurs with the aid of a viscous liquid phase and is the major densification process for the great majority of silicate systems. A viscous liquid silicate is formed at the sintering temperature and serves as a bond for the particles. The viscous liquid can originate from two sources: (1) prior glass that existed within the regolith, and (2) new glass that forms as a result of the sintering process from thermochemical reactions between minerals present in the regolith.

The driving force for viscous sintering arises from the small negative radius of curvature at the neck or contact point between two particles compared with the surface of the particles. This produces a negative pressure in the region of the neck and causes viscous flow of the glassy material into the pore region. If the liquid thoroughly wets the solid particles, densification is achieved from (1) particle rearrangement via liquid flow, and (2) solution/reprecipitation, whereby small particles dissolve and reprecipitate on larger particles. Rapid sintering results because the liquid provides for rapid transport compared to diffusion in the solid state. From an analysis originally due to Frenkel (1945), the initial shrinkage rate, $\Delta L/L$, has been derived as

$$\frac{\Delta L}{L} = \frac{3\gamma}{4\eta r} t \tag{1}$$

where t is the time, γ is the surface tension, η is the viscosity, and r is the particle radius. It is obvious from Eq. (1) that the key process variables are the viscosity and the particle size; the surface tension of silicate materials is usually a weak function of composition. The particle size has a strong effect on the sintering rate because of the inverse relationship between $\Delta L/L$ and r in Eq. (1). Of even greater importance is the viscosity, because it can be

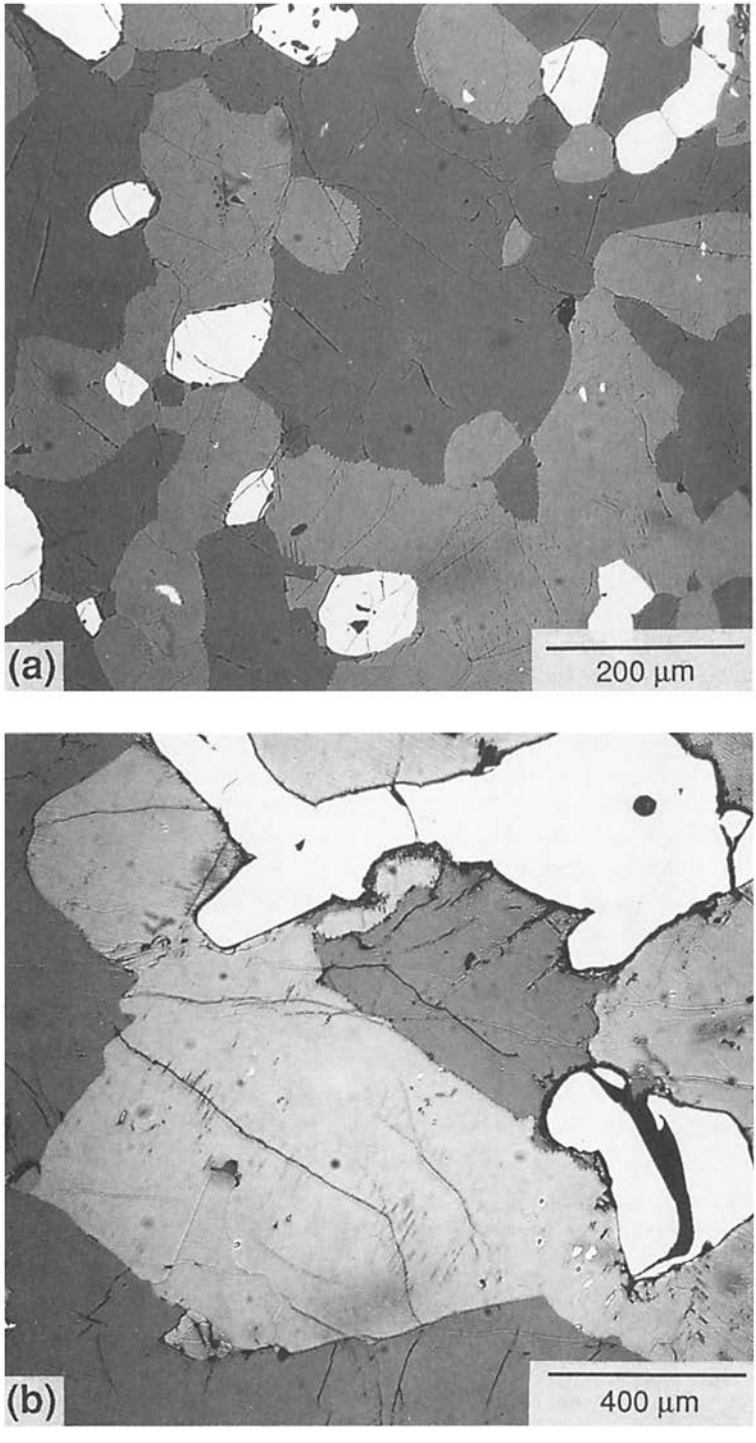


Figure 5. Optical micrographs of (a) as-received fine grain rock and (b) as-received coarse grain rock.

controlled by changing the composition of the liquid and because the viscosity is strongly dependent on temperature through the relationship

$$\eta = \eta_0 \exp\left(\frac{E_n}{RT}\right) \tag{2}$$

where E_n is the activation energy for viscous flow, R is the gas constant, η_0 is a constant, and T is the temperature. It is not uncommon for the viscosity of a glass to change by a factor of 1000 over a temperature interval of 100°C. It is also important to recognize that the viscosity in Eq. (1) is the viscosity of the overall body, which depends not only on the magnitude of the viscosity as described by Eq. (2), but also on the amount of liquid present in the system, which can also change with temperature.

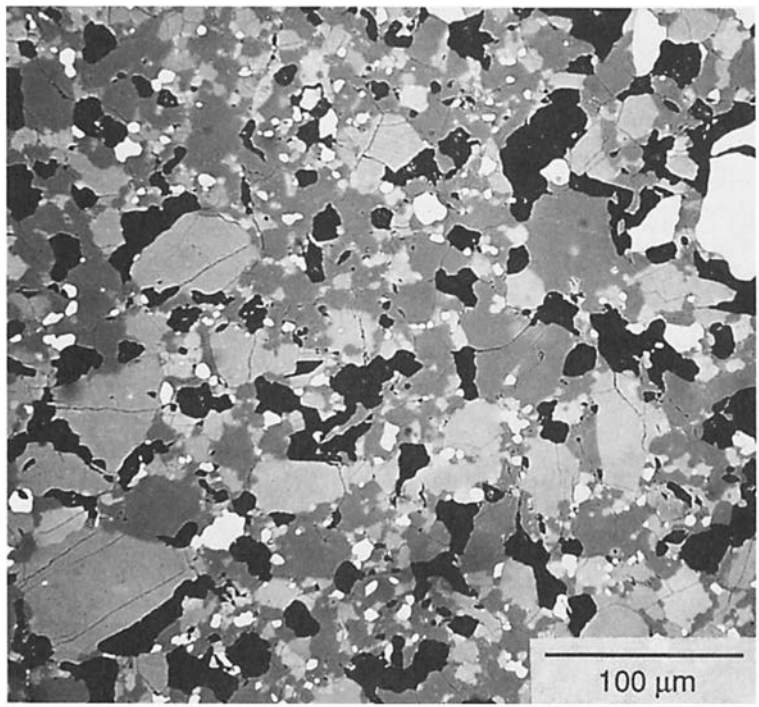


Figure 6. Representative micrograph of the sintered MLS-1 simulant.

Because of the strong temperature dependence of the viscosity, it is possible to lower the viscosity to a value whereby deformation of the ceramic body can occur from gravitational forces. A balance must be struck, therefore, in which the stresses due to surface tension (controlled by the particle size) are substantially larger than the stresses due to gravitational forces. This problem will be of less importance on the lunar surface due to the reduced gravity.

The formation of a liquid required for liquid-phase sintering can take place simply by one of the mineral components melting. However, because

of the phase equilibrium that exists between the various constituents in the regolith, formation of a liquid can also occur at temperatures much lower than the melting temperatures of the component minerals. This latter mechanism occurs in actual silicate systems. It is easiest to visualize this process using a phase diagram.

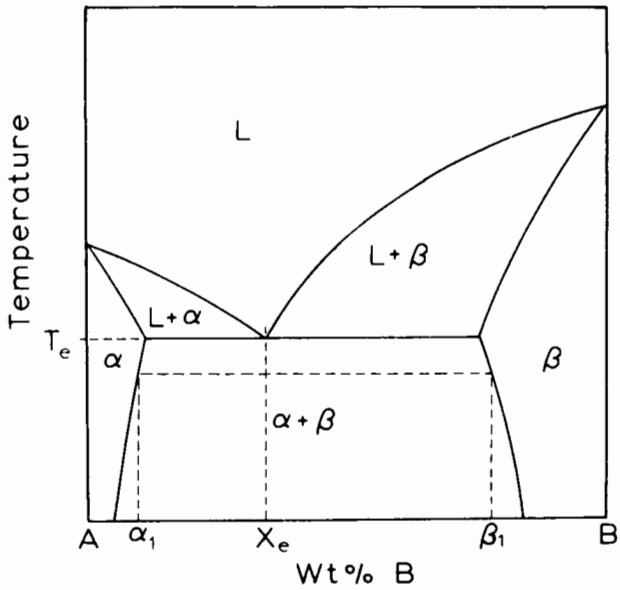


Figure 7. Hypothetical binary A-B phase diagram. See text for a discussion of the phase fields and phase compositions.

A pure substance may exist as a vapor, a liquid, or a crystalline solid, depending upon the temperature and pressure. A phase diagram indicates, at a given temperature and pressure, which phase or phases may be present. A phase diagram may be thought of, therefore, as a map indicating which stable phases exist for a given set of conditions. Phase diagrams become more complex when more than one component is present because composition now becomes a variable in addition to temperature and pressure. Often, the pressure is fixed because engineering materials are used at a constant pressure; thus, for a two-component system the only variables are temperature and composition. Such a hypothetical phase diagram is shown in Fig. 7. There are six distinct phase fields in the diagram: three single-phase regions existing over a range in composition and three two-phase fields in which the composition of each phase at a given temperature is fixed. Consider the bulk composition, X_E. At a temperature just below T_E, the structure of the material is composed of two phases, α and β. The compositions of the α and β phases are determined by drawing a horizontal line at a given temperature (a so-called

tie line) whose intersection with the phase boundaries gives the composition of the α and β phases at the desired temperature. Note that these compositions, α_1 and β_1, respectively, differ from the bulk composition. Further heating causes the compositions of the α and β phases to change slightly until at the eutectic reaction isotherm, T_E, the α and β phases melt to form a liquid (the eutectic liquid) with composition X_E. It can be clearly seen that the melting temperature of this composition is less than the pure components A and B. Such an eutectic reaction causes liquid formation, therefore, at temperatures lower than the melting point of the pure components in binary and higher-order systems.

There have been several studies in which simulant lunar material has been densified via liquid-phase sintering. Simonds (1973) showed that glass powder based on the composition of breccia 14049 can be densified at temperatures of 795°C or greater; densities ranging from ~50% to 90% were achieved for samples composed of different particle sizes. Devitrification (crystallite formation) was found to begin, however, at temperatures only about 50°C higher than required for sintering to begin. Thus, the amount of glassy material available for viscous flow decreased and densification became more difficult, if not impossible. Uhlmann and co-workers (1975) have examined the conditions necessary for the formation of lunar breccias to take place and have concluded that sintering will stop once the glass is completely crystalline. This conclusion is in agreement with the lack of densification achieved via solid state sintering in the Minnesota lunar simulant study discussed previously. Meek and co-workers (1985,1988) have also consolidated lunar simulant material utilizing microwave heating; this form of sintering will be examined in more detail shortly.

C. Processing Issues

There are a number of processing factors or issues which need to be examined if an automated facility for the production of liquid-phase sintered basalt blocks or slabs is to be constructed on the lunar surface. These issues have been considered in some detail in the Battelle report (Shirley et al. 1989) and have impacted Battelle's radiant furnace design concept. Several of these issues have already been discussed in this chapter, namely whether the strength of the compacted green body will be sufficient to allow it to be handled without the addition of binders (which would have to be imported from Earth unless inorganic binders could be used) and the density variations which result in the green body from pressure gradients arising from particle/die wall friction. In the following, other processing issues will be addressed.

The minimum time and temperature conditions for sintering need to be established in order to minimize the energy requirements for the process and to minimize the production time. The minimum sintering conditions will be a function of what mechanical properties are required for the various proposed structural components because the microstructural state controls the mechanical properties. This work would have to combine materials development

with structural engineering studies so that material properties are customized for particular structures. This approach will allow the minimum amount of energy to be used in the material fabrication process while developing a safe and efficient structure.

The other key processing issues relate to how fast the green body can be heated to the sintering temperature and how slowly the sintered product must be cooled down. The first factor results from the temperature gradients that form through the cross section of the green body due to the low thermal conductivity of the regolith; for example, Langseth et al. (1976) have reported a value of 1.5×10^{-4} W cm^{-1} K^{-1} for the thermal conductivity at a depth of 1 m at the Apollo 15 landing site. Calculations of the temperature gradients that form in a block 10 cm thick were made in the Battelle report (Shirley et al. 1989) using a value of 2.2×10^{-3} W cm^{-1} K^{-1} for the thermal conductivity. Figure 8 shows that the surface needs very little time to achieve a temperature of 1100°C, while the center of the block requires about 9 hr. This calculation represents an upper bound case, since the thermal conductivity of the basalt should increase as sintering takes place and more grain contacts are formed throughout the structure. Nonetheless, if temperature gradients such as those shown in Fig. 8 arise, not only is the processing time increased, but the temperature gradients can lead to density gradients that will adversely affect the integrity of the component. That is, because the surface reaches the sintering temperature sooner, uneven shrinkage takes place through the cross section of the body, leading to warping and possibly crack formation.

After sintering has been completed, cooling the body very rapidly can again lead to temperature gradients if the outer surface cools much more rapidly than the interior. Because of the high and often anisotropic thermal expansion that ceramic materials exhibit, this set of conditions can lead to the development of thermal stresses and can possibly cause fracture (thermal shock). Thus, cooling times of several hours are required in order to avoid this potential problem.

D. Alternate Sintering Methods

Liquid-phase sintering of lunar basalt can be achieved in a conventional radiant sintering furnace. This was, in fact, the type of furnace considered in the Battelle analysis (Shirley et al. 1989). One drawback with this approach is the long residence time in the furnace required to heat and cool the components. An alternative approach would be to use microwave heating. In contrast to radiant heating, the thermal gradient in microwave heating is reversed. That is, heat is generated within the body instead of utilizing external heating sources. Larger components may be easier to fabricate because (1) heating can take place uniformly on a macroscopic scale if the phase that is coupling strongly to the microwaves is distributed evenly within the ceramic body; and (2) thermal conduction is required over much shorter distances. Microwave sintering of basalt may involve the formation of a thin liquid on the surface of the particles as the rapid rise in absorption of microwave energy in polycrystalline ceramics

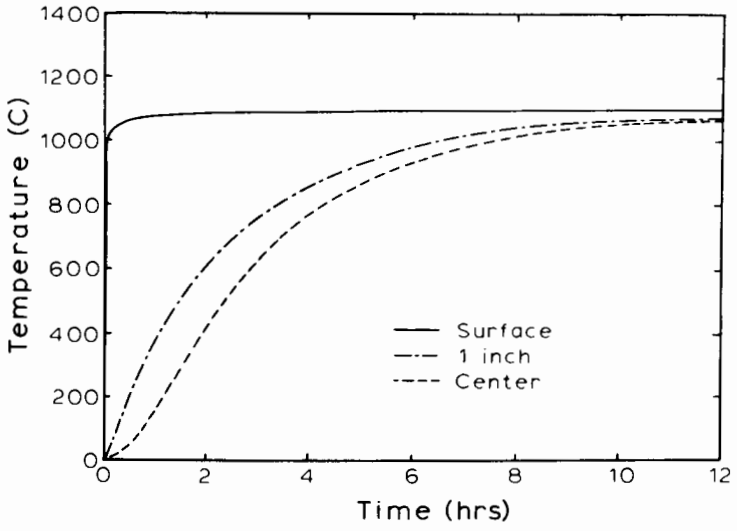


Figure 8. Calculated temperature gradients which form in a 10 cm thick block heated to 1100°C. (Reprinted with permission from F. Shirley, C. Buoni, J. W. Lennon, E. J. Mezey and A. E. Weller in "A Preliminary Design Concept for a Lunar Sintered Regolith Production Facility," Lockheed Engineering and Science Company, 1989.)

is attributed to the softening of intergranular, amorphous phases (Ho 1988). Thus, the sintering process may combine liquid-phase sintering at a local level with solid-state sintering. Meek and co-workers (1985) have found that ilmenite couples strongly with the ultrahigh frequency microwave field so that basalts can be densified. It has also been shown that as densification proceeds, heating via microwave coupling tends to concentrate at pores and other defects, leading to faster and more complete densification (Tucker et al. 1985). This latter effect can potentially allow ceramic bodies to be sintered at lower temperatures than required in conventional radiant furnaces (although the intergranular temperatures may be much higher because of the localized heating of the intergranular, amorphous phases). Densification may also be accelerated in lunar basalt because of the radiation and shock damage that exists in actual lunar materials.

Densification can also be achieved by the simultaneous application of pressure at elevated temperature. Such processing is typically done in a hot press in which powder is poured into a die (usually graphite), heated to an elevated temperature and then compressed. The application of the pressure gives rise to an additional driving force for densification by increasing vacancy transport through local changes in vacancy concentrations. This process has been shown to result in equivalent properties at lower hot-pressing temperatures than simple compaction/pressureless sintering. For example, Simonds (1973) demonstrated that extensive densification can take place at tempera-

tures about 200°C lower when glass samples are hot pressed. However, the operation of such an apparatus is more complex, less amenable to automation, and does not represent as viable an option to radiant heat sintering as microwave heating does.

V. SOLIDIFICATION OF LUNAR BASALT

Glassy materials can be formed from many silicate and oxide minerals if they are cooled at sufficiently high rates. Glass is commonly manufactured in terrestrial processing operations but is not utilized for structural applications because it is a very brittle, relatively weak material with only minimal resistance to crack propagation. It is believed that these poor mechanical properties are the result of a stress corrosion effect (hydrolytic weakening) caused by the water vapor present in the Earth's atmosphere (Michalske and Freiman 1983). It has been proposed that glass formed in an anhydrous atmosphere may possess extraordinary mechanical properties (Blacic 1985). Thus, although melting and solidification of lunar basalt is more complex than a sintering operation, the production of fully dense glassy basalt or even glass-crystalline basalt mixtures may be desirable for fabricating certain structural elements such as pipes or cables.

A. Kinetics of Lunar Glass Formation

Production of a glass requires that sufficiently high cooling rates be imposed on the melt to prevent the nucleation of crystalline material. For metallic materials, this may require cooling rates on the order of 10^5 to 10^6 °C s^{-1}. Glasses can be produced from silicate and oxide melts at cooling rates that are easy to achieve in industrial practice. The approach most commonly used to determine whether a glass will form on cooling a liquid is a kinetic approach, due in large part to the theoretical and experimental efforts of Uhlmann (1972). That is, the question of whether a glass can be formed is answered by considering how fast a particular material has to be cooled in order to prevent crystallization.

The primary tool used to determine whether a given cooling procedure will result in a glass or crystalline material is the time-temperature-transformation (TTT) diagram. This diagram was first developed to allow the decomposition of austenite to various reaction products in steels to be categorized as a function of time and temperature; samples are quenched from a higher-temperature single-phase region and isothermally reacted at a given temperature for various periods of time before being quenched to room temperature. The amount and type of decomposition product are then determined using various characterization techniques. It was recognized that a similar approach could be used to predict glass formation (Uhlmann 1983). Measured growth rates and calculated nucleation rates were used to determine the time required to obtain a certain volume fraction crystallized at various temperatures. The plotted data, shown in Fig. 9 from the work of Uhlmann

et al. (1975) for lunar composition 70019, exhibits the usual "C" curve shape for diffusion controlled transformations. At higher temperatures, sufficient diffusion or lower viscosity is available for transport (growth), but the small undercooling below the transformation or crystallization temperature does not provide a large driving force for nucleation. At much lower temperatures, greater undercooling increases the driving force for the transformation to proceed, but the decrease in diffusion or increase in viscosity limits growth. An intermediate temperature results in optimum nucleation and growth, and produces a minimum (the so-called "nose") in the TTT curve. This approach can be further modified to develop so-called continuous cooling transformation curves which take into account the effect of a continuous cooling rate (which is more realistic than the isothermal reaction kinetics of the TTT curves) on crystallite formation (Fig. 9). Based on this approach, critical cooling rates for a variety of lunar compositions could be calculated (Table III). These rates vary because the heat of fusion, liquidus or melting temperature, and glass transition temperature are functions of composition.

TABLE III
Critical Cooling Rates for Glass Formation Based on CT Curves[a]

Material	Critical Cooling Rate ($^{\circ}C\ s^{-1}$)
Apollo 15 Green Glass	89
Lunar Comp. 15286	2
Lunar Comp. 15498	1.8
Lunar Comp. 60095	7.5
Lunar Comp. 65016	16
Lunar Comp. 70019	3.8
Lunar Comp. 79155	8
Anorthite	14

[a] Uhlmann and Onorato 1979.

From a practical viewpoint, this information can be used to estimate the maximum sample size y which would be entirely glassy, using the relation (Uhlmann 1983):

$$y \approx (D_{th} t_N)^{1/2} \tag{3}$$

where D_{th} is the thermal diffusivity and t_N is the time required to reach the nose of the TTT curve. Estimating D_{th} with a value of 4×10^{-3} $cm^2\ s^{-1}$, maximum sizes on the order of 10 cm are obtained from the data of Uhlmann et al. Thus, while large glassy components are not possible, smaller glassy structural elements such as pipes and cables could be produced.

In addition to the work performed by Uhlmann et al., several studies have shown that glassy materials can be produced from simulant lunar basalt (Goforth 1976; DeLa'O et al. 1989). The work of Goforth is particularly

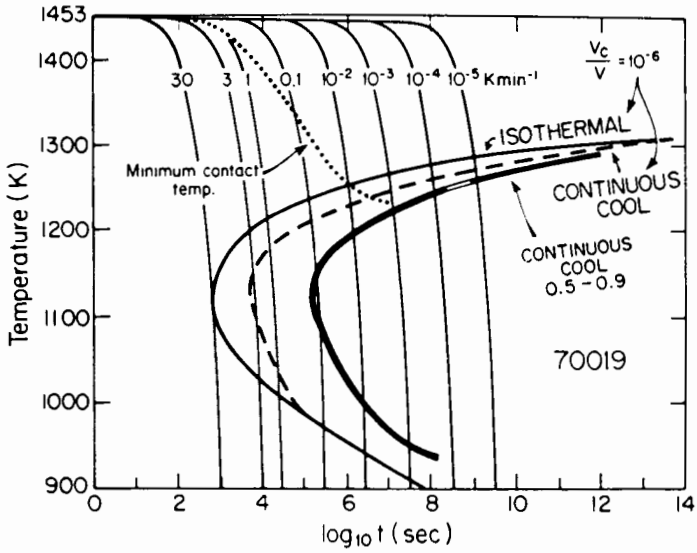


Figure 9. Isothermal (TTT) and continuous cooling (CT) curves for lunar composition 70019. (Reprinted by permission of the authors: D. R. Uhlmann, L. Klein, P. I. K. Onorato and R. W. Hopper. In *Proc. of the Sixth Lunar Science Conf.: Vol. 1: Mineralogical and Petrological Studies,* Pergamon Press, Elmsford, N. Y., 1975.)

interesting in that the TTT curve he determined would be data necessary to design a production facility.

B. Glass Forming Techniques

A variety of forming methods are used to fabricate terrestrial glass products. Typically the glass is melted in a furnace and cooled to a temperature sufficient to increase the viscosity of the glass melt so that it flows under moderate pressure, which is exerted either mechanically or with an air blast. Glass tubing and rods can be manufactured by a drawing process, while glass fibers are commonly formed by drawing glass from platinum containers with a large number of small nozzles. Thus, the technology exists to manufacture a variety of glass components, although a design concept study for a lunar glass production facility has not been undertaken similar to the Battelle report on a lunar sintered regolith production facility (Shirley et al. 1989).

C. Mechanical Properties of Lunar Glass

There is some evidence to support the hypothesis that glass processed from lunar regolith will possess mechanical properties superior to terrestrial glasses. In a recent study, the fracture toughness of simulated lunar glasses was evaluated using the microindentation technique (Carsley et al. 1992). Simulated lunar mare and highlands glass compositions shown in Table IV were prepared using reagent grade chemicals. An ultra-high vacuum system was constructed

in which melting and solidification as well as fracture toughness testing of these simulants were performed without exposure to air humidity. An electron gun heater was constructed for the purpose of melting the glass charges under vacuums of 10^{-7} torr. After solidification of the glasses, indentation was conducted with both Vickers and Knoop microhardness indenters. Measurements of the radial surface traces of half-penny cracks which develop along the Vickers indenter diagonals (see Fig. 10) were used to calculate fracture toughness (Anstis et al. 1981).

TABLE IV
Glass Compositions

	Mare Glass (10089)	Highlands Glass (69999)
SiO_2	41.5	45.4
TiO_2	7.5	0.4
Al_2O_3	15.8	26.8
FeO	15.8	5.6
MnO	0.3	0.1
MgO	7.7	6.1
CaO	11.1	14.5
Na_2O	0.5	0.6
K_2O	0.1	0.1
Cr_2O_3	0.3	0.1
P_2O_5	0.1	0.1

Fracture toughness (K_{IC}) values for the highlands and mare simulants, which were melted and tested under different environmental conditions, are shown in Table V. Although these data support the hypothesis that increases in the toughness of glasses occur in anhydrous environments (melted and tested in vacuum), there was a significant change in the composition of the vacuum-melted samples. The simulants lost all volatile elements (Na, K, S and P) as well as significant amounts of iron, silicon, and chromium. Repeated melting of the glass charges in vacuum resulted in the formation of spinel crystallites on both the free and substrate surfaces of the glasses, as shown in Fig. 11. Therefore, the changes in fracture toughness may be due to a combination of the reduction in hydrolytic action, the change in composition, and the influence of the spinel crystals on the indentation crack morphology. Clearly, more experimental work is warranted to investigate these results.

D. Glass-Ceramic Basalt Materials

Although any ceramic which contains a mixture of glassy and crystalline material could be called a glass ceramic, this term is usually reserved for materials produced by the controlled crystallization of certain glasses to yield a structure composed of predominantly small crystallites (usually less than 1 μm in size) with a small amount of residual glass (typically 2–5 vol%) in a pore-free

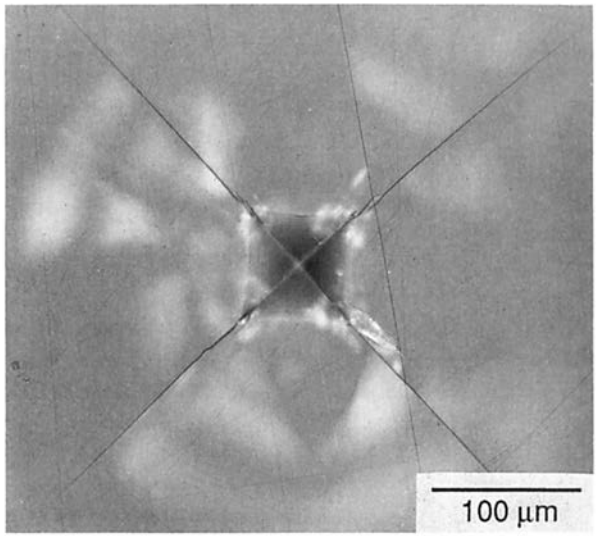


Figure 10. Optical micrograph illustrating the crack morphology about a Vickers indentation in the polished surface of an argon-melted simulant mare glass.

material. Conventional crystallization in glasses invariably initiates at external surfaces that serve as heterogeneous nucleation sites. This is followed by crystal growth into the glassy phase, producing a nonuniform microstructure of large grain size. In order to achieve the desired small crystallite size, which usually enhances the mechanical properties of the material, a high density of nucleating sites (agents) must be introduced during the melting operation. This is then followed by a controlled heat treatment in which a large dispersion of fine nuclei is achieved by heat treatment at a lower temperature than the subsequent growth step of the crystallization procedure (see Fig. 12).

Nucleation in terrestrial glass ceramics is achieved by the addition of Pt group and noble metal nucleating agents as well as oxides such as TiO_2 and ZrO_2. The metal nucleating agents seem to function by directly forming a crystalline nucleating phase. One of the most important differences between terrestrial and lunar basalt is the ubiquitous presence of native iron as a stable phase. The iron particles are on the scale of 40 to 200Å and form as a product of reduction by solar-wind-implanted hydrogen during meteoroid impact melting. Thus, a potential nucleating agent of required microstructural scale is already present in the lunar basalt. The processing of lunar basalt by glass-forming techniques, therefore, may lead invariably to the production of lunar glass-ceramic basalt possessing mechanical properties superior to the lunar glass. Unfortunately, no heat treatment of actual returned lunar material or of a suitable lunar simulant has been conducted to assess whether such

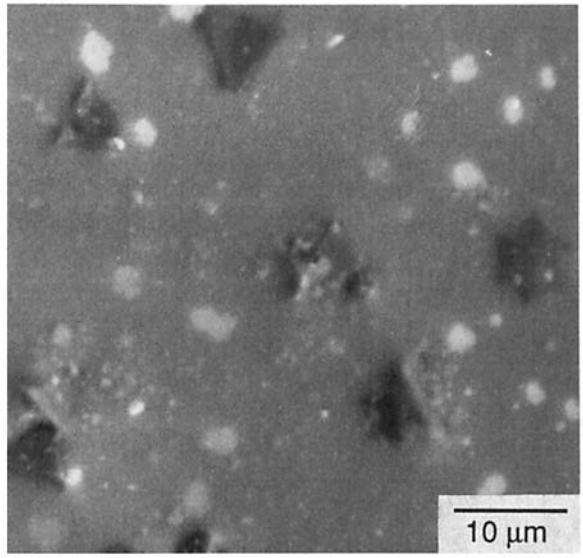


Figure 11. Scanning electron micrograph of the free surface of a vacuum-melted simulant mare glass. The angular (triangular and star-shaped) particles are spinel crystals.

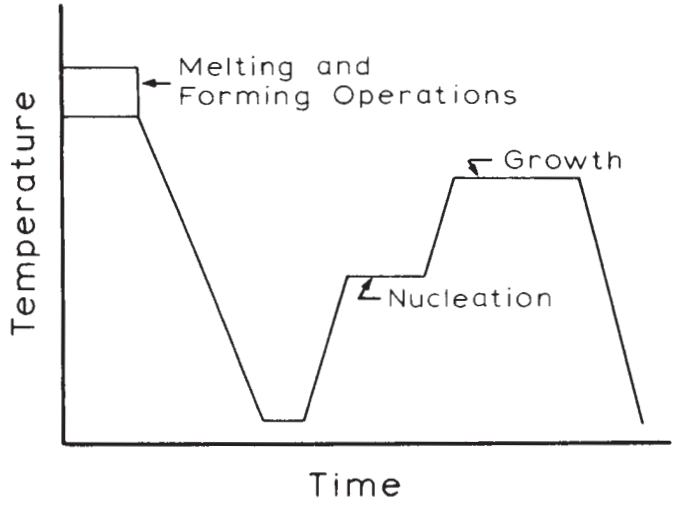


Figure 12. Schematic time-temperature profile for the controlled crystallization of a glass ceramic.

TABLE V
Fracture Toughness Values of
Lunar Simulant Glasses

Test Condition	Fracture Toughness: MPa · $m^{1/2}$	
	Mare Glass	Highlands Glass
Melted in argon and tested in air	0.87	0.95
Melted in vacuum and tested in air	1.01	1.09
Melted in vacuum and tested in vacuum	2.64	2.83

materials can be produced.

VI. SUMMARY

The processing of slabs or blocks of lunar basalt should be possible utilizing liquid-phase sintering. A number of key processing issues were identified that will have an impact on the design of a sintered lunar basalt-production facility. Microwave heating may be a viable alternate sintering technique, although radiant heating can be used. In addition, the production of glassy basalt utilizing melting/solidification also appears to be feasible, although the maximum cross section for an entirely glassy material is limited. The presence of native Fe in the regolith may make it possible to obtain controlled crystallization of glasses to produce glass ceramics. Some preliminary data were presented which indicate the fracture toughness of lunar glass may be superior to that of terrestrial glass. It seems, therefore, that the lunar basalt can be used as a feed stock to produce structural elements. Many more studies, however, utilizing simulant and returned lunar materials are required in order to develop the database necessary so that the design and construction of pilot plant facilities are possible.

Acknowledgments. Financial support for a portion of the work performed in the author's laboratory was provided by the National Aeronautics and Space Administration. This support is gratefully acknowledged.

REFERENCES

Anstis, G. R., Chantikul, P., Lawn, B. R., and Marshall, D. B. 1981. A critical evaluation of indentation techniques for measuring fracture toughness: I, Direct

crack measurements. *J. Amer. Ceramic Soc.* 64:533–538.

Bardhan, P., and McNally, R. N. 1980. Fusion-casting and crystallization of high temperature materials. *J. Mater. Sci.* 15:2409–2427.

Blacic, J. D. 1985. Mechanical properties of lunar materials under anhydrous, hard vacuum conditions: Applications of lunar glass structural components. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 487–495.

Carsley, J. E., Blacic, J. D., and Pletka, B. J. 1992. Vacuum melting and mechanical testing of simulated lunar glasses. In *Engineering, Construction, and Operations in Space III: Proc. Space 92*, vol. 2, eds. W. Z. Sadeh, S. Sture and R. J. Miller (New York: American Soc. for Civil Engineers), pp. 1219–1231.

Cutler, A. H., and Krag, P. 1985. A carbothermal scheme for lunar oxygen production. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 559–569.

Dalton, C., and Hohmann, E., eds. 1972. *Design of a Lunar Colony.* Report prepared by the University of Houston, NASA/Manned Spaceflight Center and Rice University under NASA Grant NGT 44-005-114.

DeLa'O, J. D., Hellawell, A., Pletka, B. J., and Rose, W. I. 1989. Solidification and crystallization kinetics of synthetic and lunar simulant basalts. In *Space Manufacturing 7, Proc. of the Ninth Princeton/AIAA/SSI Conf.: Space Resources to Improve Life on Earth*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 351–357.

Frenkel, J. 1945. Viscous flow of crystalline bodies under the action of surface tension. *J. Physics (USSR)* 9:385–391.

Goforth, R. E. 1976. Thermally activated nucleation and grain growth of basalt. *SAMPE Quart.*, pp. 34–42.

Heiken, G., Vaniman, D., and French, B. 1991. *Lunar Sourcebook: A User's Guide to the Moon* (New York: Cambridge Univ. Press).

Ho, W. W. 1988. High temperature dielectric properties of polycrystalline ceramics. In *Microwave Processing of Materials*, vol. 124, eds. W. H. Sutton, M. H. Brooks and I. J. Chabinsky (Pittsburgh: Materials Research Soc.), pp. 137–148.

Kingery, W. D., Bowen, H. K., and Uhlmann, D. R. 1976. *Introduction to Ceramics* (New York: J. Wiley & Sons), pp. 469–513.

Kopecky, L., and Voldan, J. 1965. *The Cast Basalt Industry*, Annals of the New York Academy of Sciences, vol. 123, pp. 1086–1105.

Langseth, M. G., Keihm, S. J., and Peters, K. 1976. Revised lunar heat flow values. *Proc. Lunar Sci. Conf.* 7:3143–3177.

McKay, D. S., and Ming, D. W. 1989. Mineralogical and chemical properties of the lunar regolith. In *Lunar Base Agriculture: Soils for Plant Growth*, eds. D. W. Ming and D. L. Henninger (Madison, Wis.: American Soc. of Agronomy), pp. 45–68.

Meek, T. T., Vaniman, D. T., Cocks, F. H., and Wright, R. A. 1985. Microwave processing of lunar materials: Potential applications. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 479–486.

Meek, T. T., Fayerweather, L. A., Godbole, M. J., Vaniman, D. T., and Honnell, R. 1988. Sintering lunar simulants using 2.45 GHz radiation. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers). pp. 102–110.

Michalske, T. A., and Freiman, S. W. 1983. A molecular mechanism for stress corrosion in vitreous silica. *J. Amer. Ceramic Soc.* 66:284–288.

Mitchell, J. D., Houston, W. N., Scott, R. F., Costes, N. C., Carrier, W. D. III, and Bromwell, L. G. 1972. Mechanical properties of lunar soil: Density, porosity,

cohesion and angle of internal friction. *Proc. Lunar Sci. Conf.* 3:3235–3253.

Papike, J. J., Simon, S. B., and Laul, J. C. 1982. The lunar regolith: Chemistry, mineralogy, and petrology. *Rev. Geophys. Space Phys.* 20:761–826.

Reed, J. S. 1988. *Introduction to the Principles of Ceramic Processing* (New York: J. Wiley & Sons).

Shirley, F., Buoni, C. M., Lennon, J. W., Mezey, E. J., and Weller, A. E. 1989. A Preliminary Design Concept For A Lunar Sintered Regolith Production Facility (Columbus, Oh.: Battelle). P. O. No. 0200117854 for Lockheed Engineering and Sciences Co.

Simonds, C. H. 1973. Sintering and hot pressing of Fra Mauro composition glass and the lithification of lunar breccias. *Amer. J. Sci.* 273:428–439.

Steurer, W. 1982. Extraterrestrial Materials Processing. First Annual Report on RTOP 179-29-62-01. JPL Publ. 82-41.

Taylor, L. A. 1990. Rocks and minerals in the regolith of the Moon: Resources for a lunar base. In *Advanced Materials: Applications of Mining and Metallurgical Processing Principles*, ed. V. I. Lakshmanan (Littleton, Co.: Soc. for Mining, Metallurgy, and Exploration), pp. 29–47.

Tucker, D. S., Vaniman, D. T., Anderson, J. L., Clinard, F. W. Jr., Feber, F. C. Jr., Frost, H. M., Meek, T. T., and Wallace, T. C. 1985. Hydrogen recovery from extraterrestrial materials using microwave energy. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 583–590.

Uhlmann, D. R. 1972. A kinetic treatment of glass formation. *J. Non-Cryst. Solids* 7:337-348.

Uhlmann, D. R. 1983. Glass formation, a contemporary view. *J. Amer. Ceramic Soc.* 66:95–100.

Uhlmann, D. R., and Onorato, P. I. K. 1979. A simplified model for glass formation. *Proc. Lunar Planet. Sci. Conf.* 10:375–381.

Uhlmann, D. R., Klein, L., Onorato, P. I. K., and Hopper, R. W. 1975. The formation of lunar breccias: Sintering and crystallization kinetics. *Proc. Lunar Sci. Conf.* 6:693–705.

Weiblen, P. W., and Gordon, K. L. 1988. Characteristics of a simulant for lunar surface materials. In *Papers Presented to the Symp. on Lunar Bases and Space Activities in the 21st Century*, April 5–7, Houston, Tex., p. 254 (abstract).

Weiblen, P. W., Murawa, M. J., and Reid, K. J. 1990. Preparation of simulants for lunar surface materials. In *Engineering, Construction, and Operations in Space II: Proc. Space 90*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 98–106.

Williams, R. J. 1985. Oxygen extraction from lunar materials: An experimental test of an ilmenite reduction process. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 551–558.

Williams, R. J., and Jadwick, J. J. 1980. *Handbook of Lunar Materials*, NASA Ref. Publ. 1057.

REFRACTORY MATERIALS FROM LUNAR RESOURCES

W. HOWARD POISL and BRIAN D. FABES
University of Arizona

Any lunar outpost will generate a strong demand for refractory materials. We have studied the feasibility of using lunar resources to produce two classes of refractories—refractory bricks and aerobrake heat shields—representing two extremes of compositional and processing complexity. We conclude that refractory bricks for low temperature applications can be produced on the Moon, though not every step of the process is clear, e.g., the amount, type and source of binders. More advanced refractory materials, such as aerobrake heat shields, will initially have to be transported from Earth. However, as the lunar outpost grows and novel processing techniques are developed, even complicated, high temperature refractory structures could be produced on the Moon.

I. INTRODUCTION

Refractories are materials that are used to contain solids, liquids and gases at high temperature. On Earth refractories are part of all high temperature manufacturing processes. They are used in the iron and steel, nonferrous metal, glass and cement industries to confine melts, to form molds, and to provide structural support for materials in contact with hot or molten material. Refractories are also used in more advanced applications in the electronic and aerospace industries, where they serve such diverse functions as substrates for high-performance computer chips and insulating tiles for the Space Shuttle. With such a wide range of applications, it is not surprising that refractories are made using a variety of both compositions and processes. Oxides, especially those of silicon, aluminum, calcium and magnesium, are by far the most widely used class of materials for refractories. And as the Moon consists principally of these oxides, it seems natural to investigate the feasibility of producing refractory materials from lunar oxides.

Any lunar outpost will generate a strong demand for refractories. Components in high-temperature manufacturing processes, heat exchangers, solar furnaces, nuclear reactors, spacecraft launch and landing pads, aerobrake heat shields, and insulation for human dwellings will all require some refractory components. Whether it is more attractive to manufacture refractories on the Moon, or to bring them to the Moon from Earth will depend on the application.

The concept of using lunar derived ceramics on the Moon is not new. Lunar ceramics have been suggested for structural (Phinney et al. 1977; Simonds 1988; Waldron et al. 1979) as well as refractory uses (Khalili 1985; Mackenzie and Claridge 1979; Shirley et al. 1989). Most of these papers have noted the usefulness of ceramics such as SiO_2, Al_2O_3 and MgO for applications on the Moon. However, in order to obtain these ceramics extensive beneficiation and processing of the lunar regolith would have to occur. Also, little attention has been given to the processing requirements that will be necessary to produce useful ceramic articles on the lunar surface. Exceptions to this are the papers by Shirley et al. (1989) and Mackenzie and Claridge (1979). These authors have identified some of the issues concerning the general processing of ceramic materials on the lunar surface, such as grain size distributions, compositions and binders. These issues will be addressed further in this chapter with respect to two classes of refractories—refractory bricks and aerobrake heat shields—which represent two extremes of compositional and processing complexity.

In each of the above cases we will examine first whether or not the components necessary to make the product are available on the Moon. We have attempted to consider only those minerals which are present in large quantities at various sites on the lunar surface, or which could, in our opinion, be obtainable easily from beneficiation of the lunar regolith. We have not considered the mining or beneficiation techniques that would have to be used on the lunar surface, as this has been addressed by other authors (Burt 1989; Chapter by Chamberlain et al.).

Second, we will look at what processes are available, or are likely to be developed easily, on the Moon. We will also examine some special advantages that the lunar environment might provide for developing new fabrication technologies. Not surprisingly, we will see that most of the necessary materials and processing technology already exist for producing refractory bricks on the Moon, while the prospects for using lunar resources to produce aerobrake heat shields seem remote at present. Finally, we will propose that the development of a process to make ceramic foams from molten regolith is likely to be useful in a variety of applications.

II. REFRACTORY (FURNACE) BRICK

A. Compositions

Terrestrial refractory bricks are made from a variety of compositions. Chrome ores [$(Mg,Fe)(Al,Cr)_2O_4$], periclase (MgO), calcined dolomite (CaO-MgO), olivine [$(Mg,Fe)_2SiO_4$] and mixtures of these account for most common brick (Kingery et al. 1976). For less demanding (lower temperature) applications, the composition is not critical. As a result, the composition of refractory brick depends on the geographic region from which the brick is made. In general, iron, alkali, and alkaline earth impurities lower the use temperature by lowering the melting point of the brick. Sensitivity to these impurities,

therefore, depends on the final use temperature; higher concentrations of impurities can be tolerated for lower use temperatures.

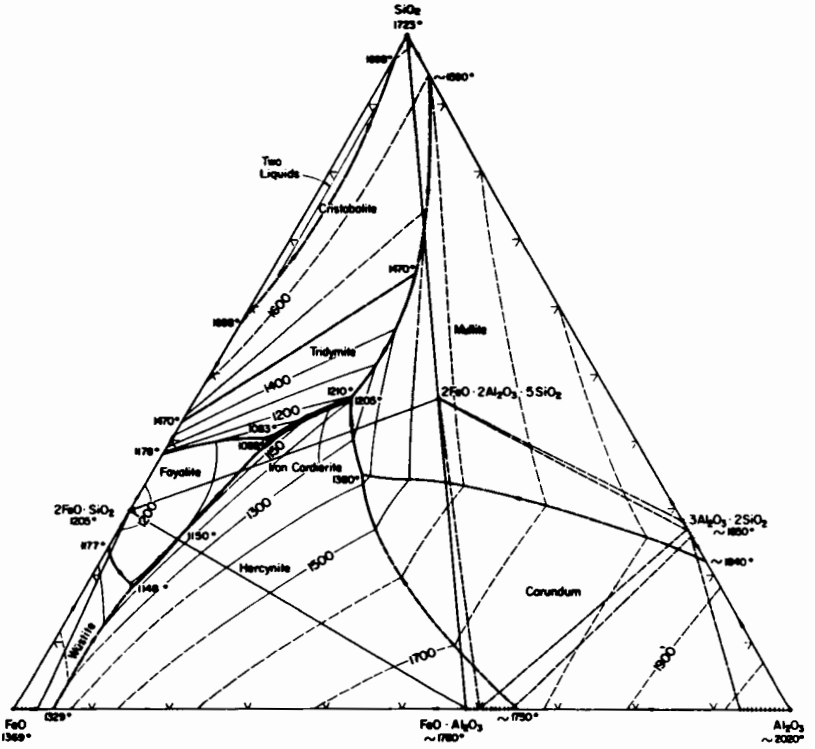


Figure 1. The FeO-SiO_2-Al_2O_3 phase diagram. Notice that the addition of FeO to a SiO_2-Al_2O_3 mixture causes a decrease in the liquidus, seen as a decreasing temperature from right to left (figure from Levin et al. 1964).

It is likely that the composition of bricks for low performance applications on the Moon will depend, as it does on Earth, on the region from which the brick is made. The average composition of the lunar regolith is high in both silica (SiO_2, melting point 1723°C) and alumina (Al_2O_3, melting point 2020°C), upon which many high temperature refractories are based. Much of the regolith contains a significant quantity (5–16 wt%) of iron oxide (FeO), however, which significantly decreases the melting temperature of a crystalline ceramic. Fayalite Fe_2SiO_4, for example, melts at 1205°C, and a fayalite/wustite (FeO)/hercynite ($FeAl_2O_4$) eutectic occurs at 1148°C (Fig. 1). In addition, it is common for refractory brick to contain a residual glassy phase at the grain boundaries. On the Moon the softening temperature of this glass phase will be reduced by additions of FeO, which will define the upper use temperature of the brick (Fabes and Poisl 1991). The magnitude of this effect

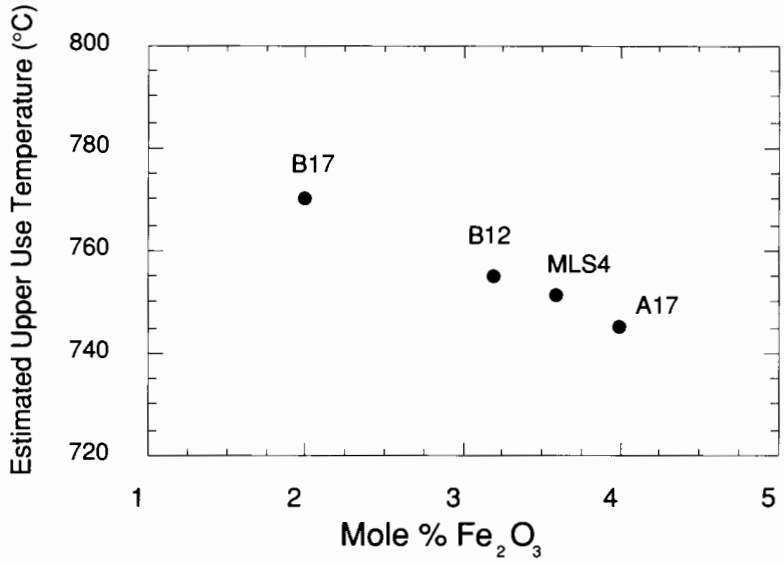


Figure 2. Effect of iron oxide content on the upper use temperature (as defined by the temperature at which the viscosity drops to 10^{11} poise) of simulated lunar materials. See Table I for compositions (figure from Fabes and Poisl 1991).

is shown in Fig. 2, where the change in use temperature (which is defined here by the temperature at which the viscosity drops to 10^{11} poise) is shown for four simulated lunar glasses with varying contents of Fe_2O_3. [Iron, of course, is present in a reduced (2+) state on the Moon. This is likely to alter the detail, but not the trends, shown in Fig 2.]

The compositions of the simulants are shown in Table I. As a guide, refractory bricks made from unprocessed regolith are expected to have a maximum use temperature of no more than 1000°C.

For many applications, an upper use temperature of around 1000°C is quite adequate, and relatively unprocessed regolith should work well. For more demanding applications, however, it will be necessary to perform some sort of beneficiation of the regolith. Olivine [$(Mg,Fe)_2SiO_4$] is normally present in the regolith at concentrations ranging from only 2 to 10 vol% (Papike et al. 1991). However, it has been suggested that some highland impact crater ejecta have associated with them even higher olivine concentrations (Pieters 1989). Beneficiation of this soil would allow bricks based on olivine to be made. As shown in Fig. 3, a refractory brick with fayalite (Fe_2SiO_4), one end-member in the olivine system, as the lowest-temperature phase should be stable to nearly 1200°C. As the fraction of forsterite (Mg_2SiO_4), the other end-member in the olivine system, is increased (Fig. 4) the amount of liquid at temperatures above 1200°C decreases, so that bricks become stable to higher temperatures.

Pyroxenes [$(Ca,Mg,Fe)_2Si_2O_6$] are present in the lunar regolith at con-

TABLE I
Composition of the Simulants in Figure 2 (in wt%)[a]

	A17	B12	B17	MLS-4
SiO_2	46.8	50.2	49.4	51.8
Al_2O_3	14.7	14.6	15.5	14.0
CaO	12.6	11.4	12.3	12.9
MgO	11.0	10.5	11.6	7.3
TiO_2	5.0	5.3	5.3	5.0
Fe_2O_3	9.9	8.0	5.0	8.9
Total	100.0	100.1	99.1	99.9

[a] Analysis based on starting composition of reagent grade chemicals.

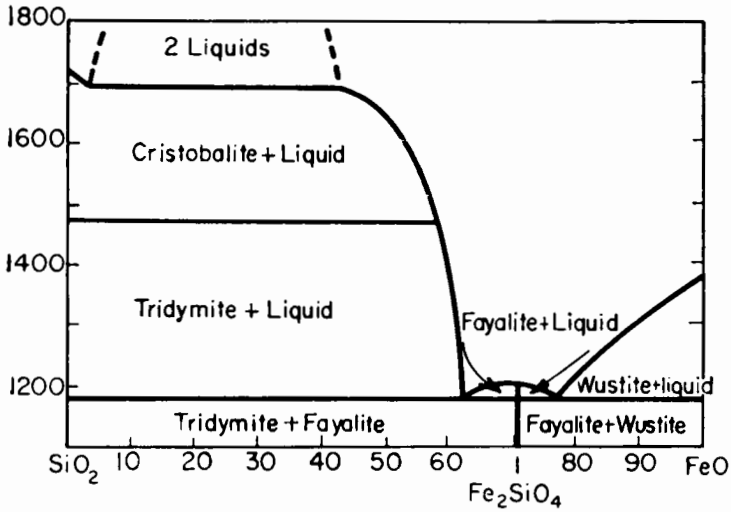


Figure 3. The FeO-SiO_2 phase diagram, showing the effect of adding FeO to SiO_2 (figure from Levin et al. 1964).

centrations ranging from approximately 10 to 60 vol% percent (Papike et al. 1991). Beneficiation of pyroxenes low in calcium and iron from the regolith could be used to prepare bricks which would be stable to about 1300°C.

For applications demanding stability in the 1500°C range, anorthitic rock ($CaAl_2Si_2O_8$) could be used. This mineral is present in large quantities in relatively high purity at some highland sites throughout the Moon (Vaniman

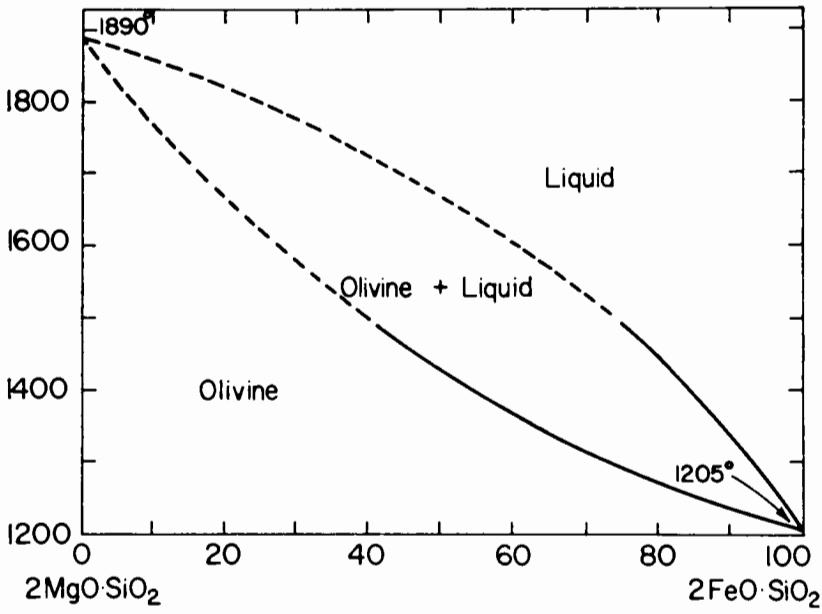


Figure 4. The Mg_2SiO_4-Fe_2SiO_4 phase diagram. Note the entire region of solid solution and the increasing solidus temperature with increasing forsterite (Mg_2SiO_4) content (figure from Levin et al. 1964).

and Heiken 1989). It has a melting temperature of over 1550°C, and as long as there is no excess silica (which forms a glassy grain-boundary phase that melts at lower temperatures), small amounts of alkali, alkaline earth and iron impurities can be tolerated in the anorthite structure.

Finally, for extremely high-temperature applications, fused alumina brick could be made. While this would certainly require significant, energy intensive beneficiation of the soil, it is only for the most demanding applications (use temperatures greater than 1500°C) that such extensive beneficiation would be necessary.

B. Processing

Generally speaking, refractory bricks provide heat containment because of low thermal conductivity k and low heat capacity C_p. Interestingly, both k and C_p are relatively insensitive (on a molar basis) to composition, at least for the range of oxide ceramics that could be obtained easily on the Moon. Refractory bricks contain a large volume fraction of porosity (almost always greater than 25 vol%), however, which decreases both k and C_p on a volume basis. It is, therefore, the amount and type of porosity that, to a large extent, define the insulating properties of different refractory bricks. Hence, in addition to choosing the proper composition, it is important to process refractory bricks so that the proper amount and distribution of porosity are present.

The grain-size distribution of the particles used to produce refractory brick has an effect on the porosity of the brick. Various combinations of coarse, medium and fine-grained particles can result in variations of porosity from 22 to 40 vol% (Chesters 1973). This grading of particle sizes is usually done to produce a brick with low porosity. For example, mixing 45 wt% of about 0.13 mm particles, 10% of about 0.2 mm particles, and 45% of particles about 2 mm results in a porosity of about 23%. Higher porosities can be obtained by using a more highly monosized distribution of particles or by the addition of organic materials which burn out during firing.

The grain-size distribution of the lunar regolith at several Apollo landing sites is shown in Fig. 5 (Mitchell et al. 1972). This figure illustrates that the regolith contains a number of particles with sizes greater than 3 mm. Such particles would likely need to be removed from the lunar regolith by screening or sieving. Finer particle separation would probably not be necessary, although this will depend somewhat on the compositions of the different-sized particles.

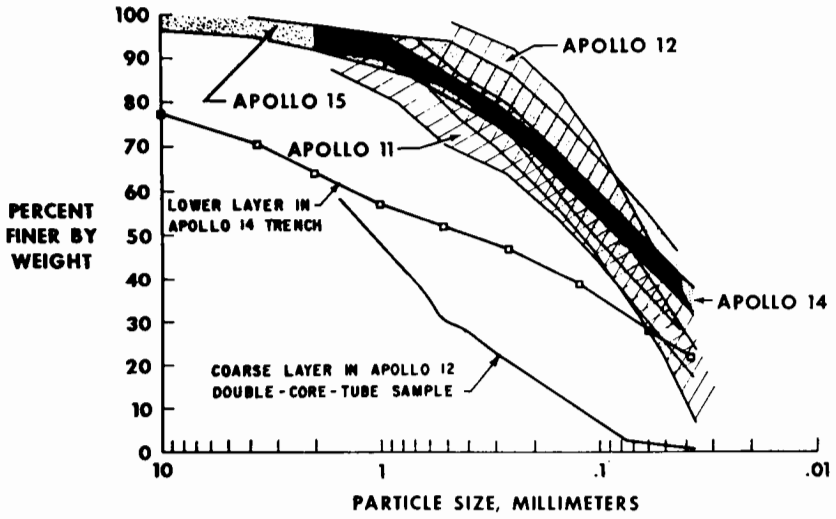


Figure 5. Grain size distribution ranges (less than 1 cm) of samples from various Apollo sites (figure from Mitchell et al. 1972).

On Earth, refractory bricks are made by mixing oxide particles with 5 to 15 wt% of a binder—clay and water or organic materials, such as glycerol, stearic acid or epoxy resins (Gilchrist 1977; Shaw 1972). The binder serves two purposes. First, it gives the mixture of oxide particles a small amount of strength (known as green strength) for molding and handling purposes. Second, during firing, the binder burns out, leaving the porosity in the volume that it or its combustion products had filled.

Unfortunately, neither water and clay mixtures nor organic binders are

available on the Moon (Shirley et al. 1989). They are not required in large amounts though, and a single 100 kg shipment of an organic resin, added at 5 wt%, could be used to produce approximately 3,000 bricks measuring $5 \times 10 \times 20$ cm. It is likely that judicious use of binder will allow bricks to be made with less than the 5 wt% which is commonly used on Earth. It is also likely that a process for recovering binder as it volatilizes could be developed on the Moon. Such a process should further decrease the amount of binder that would need to be transported from Earth.

D. T. Vaniman (personal communication) has suggested that sulfur could be used as a binder for making refractory bricks. Sulfur is present on the Moon in the form of sulphide minerals, predominantly troilite (FeS), or surface correlated sulfur, and can be extracted by heating of the regolith (Vaniman et al. 1992). Pure sulfur is a fluid liquid over a broad range of temperatures (112–160°C) and could be used to produce bricks with sufficient green strength for processing. However, the process of removing sulfur during firing is not known. Neither are the effects of any residual sulfur on the chemical or mechanical properties of the brick. The use of sulfur as a binder for refractories is an area which should be investigated further.

The lunar environment also provides some unique processing capabilities, two of which might obviate the need for binders. First, by melting the refractory in a closed vessel, and then opening up the container to the natural vacuum of the lunar environment, it might be possible to foam and quench a melt simultaneously, which would provide a highly porous, insulating material. Alternatively, it might be possible to fill molds with unprocessed regolith, which could then be heated rapidly (using either solar or conventional, resistive heating) to sinter only partially (Shirley et al. 1989), leaving some porosity to provide for low k and C_p. Neither of these approaches has been investigated extensively on Earth, principally because binders are so readily available.

III. AEROBRAKE HEAT SHIELDS

With respect to compositional and processing complexity, aerobrake heat shields would seem to be about as far removed from refractory brick as one can get. Nonetheless, the purpose of the heat shield is essentially the same as refractory brick: to confine heat and protect underlying materials. Two differences between furnace brick and aerobrake heat shields, though, stand out. First, aerobrake heat shields must tolerate mechanical, as well as thermal loads. And second, the tolerance for failure is quite low for the aerobrake. A spilled pool of glass is an extreme nuisance, but a burned-up spacecraft can be tragic.

Aerobrakes, as envisioned in the form of structures with heat shields deployable from the Moon or from an easily accessible orbit, are currently under development (Klang and Washington 1990). Although such systems are not currently made, even on Earth, the thermal protection system (TPS)

of the Space Shuttle provides a convenient terrestrial example from which we can infer information on the feasibility of processing the heat-shielding components of aerobrakes from lunar resources.

A. Space Shuttle Thermal Protection System (TPS)

As drawn schematically in Fig. 6, the Space Shuttle TPS can be broken down into three main components: ceramic tiles, a nomex[a] felt strain isolation pad, and silicone impregnated nomex filler bar spacers. The ceramic tiles make up the heart of the TPS. They come in two types: high temperature reusable surface insulation (HRSI, used for temperatures between 650 and 1350°C) and low-temperature reusable surface insulation (LRSI, used for temperatures between 370 and 650°C) (Korb et al. 1981). Both are made of high-purity (99.62%) amorphous silica fibers, and in some cases aluminoborosilicate fibers, fused together to form tiles. On their external surfaces is a borosilicate coating, which provides a smooth, aerodynamic surface during landing. The main difference between the black (HRSI) and white (LRSI) tiles is the use of SiB_4, which is added to the HRSI tiles to provide high emissivity during re-entry. The LRSI tiles, on the other hand, are designed for thermal control during orbit.

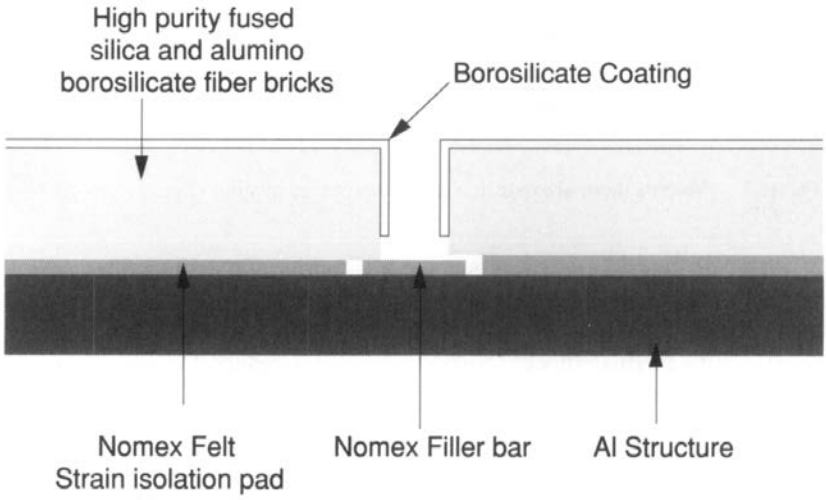


Figure 6. Schematic of the Space Shuttle thermal protection system, showing ceramic tiles with coating, strain isolation pads, and filler bars (figure after Thuss et al. 1971).

To produce a tile, fibers are slurry cast, pressed and sintered at approximately 1370°C (Leiser et al. 1981). Mechanical strength is provided by the

[a] Poly-m-phenylene isophthalamide, manufactured by E. I. DuPont deNemours Co., Wilmington, Del.

fiber-fiber bonding which occurs at this temperature. The resulting tile is as much as 93% porous and has low thermal expansion and low thermal conductivity, resulting in a tile which has good thermal shock properties. With the exception of a binder and the slurry, it should be possible to transfer this firing process to the Moon. The lack of water in the lunar environment, on the other hand, is likely to increase the adhesion of fibers due to the absence of adsorbed surface contaminants and may obviate the need for binders (Shirley et al. 1989).

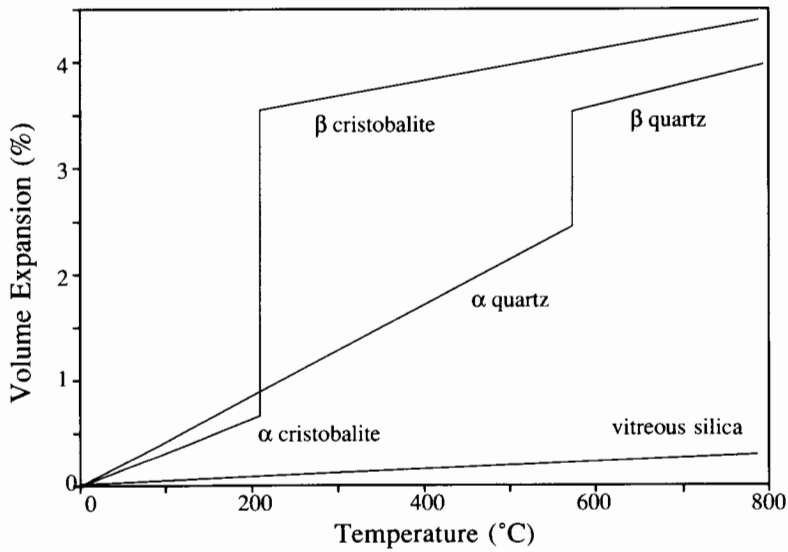


Figure 7. Volume thermal expansion of various phases of silica (figure after Gilchrist 1977).

The key to the tiles' success is the high purity of the silica fibers. High purity is important for a number of reasons. First, high-purity silica has an extremely low coefficient of thermal expansion, which provides superior resistance to thermal shock. Second, impurities reduce the maximum use temperature of silica by decreasing the softening point of the glass. And third, impurities cause the glass to crystallize (Ainslie et al. 1962; Thuss et al. 1971), which can have devastating consequences on the mechanical properties of the tiles. As shown in Fig. 7, crystalline silica undergoes displacive phase changes from β quartz to α quartz at 573°C and from β cristobalite to α cristobalite at 210°C (Gilchrist 1977). A large change in volume is associated with these transformations, which can lead to cracking and fracture of the tiles. In fact, if the tiles crystallize at high temperatures, they break apart into powder when they are cooled (H. Goldstein, personal communication).

Amorphous silica, on the other hand, contracts smoothly as it is cooled and heated from high temperature to room temperature and back. The devitrification (crystallization) rate of amorphous silica is a complex function

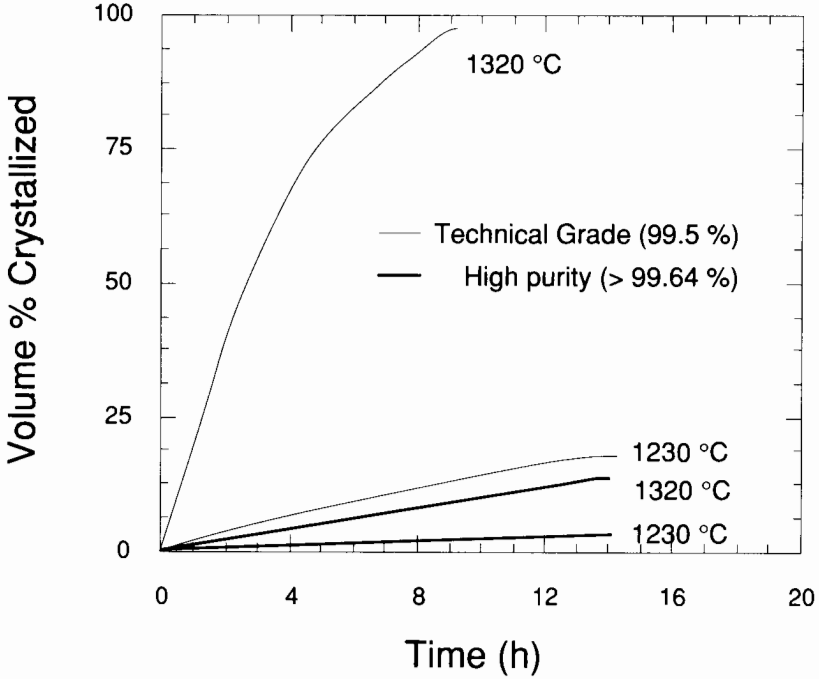


Figure 8. Volume percent crystallized as a function of time for high purity and technical grade amorphous silica at various temperatures. Note the large effect of impurities on devitrification of silica (figure after Thuss et al. 1971).

of temperature, impurities and surface area. Figure 8 shows the marked effect of purity on the growth rate of cristobalite in high-purity amorphous silica (approximately 99.9% pure) and in technical grade amorphous silica of lower-purity (approximately 99.5% pure). Notice that at all temperatures the lower-purity material (which is still 99.5% pure) has a much greater rate of devitrification. This low-purity material, therefore, tends to crack and disintegrate when it is cycled from high to low temperatures and back.

B. Lunar Processing of Aerobrake Heat Shields

The raw materials needed to process aerobrake heat shields are present on the Moon. Silica, for example, is abundant, although at nowhere near the purity (>99.5%) that is required. Because the mass of lunar-processed aerobrake tiles will not be as critical as the mass of terrestrial-processed tiles, due to the lower lunar gravity well, there is little reason to worry about optimizing the density of the lunar-derived materials. Oxides with slightly higher densities (e.g., mullite, alumina and anorthite) should be considered. Mullite, alumina and anorthite, for example, all have high thermal stability and low thermal conductivities. Relatively high purity of all of these materials would be

desired, because impurities reduce the temperature at which these materials can be used. For example, as shown in Fig. 9, the addition of several percent of anorthite to alumina will produce approximately 10% liquid at 1600°C. This amount of liquid is certain to ruin the mechanical integrity of the refractory. Figure 10 shows that the addition of several percent silica to anorthite will produce approximately 10% liquid at 1400°C. Separating anorthite from the lunar regolith at purity levels greater than 95% would therefore be necessary. Even though such beneficiation would require a large expense in energy, it is probably more achievable than the greater than 99.5% purity needed for silica.

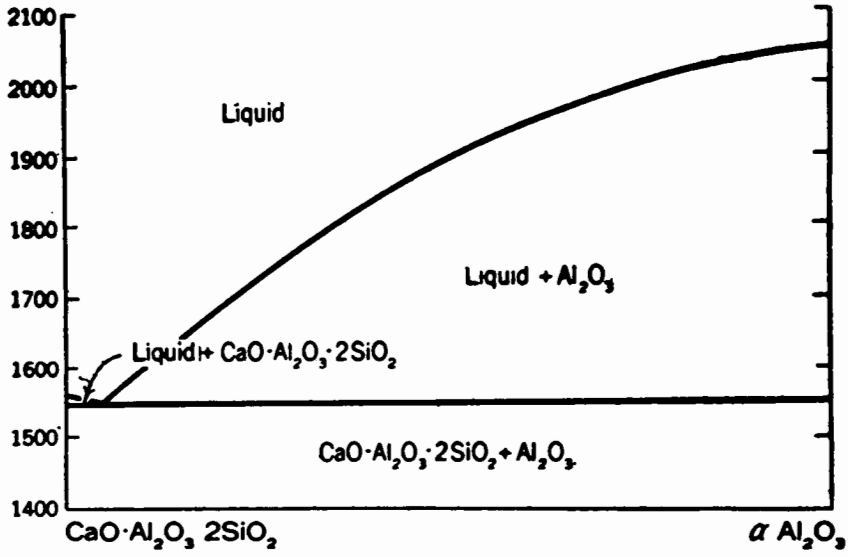


Figure 9. The Al_2O_3-$CaAl_2Si_2O_8$ phase diagram. Note the initial decrease in liquidus as Al_2O_3 is added to $CaAl_2Si_2O_8$ (figure from Levin et al. 1964).

Assuming that the appropriate purity can be obtained, fibers might be spun, cast and then sintered into the desired shape. Mackenzie and Claridge (1979) were able to spin glass fibers from melt compositions corresponding to bulk regolith. Fiber spinning is a relatively complicated process, however, and it is likely that a large effort would be needed to tailor the process for both the lunar environment and for the compositions being spun. Alternatively, a foamed ceramic with low C_p and k might be produced by melting regolith of the proper composition in a containment vessel and then exposing this melt to the lunar vacuum, as discussed previously for refractory brick.

C. Coatings

The borosilicate coating applied to the nominally 6″ × 6″ square tiles provides a high emittance surface as well as handling capability (the ability to pick

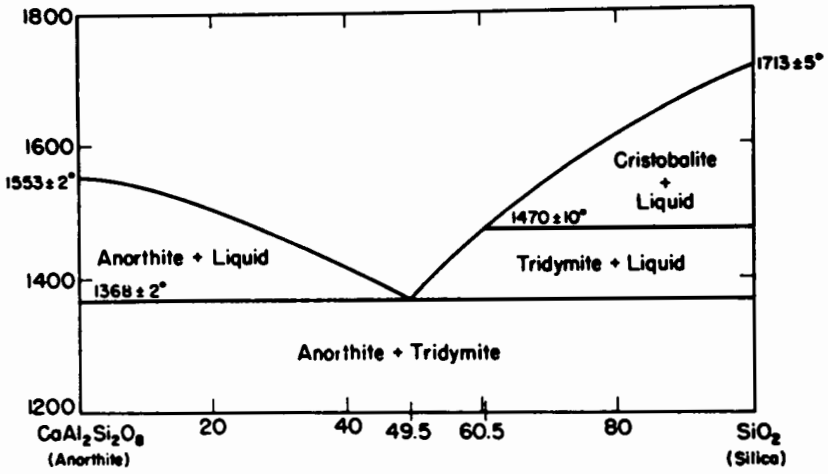


Figure 10. The SiO_2-$CaAl_2Si_2O_8$ phase diagram. Note the initial decrease in liquidus as SiO_2 is added to $CaAl_2Si_2O_8$ (figure from Levin et al. 1964).

up without crushing) and moisture-resistance (sealing of pores from water adsorption). The coating is applied to the external surface and partway down the four sides of each tile in thicknesses ranging from 0.23 to 0.38 mm, and is designed to be soft at its operating temperature. This provides the tiles with a smooth surface during re-entry, which prevents localized overheating due to disruptions of the laminar flow region over the surface.

The borosilicate glass used for shuttle coatings would be difficult to reproduce on the Moon, because there is very little boron in the lunar regolith. The mass of the coating material, though, is not likely to be large. At a thickness of 0.3 mm, 100 kg of glass should be able to coat approximately 140 m^2 of tile. Hence, it is not unreasonable to anticipate that the coating material could be fabricated on Earth and transported to the Moon.

IV. SUMMARY AND CRITICAL ISSUES

Any lunar outpost will provide a demand for a range of refractory materials and it seems that a number of these could be made using native resources, especially for the less demanding (lower-temperature) applications, such as refractory brick. Even for these applications, though, not every step of the process is clear. For example, the choice and amount of a binder need to be determined, including the use of sulfur, as well as the possibilities provided by binderless processing using solar heating of refractories in molds or foaming of ceramic melts.

For more complicated applications, such as aerobrake heat shields, issues in compositional purity as well as processing complexity become more important. Even if minerals with the requisite purity can be obtained, the fab-

rication technology (fiber spinning) will be complicated. Here, again, there are opportunities for novel processing, such as foaming of high-purity-melts, to be used.

Finally, as more information concerning the mineralogy, stratigraphy and properties of the lunar crust becomes available, a better understanding of the beneficiation required of the lunar regolith and the processing techniques necessary to produce a range of refractory materials will be obtained. We anticipate, therefore, that the use of lunar resources for refractory applications will evolve slowly—from an initial stage, where only the most simple materials are made on the Moon and most are transported from Earth, to a point where even extremely high temperature, complicated structures, such as aerobrake heat shields can be produced entirely on the Moon.

Acknowledgments. H. Goldstein and D. Lynch provided critical reviews of the manuscript along with many helpful suggestions. The authors are supported by the NASA Space Engineering Research Center for the Utilization of Local Planetary Resources at the University of Arizona.

REFERENCES

Ainslie, N. G., Morelock, C., and Turnbull, D. 1962. Devitrification kinetics of fused silica. In *Symp. on Nucleation and Crystallization of Glasses and Melts* (Columbus, Oh.: American Ceramic Soc.), pp. 97–108.

Burt, D. M. 1989. Mining the Moon. *American Scientist* 77:574–579.

Chesters, J. H. 1973. *Refractories: Production and Properties* (London: Iron and Steel Inst.).

Fabes, B. D., and Poisl, W. H. 1991. Glass-ceramics from lunar resources. In *Space Manufacturing 8, Proc. of the Tenth Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 352–357.

Gilchrist, J. D. 1977. *Fuels, Furnaces and Refractories* (Oxford: Pergamon Press).

Khalili, E. N. 1985. Magma, ceramic, and fused adobe structures generated in situ. In *Lunar Bases and Space Activities of the 21st Century,* ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 399–403.

Kingery, W. D., Bowen, H. K., and Uhlmann, D. R. 1976. *Introduction to Ceramics,* 2nd ed. (New York: Wiley & Sons).

Klang, E., and Washington, G. 1990. Aerobrake construction concepts for the Mars mission. In *Engineering, Construction, and Operations in Space II: Proc. Space 90,* eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Chemical Engineers), pp. 738–747.

Korb, L. J., Morant, C. A., Calland, R. M., and Thatcher, C. S. 1981. The Shuttle Orbiter Thermal Protection System. *Ceramic Bull.* 60:1188–1193.

Leiser, D. B., Smith, M., and Goldstein, H. E. 1981. Developments in fibrous refractory composite insulation. *Ceramic Bull.* 60:1201–1204.

Levin, E., Robbins, C., and McMurdie, H. 1964. *Phase Diagrams for Ceramics* (Columbus, Oh.: American Ceramic Soc.).

Mackenzie, J. D., and Claridge, R. C. 1979. Glass and ceramics from lunar materials. In *Space Manufacturing Facilities III*, eds. J. Grey and C. Krop (New York: AIAA), pp. 135–140.

Mitchell, J., Houston, W., Scott, R., Costes, N., Carrier, W., and Bromwell, L. 1972. Mechanical properties of lunar soil: Density, porosity, cohesion, and angle of internal friction. *Proc. Lunar Sci. Conf.* 3:3235–3253.

Papike, J., Taylor, L., and Simon, S. 1991. Lunar minerals. In *Lunar Sourcebook: A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (Cambridge: Cambridge Univ. Press), pp. 121–182.

Phinney, W. C., Criswell, D. R., Drexler, E., and Garmirian, J. 1977. Lunar resources and their utilization. In *Space Manufacturing Facilities II*, ed. J. Grey (New York: AIAA), pp. 171–180.

Pieters, C. M. 1989. Compositional stratigraphy of the lunar highland crust. *Lunar Planet. Sci.* XX:848–849 (abstract).

Shaw, K. 1972. *Refractories and Their Uses* (London: Applied Science).

Shirley, F., Buoni, C. M., Lennon, J. W., Mezey, E. J., and Weller, A. E. 1989. A Preliminary Design Concept For A Lunar Sintered Regolith Production Facility (Columbus, Oh.: Battelle). P. O. No. 0200117854 for Lockheed Engineering and Sciences Co.

Simonds, C. H. 1988. Hot pressing of lunar soil and qualification for manned applications. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc. of Civil Engineers), pp. 90–101.

Thuss, R. C., Thibault, H. G., and Hiltz, A. 1971. The utilization of silica based surface insulation for the Space Shuttle Thermal Protection System. In *Space Shuttle Materials*, vol. 3 (New York: Soc. for the Advancement of Material and Process Engineering), pp. 445-456.

Vaniman, D. T., and Heiken, G. H. 1989. The unknown lunar anorthositic regolith. Geological Soc. of America, Annual Meeting, Nov. 6–9, St. Louis, Mo., Abstract book, p. A368.

Vaniman, D. T., Pettit, D., and Heiken, G. 1992. Uses of lunar sulfur. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 1, pp. 429–435.

Waldron, R. D., Erstfeld, T. E., and Criswell, D. R. 1979. Overview of methods for extraterrestrial materials processing. In *Space Manufacturing Facilities III*, eds. J. Grey and C. Krop (New York: AIAA), pp. 113–127.

LUNAR VOLATILES: IMPLICATIONS FOR LUNAR RESOURCE UTILIZATION

BRUCE FEGLEY, JR.
Washington University

and

TIMOTHY D. SWINDLE
University of Arizona

We critically review from a resources perspective the available data on the abundances of the noble gases, hydrogen, carbon, nitrogen, sulfur, fluorine and chlorine in lunar samples. The analytical and mineralogical data relevant to the water content of lunar materials are also reviewed. The factors affecting the abundances of the solar-wind-implanted volatiles are discussed. The extensive analytical data on the Apollo and Luna samples are used to estimate average global inventories of these volatiles in the lunar regolith. The analytical data are also used to discuss some implications of different volatile extraction schemes such as heating and grain size sorting. The chemistry of lunar volcanic gases is discussed with an emphasis on using the chlorofluorocarbon gases as chemical probes of the water abundance in the lunar interior and on the transport of ore-forming metals. Finally, we highlight some of the major unanswered questions about using lunar volatiles as resources and make recommendations for future resource related work on lunar volatiles.

I. INTRODUCTION

This chapter presents a critical review from a resources perspective of the available data on the abundances of the noble gases and several important chemically reactive volatile elements (H, C, N, S, F, Cl) in lunar materials. These data are an essential input to studies of lunar resource utilization, but in the past have been difficult to use because they are scattered throughout the vast literature on studies of the Apollo and Luna samples.

The chapter is organized as follows. In Sec. II some general information about the factors affecting the abundances of the solar-wind-implanted volatiles is presented. In Sec. III the extensive analytical data on lunar samples are compiled and critically assessed. The mean abundances of the different volatiles in lunar soils at the different Apollo landing sites, in breccias, and in basalts are derived from this assessment. Where data are available, the mean volatile abundances at the different Luna landing sites are also derived. The volatile abundances determined by different laboratories on lunar soils,

breccias, and basalts are also tabulated. The mean abundance data are then combined with models of volatile abundance as a function of depth in the lunar regolith to give average global inventories for the volatiles. In Sec. IV, some volatile extraction methods are discussed with an emphasis on thermal extraction and grain size sorting applied to lunar soils. In Sec. V, the isotopic and mineralogical evidence for and against indigenous lunar water is reviewed. In Sec. VI, thermodynamic models of lunar volcanic gases are presented and the implications for the transport of ore-forming metals in lunar volcanic gases are discussed. These models also show that the water abundance in the lunar interior (at the time that volcanism occurred) can potentially be probed by analyzing trapped lunar volcanic gases in lunar samples for chlorofluorocarbon (CFC) gases. Finally, in Sec. VII we summarize some of the major remaining unanswered questions about lunar volatiles and recommend studies to answer these questions.

II. SOLAR-WIND VOLATILES: GENERAL CONSIDERATIONS

The solar wind consists of energetically charged particles that flow radially outward from the solar corona. Essentially, it represents the continuous expansion of the outer layers of the solar atmosphere. The solar wind can be deflected either by a magnetic field or by an atmosphere, but on a body such as the Moon, which has little of either, the solar wind is only stopped by the surface material. The elemental composition of the solar wind is that of the Sun—predominantly H and He, with the abundance of heavier elements generally decreasing with increasing mass. The only elements for which the solar wind makes the largest contribution to lunar inventories are those which are extremely depleted on the Moon: the noble gases and the volatiles H, C and N. Because the abundances of these elements tend to correlate with one another, we will discuss them together. For a more detailed discussion of solar-wind implantation, focusing on the scientific aspects rather than resource evaluation, see Haskin and Warren (1991).

A. Factors Affecting Abundances

When a solar-wind particle hits a rock (or mineral grain) on the Moon's surface, it penetrates less than 1 μm, so only material which has been within 1 μm of the lunar surface will contain implanted solar-wind particles. One result of the shallow implantation depth of the solar wind is that solar-wind-derived volatiles are concentrated in the finest grain sizes, where the surface-to-volume ratio is largest. This is illustrated in Fig. 1 where analytical results for the abundance of ^{3}He in different size fractions in lunar soil 10084 are displayed. Analogous studies of the hydrogen, carbon and nitrogen abundance as a function of grain size in lunar soils have been conducted by several groups (see, e.g., Becker and Clayton 1975,1977; Bustin et al. 1984,1986; DesMarais et al. 1973; Frick et al. 1988; Gibson et al. 1987; Goel and Kothari 1972; Goel et al. 1974*a*, *b*; Kaplan et al. 1970; Müller 1972,1973,1974). These results

generally show that H, C and N are also concentrated in the finer grain size fractions of lunar soils. Other investigations in which chemical etching (e.g., by F_2) or depth sensitive analytical techniques are employed also show that the solar-wind-derived volatiles are concentrated on the outer surfaces of grains (see, e.g., Becker 1980; Epstein and Taylor 1975; Filleux et al. 1977,1978; Goldberg et al. 1975; Leich et al. 1973,1974). We note that there is some disagreement between the results of the different F_2 stripping experiments (Becker 1980; Epstein and Taylor 1975) with regard to the exact amount of hydrogen on grain surfaces and its speciation. Overall, the large body of analytical data on the noble gases and the chemically reactive solar-wind-derived volatiles show that in principle, beneficiation by grain-size sorting is possible. However, as we discuss below, it is not clear whether this beneficiation method will prove to be practical.

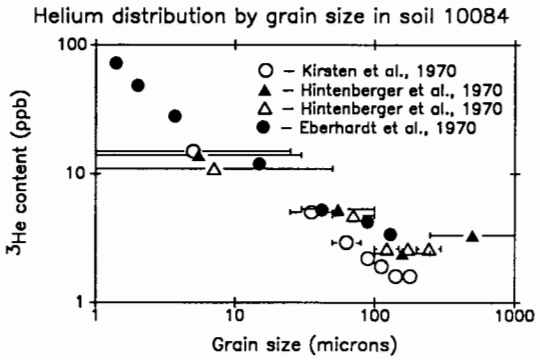


Figure 1. The ^{3}He content of Apollo 11 soil 10084 as a function of grain size. The larger contents are found in the smaller particles which have larger surface to volume ratios (figure modified from Swindle et al. 1990).

Another factor affecting the abundance of solar wind volatiles is the "maturity" of a soil—the amount of natural processing that it has undergone. The abundances of the solar-wind-derived volatiles generally correlate with several other parameters that relate to exposure to either the solar wind or impacting micrometeorites. As surface exposure progresses, volatile abundance increases, the mean grain size of the soil decreases, the abundances of agglutinates (composite particles welded together by impacts) increases, and the ratio of fine-grained reduced iron (presumably produced by impact processes) to oxidized iron increases (see, e.g., Morris 1978). Correlations between the abundances of several solar-wind-derived volatiles with the last parameter, usually denoted by I_s/FeO, are shown in Fig. 2. Techniques which can remotely sense one of these other parameters (see, e.g., Charette et al. 1976) might be useful in evaluating the resource potential of different areas of the Moon. On the other hand, the maturity of the surface layer, which is all that can be detected by remote sensing, might be unrelated to the maturity

of deeper material, in which case knowledge of surface maturity variations would not be important (Taylor 1991).

The integrated solar-wind fluence varies with location on the Moon as a result of two effects. First, the direction of flow of the solar wind is nearly constant and nearly parallel to the plane of the Moon's rotation, so higher latitudes receive less solar wind per unit area than do equatorial latitudes. Second, the Moon spends about one-quarter of its time (near full Moon) in the tail of the Earth's magnetosphere, which deflects the solar wind (Swindle et al. 1992), so the nearside receives less solar wind exposure than does the farside. The combined result is shown in Fig. 3, which indicates that the central nearside receives <30% of the exposure than that of the central farside. No clear pattern of geography-related variation is apparent in the Apollo samples. The absence of an effect in the Apollo samples might simply reflect the fact that these come from a restricted area on the Moon's nearside. On the other hand, a higher fluence would not translate into a higher volatile abundance of a gas if individual grains are saturated with that gas at the fluence received by the Apollo samples (see Futagami et al. [1990], and Wieler et al. [1980] for different viewpoints and references to other discussions). Currently available data on Luna samples, which come from nearer the eastern limb of the nearside, have higher mean concentrations of all five noble gases and N than do Apollo samples, but, as discussed below, the Luna data only represent a few measurements of mature soils.

The chemistry of the lunar soil seems to be a significant factor only in the case of the light noble gases. Helium (and Ne) contents are much higher (by as much as two orders of magnitude) in ilmenite than in co-existing minerals such as olivine, pyroxene and plagioclase, presumably because diffusion is slower in ilmenite than in the other minerals (see, e.g., Basford 1974; Frick et al. 1988; Hübner et al. 1975; Kirsten et al. 1971; Signer et al. 1977). Because most lunar Ti is in ilmenite, the He and Ne abundances tend to correlate with the Ti abundance (Fig. 4). If maturity effects are taken into account, the correlation is even stronger (Jordan 1989) as illustrated in Fig. 5. As the Ti abundance can be determined remotely by either gamma-ray (Davis 1980) or visible-infrared (Pieters 1978; Johnson et al. 1991) spectroscopy, the He abundance can also be mapped, with some assumptions about maturity variations.

Although there are presumably mineral-dependent abundance effects for the heavier noble gases and H, C and N as well, these are apparently much smaller than the effects for He and Ne (see, e.g., Frick et al. 1988). Therefore, the variations from location to location would be expected to be smaller. We took mean measured volatile abundances and major and minor element abundances for more than 60 Apollo soils, and found that, even if the strong correlations with maturity are taken into account (e.g., by calculating residuals from best-fit lines in plots like Fig. 2), there are few statistically significant (90% level) correlations between the heavy noble gases, H, C or N and other elements. The few correlations present were mostly with Ti, although they

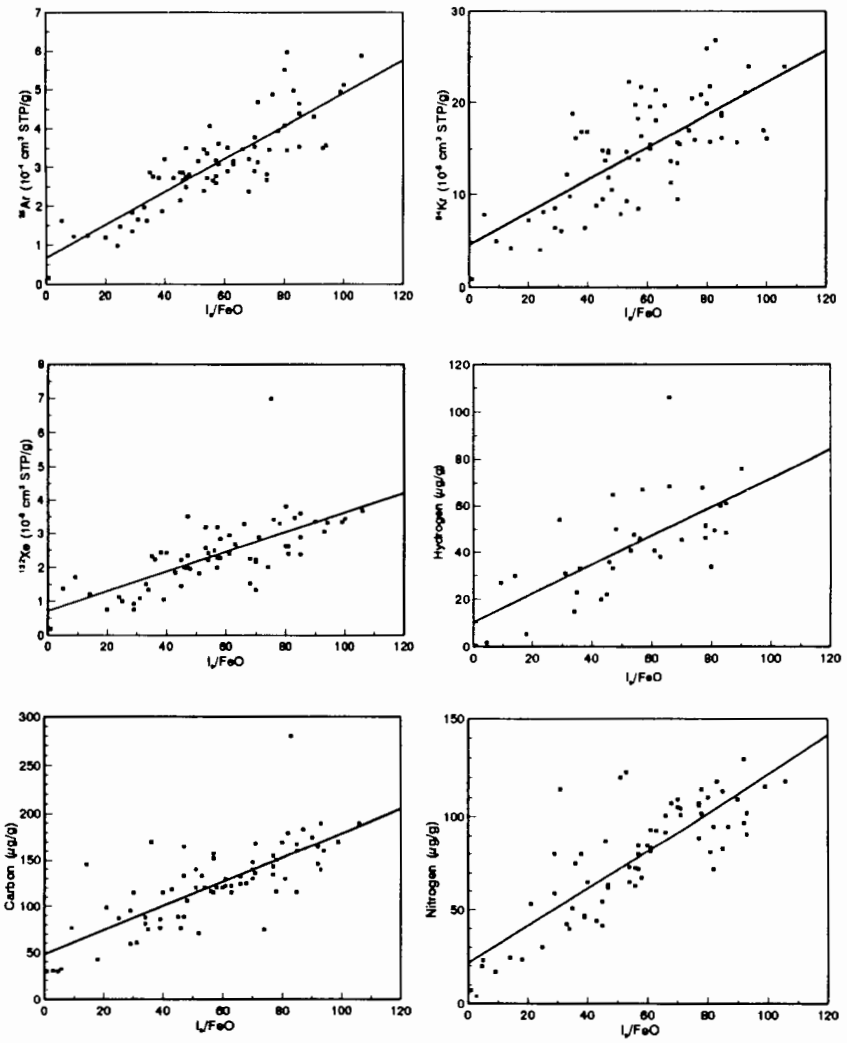


Figure 2. (a) top left: The ^{36}Ar content of lunar soils as a function of maturity as measured by I_s/FeO, the ratio of reduced to oxidized iron. (b) top right: The same for ^{84}Kr. (c) middle left: The same for ^{132}Xe. (d) middle right: The same for hydrogen. (e) bottom left: The same for carbon. (f) bottom right: The same for nitrogen. Data for ^{36}Ar, ^{84}Kr and ^{132}Xe are taken from soil means (references given by Swindle et al. 1990). The data for H, C, and N are taken from literature data compiled by B. Fegley as part of writing this chapter. The I_s/FeO data are generally taken from the compilation by Morris (1978). The lines in the figures are unweighted linear least-squares fits to all of the data points in each figure.

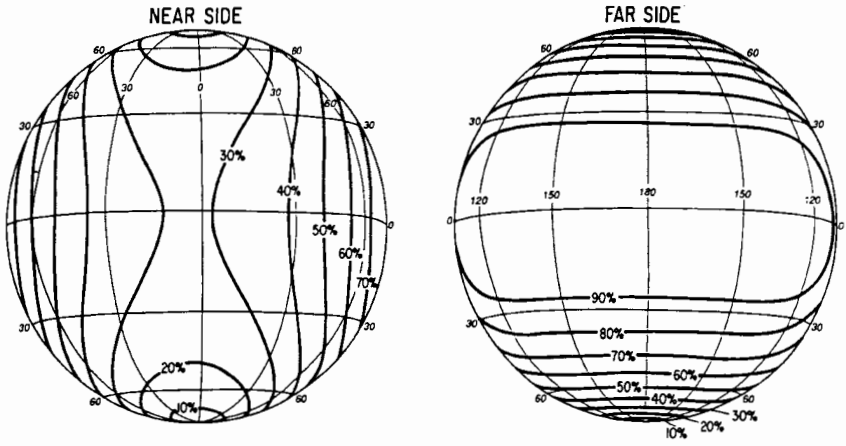


Figure 3. The relative amount of solar-wind exposure for lunar regions as a function of location. The calculations, which are from Swindle et al. (1992), assume that the Moon is completely shielded from the solar wind by the terrestrial magnetotail for 25% of each month, and that the flux is constant otherwise. The contours give percent of maximum (equatorial farside) fluence.

were weaker (and would translate into much smaller abundance variations) than the correlations of these elements with I_s/FeO or the correlations of He and Ne with Ti.

Finally, some regions with distinctive swirl-like albedo patterns could be the result of deflection of solar-wind ions by local magnetic anomalies (Hood and Schubert 1980; Hood and Williams 1989). If so, there would be corresponding regional variations in the abundance of solar-wind-implanted volatiles, which could be easily mapped.

III. ESTIMATED INVENTORIES OF LUNAR VOLATILES

In this section, we use the extensive analytical data on the Apollo and Luna samples to make reasonable estimates of the typical abundances of He, Ne, Ar, Kr, Xe, H, C, N, S, F and Cl in soils collected near the surface of the central nearside. These estimates will then be combined with analytical results (where available) and theoretical models on the depth distribution of these elements in the lunar regolith. As a result, we derive average global inventories for each element considered. We also review the abundances of the chemically reactive volatiles H, C, N, S, F and Cl in lunar basalts and breccias. These data, in combination with the data on lunar soils are then used to discuss the water content of the Moon, compositional models for lunar volcanic gases, and theoretical studies of the transport of metals by lunar volcanic gases.

The literature on analyses of the noble gases, H, C, N, S, F and Cl in lunar samples was reviewed by the two authors with Swindle compiling and

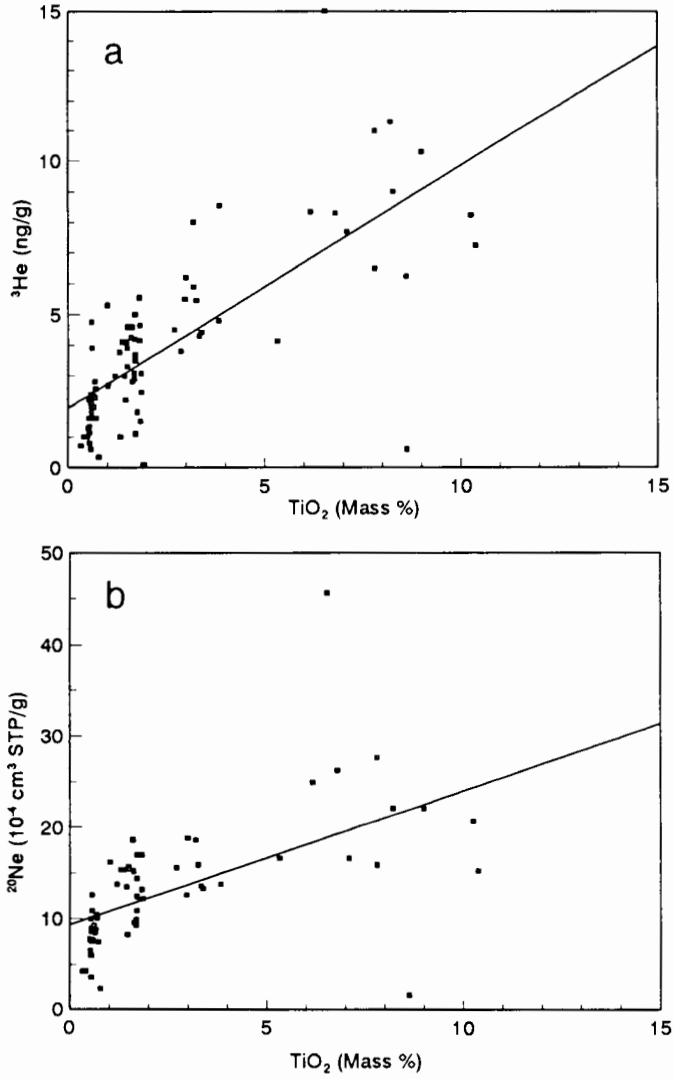


Figure 4. (a) top: A plot of ^{3}He content (ng g^{-1}) vs Ti content (% TiO_2 by mass) in lunar soils. The solid diagonal line is a linear least-squares fit to the values for individual soils. Plot modified from Swindle et al. (1990). (b) bottom: A plot of ^{20}Ne content (cm^3 STP g^{-1}) vs Ti content (% TiO_2 by mass) in lunar soils. The solid diagonal line is a linear least-squares fit to the values for individual soils.

assessing the analytical data for the noble gases and Fegley compiling and assessing the data for the chemically reactive volatiles. The results of this critical assessment are described below starting with the noble gases.

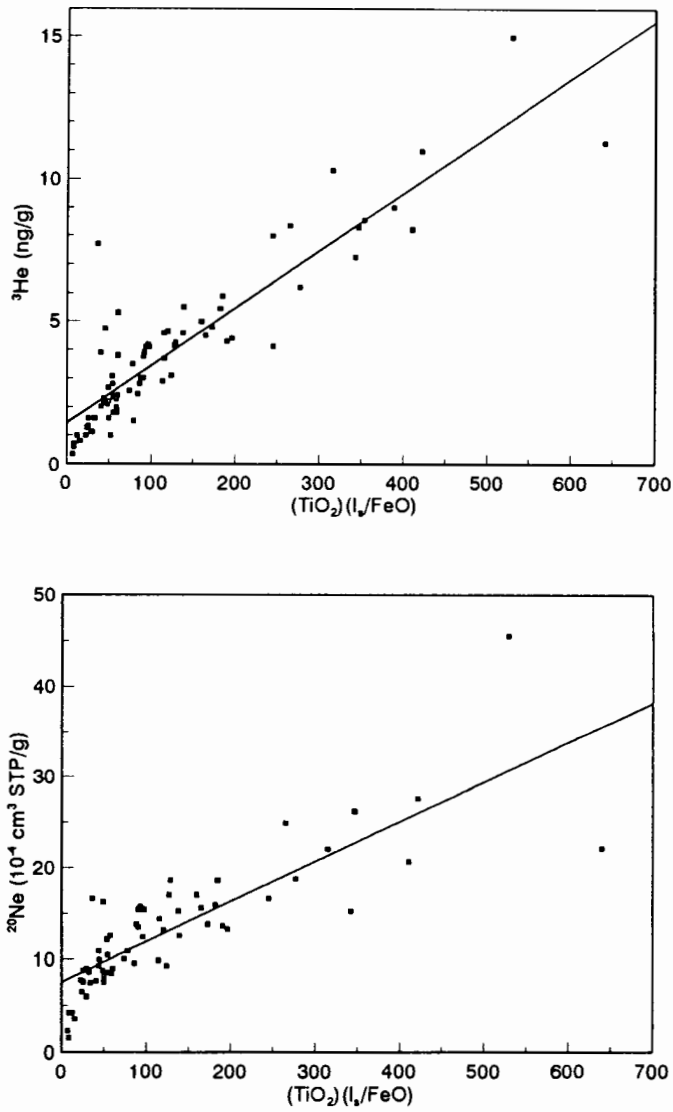


Figure 5. (top) The ^{3}He content (ng g^{-1}) plotted against the product of Ti content (% TiO_2) and I_s/FeO. The solid diagonal line is a linear least-squares fit. All soils for which the relevant data are available are included in this plot. Note the smaller amount of scatter relative to that in Fig. 4. Data from compilations by Swindle et al. (1990). (bottom) A similar plot of the ^{20}Ne content (cm^3 STP g^{-1}) in lunar soils. Again, note the smaller amount of scatter relative to that in Fig. 4.

A. Noble Gases

The noble gas data are presented in Tables I (Apollo sites) and II (Luna 16 and 20 sites). The Apollo values in Table I were calculated by first taking the mean of all bulk soil measurements for each soil that has been analyzed (references are given in Swindle et al. 1990), and then combining the values for individual soils to get means and medians for each site. The number of analyses available for a single soil ranges from one (nearly half the soils) to 11 (10084). For the Luna samples, where different identification systems have been used in different reports, all analyses of each site were averaged.

Some values in Table I are skewed by single measurements or single soils. In particular, the Apollo 12 He and Ne give much different means (4.5 ng g^{-1} and 13.0 cm^3 STP g^{-1}, respectively) if 12030 is deleted. The 12030 numbers come from a single measurement by Heymann et al. (1972). They are about four times as high as the second highest value measured in an Apollo sample (79221). However, the values reported by Heymann's laboratory are not systematically different from other laboratories on the 11 soils where direct comparisons can be made. The Apollo 11 Xe similarly is driven up by one of two measurements of 10010 by Funkhouser et al. (1970) that is about twice as high as most other soils with high Xe contents. Finally, the low end of the Apollo 17 samples is always 74220, the orange soil. Although the 74220 analyses are confirmed by multiple measurements, the soil is certainly unrepresentative. However, enough Apollo 17 soils have been analyzed that it has little effect on the means or medians.

For Ar, Kr and Xe, there is no obvious relationship with landing site for the Apollo sites. Except for Apollo 11, the mean value from each Apollo site is within 1σ of the mean of every other site. The values for Apollo 11 are slightly higher, but this reflects the fact that the few soils analyzed are all mature, and these gases correlate strongly with maturity (Figs. 2a–c). For He and Ne, the Apollo 11 and Apollo 17 means are clearly higher than the other sites. In this case, it reflects the higher mean Ti content of these sites and the correlation of He and Ne with Ti (Fig. 4). The means for the Luna data are higher than the Apollo means for all gases. Although this could indicate a geographic dependence (Sec. II), the Luna 16 and 20 soils are apparently mature (Swindle et al. 1992), and the scatter in reported values is surprisingly high for such a small number of samples.

B. Hydrogen

The hydrogen contents of lunar samples have generally been determined by four techniques: (1) pyrolysis followed by H_2 analysis by gas chromatography (see, e.g., Bustin et al. 1984,1986; DesMarais et al. 1974; Gibson et al. 1987); (2) pyrolysis followed by manometric measurement of H_2 (see, e.g., Chang et al. 1974*b*); (3) acid hydrolysis followed by manometric measurement (see, e.g., Chang et al. 1974*b*) and (4) pyrolysis followed by separation of H_2 and H_2O which are measured manometrically and isotopically (see, e.g., Epstein and Taylor 1970,1971,1972,1973,1974,1975; Friedman et al.

TABLE I

Noble Gas Contents of Bulk Lunar Soils: Apollo Sites

Mission	No. of Soils	Abundance Min.	Max.	Median	Mean $\pm 1\sigma$
			^{3}He (ng g^{-1})		
Apollo 11	3	8.0	11.3	9.2	9.5 ± 1.7
Apollo 12	8	2.6	21.0	4.9	6.6 ± 5.9
Apollo 14	7	2.8	4.6	3.5	3.7 ± 0.8
Apollo 15	19	1.0	5.6	3.7	3.5 ± 1.1
Apollo 16	29	0.4	2.8	1.7	1.7 ± 0.6
Apollo 17	24	0.6	15.0	7.2	6.7 ± 3.2
All Soils	90	0.4	21	3.6	4.2 ± 3.4
			^{20}Ne (10^{-4} cm^3STP g^{-1})		
Apollo 11	3	22.1	25.6	23.9	23.9 ± 1.8
Apollo 12	6	7.3	61.4	14.6	21.1 ± 20.0
Apollo 14	7	9.3	15.2	10.9	11.5 ± 2.2
Apollo 15	12	8.2	17.0	14.9	14.4 ± 2.4
Apollo 16	29	2.4	12.6	7.9	7.6 ± 2.5
Apollo 17	18	1.6	45.6	18.7	19.7 ± 8.7
All Soils	75	1.6	61.4	12.6	13.7 ± 9.0
			^{36}Ar (10^{-4} cm^3STP g^{-1})		
Apollo 11	3	3.4	3.9	3.8	3.7 ± 0.2
Apollo 12	6	1.0	2.9	2.6	2.2 ± 0.8
Apollo 14	7	2.4	4.4	3.2	3.2 ± 0.6
Apollo 15	12	1.6	3.6	2.8	2.7 ± 0.7
Apollo 16	29	1.2	5.9	3.5	3.4 ± 1.4
Apollo 17	18	0.2	6.0	3.1	2.9 ± 1.2
All Soils	75	0.2	6.0	3.1	3.1 ± 1.2
			^{84}Kr (10^{-8} cm^3STP g^{-1})		
Apollo 11	3	20.5	22.0	20.9	21.1 ± 0.8
Apollo 12	6	4.0	19.8	13.8	12.0 ± 6.4
Apollo 14	6	9.3	19.7	17.7	15.8 ± 4.4
Apollo 15	10	5.9	23.9	12.0	13.5 ± 6.2
Apollo 16	28	4.9	26.8	14.8	14.1 ± 5.7
Apollo 17	18	0.8	22.3	16.3	14.9 ± 6.2
All Soils	71	0.8	26.8	15.5	14.5 ± 5.8
			^{132}Xe (10^{-8} cm^3STP g^{-1})		
Apollo 11	3	3.0	7.0	3.1	4.4 ± 2.3
Apollo 12	6	1.1	2.5	2.0	1.7 ± 0.6
Apollo 14	6	2.0	3.6	2.9	2.8 ± 0.6
Apollo 15	10	0.8	3.3	1.9	2.0 ± 0.8
Apollo 16	28	0.8	3.8	2.4	2.4 ± 0.9
Apollo 17	18	0.2	3.1	2.3	2.1 ± 0.7
All Soils	71	0.2	7	2.4	2.4 ± 1.0

TABLE II
Noble Gas Contents of Bulk Lunar Soils: Luna Sites

Mission	No. of Soils	Abundance Min.	Max.	Median	Mean $\pm 1\sigma$
			^{3}He ($ng\ g^{-1}$)		
Luna 16	5	3.4	10.8	8.4	7.5 ± 2.9
Luna 20	3	1.5	3.0	2.3	2.3 ± 0.8
			^{20}Ne (10^{-4} $cm^3STP\ g^{-1}$)		
Luna 16	5	11.2	75.1	25.3	32.6 ± 25.3
Luna 20	4	9.2	17.8	11.5/14.3	13.2 ± 3.7
			^{36}Ar (10^{-4} $cm^3STP\ g^{-1}$)		
Luna 16	5	2.7	8.6	5.4	5.4 ± 2.2
Luna 20	4	4.8	6.5	4.9/5.8	5.5 ± 0.8
			^{84}Kr (10^{-8} $cm^3STP\ g^{-1}$)		
Luna 16	4	9.0	35.2	22.0/25.3	22.9 ± 10.8
Luna 20	3	13.8	23.6	23.1	20.2 ± 5.5
			^{132}Xe (10^{-8} $cm^3STP\ g^{-1}$)		
Luna 16	4	1.4	8.5	2.2/5.6	4.4 ± 3.3
Luna 20	3	1.9	3.4	3.0	2.8 ± 0.8

1970,1971,1972; Merlivat et al. 1974,1976; Stievenard et al. 1990). The results of hydrogen determinations on bulk lunar soils are listed by Apollo mission in Table III and by analysis in Table IV.

TABLE III
Hydrogen Contents of Bulk Lunar Soils

Mission	No. of Soils	Hydrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	1	38	66	46/54.2	52 ± 10
Apollo 12	5	1.9	46	30/38	31 ± 14
Apollo 14	5	35.9	106	61/67	67 ± 25
Apollo 15	14	28	120	52.6	57 ± 22
Apollo 16	15	19	79	36.3/39	46 ± 17
Apollo 17	14	0.2	53	22/25.7	26 ± 18
All Soils[a]	54	0.2	120	45	46 ± 16

[a] Excluding the orange soil 74220, the minimum value becomes $1.9\ \mu g\ g^{-1}$ and the median and mean values are unchanged.

The first thing to note is that the reported hydrogen contents range over a factor of 600 from a low of $\sim 0.2\ \mu g\ g^{-1}$ reported by Epstein and Taylor (1973) for the orange soil 74220 to a high of $\sim 120\ \mu g\ g^{-1}$ reported by Friedman et

TABLE IV
Comparison of Analytical Results for Hydrogen in Bulk Lunar Soils

Reference	No. of Analyses	Hydrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Bustin et al. 1984	16	17	106	47/50	49 ± 22
Bustin et al. 1986	5	26	66	39	41 ± 16
Chang et al. 1974*b*	7	0.6	22	7.8	11 ± 8
DesMarais et al. 1974	18	31	92	62	59 ± 15
Epstein and Taylor 1970	2	45	46	—	46
Epstein and Taylor 1971	3	1.9	40.7	38	27 ± 23
Epstein and Taylor 1972	3	35.9	52.6	49.4	46 ± 10
Epstein and Taylor 1973	4	0.2	61.7	7.9/28.8	25 ± 30
Epstein and Taylor 1975	13	28.2	58.5	36.3	40 ± 9
Friedman et al. 1971	2	30	46	—	38
Friedman et al. 1972	5	66	120	85	89 ± 20
Gibson et al. 1987	5	25.7	54.2	39.2	41 ± 11
Merlivat et al. 1974	6	28	53	34.5/36.1	38 ± 9
All Soils[a]	89	0.2	120	45	46 ± 22

[a] Excluding the two analyses of the orange soil 74220 (0.2 and 0.6 $\mu g\ g^{-1}$) the minimum value becomes 1.9 $\mu g\ g^{-1}$ and the median and mean values are unchanged.

al. (1972) for lunar soil 15271. The latter value appears anomalously high by comparison with another split of the same soil which has a hydrogen content of 38 $\mu g\ g^{-1}$ (Bustin et al. 1984). However, the Friedman et al. (1972) value is a combination of hydrogen found as elemental hydrogen and hydrogen found combined in water. If the Friedman et al. (1972) analysis of 15271 were not included in the data set, the highest hydrogen content in a lunar soil would then be 106 $\mu g\ g^{-1}$ for 14003 (Bustin et al. 1984).

On the other hand, the low hydrogen content of 74220 is apparently real. Chang et al. (1974*b*) reported a value of 0.6 $\mu g\ g^{-1}$ hydrogen for another split of the orange soil. This soil also has an unusually low carbon content of ~4 to 11 $\mu g\ g^{-1}$ (Epstein and Taylor 1973; Chang et al. 1974*b*; Moore et al. 1974). However, the low hydrogen content of 74220 is probably not representative of the hydrogen contents of most bulk lunar soils. As Table III shows, the median hydrogen content of 54 soils is 45 $\mu g\ g^{-1}$ and the mean hydrogen content of the same soils is 46 ± 16 $\mu g\ g^{-1}$. These values do not change if the data for 74220 are excluded from the statistical calculations. Excluding the possibly suspect value of 120 $\mu g\ g^{-1}$ for 15271 also has only a small effect on the calculated mean which decreases to 45 $\mu g\ g^{-1}$.

Another important point illustrated by the data in Table III is that the hydrogen contents of bulk lunar soils at the different Apollo sites are about the same. To a first approximation, the mean hydrogen contents of bulk soils at the Apollo 12 and Apollo 17 sites appear slightly lower than the mean

hydrogen contents of bulk soils at the other four Apollo sites. This difference is not significant at the 1σ level, but it is important to remember that this simple comparison does not attempt to take any chemical and mineralogical differences into account.

When an attempt is made to consider variations in soil properties a slightly different picture emerges. Hydrogen, unlike He, does not correlate with either the Ti content of lunar soils or the product of Ti content and I_s/FeO. However, as Fig. 2d illustrates, the hydrogen content does correlate with I_s/FeO itself, with more hydrogen being found in more mature soils with higher I_s/FeO values. This correlation may be responsible for the generally low hydrogen contents reported by Chang et al. (1974*b*) in immature and submature soils from the Apollo 17 landing site. But the considerable scatter in Fig. 2d suggests that it may not be economically advantageous to beneficiate soils on the basis of maturity for hydrogen mining.

The hydrogen contents of lunar basalts and breccias (soil and regolith) are given in Table V. The hydrogen contents of breccias are variable and range over 2 orders of magnitude from a low of 1 $\mu g\ g^{-1}$ in breccia 76015 (Gibson et al. 1987) to a high of 101 $\mu g\ g^{-1}$ in breccia 10061 (Epstein and Taylor 1970). The median and mean values of 20 $\mu g\ g^{-1}$ and 28 $\mu g\ g^{-1}$, respectively, are slightly lower than the corresponding values for bulk lunar soils, although the 1σ uncertainties on the means for the soils and breccias overlap.

In contrast, the hydrogen contents of lunar basalts are only a few $\mu g\ g^{-1}$ or less with a maximum value of 12 $\mu g\ g^{-1}$ for 15058 (DesMarais et al. 1974) and a minimum value of $<0.6\ \mu g\ g^{-1}$ for 10072 (Epstein and Taylor 1970). The most recent study of hydrogen in lunar basalts by Stievenard et al. (1990) gives only $\sim 1\ \mu g\ g^{-1}$ hydrogen in 70215, a basalt which has been studied extensively by Merlivat et al. (1974,1976) as well. Stievenard et al. (1990) note that the hydrogen in 70215 contains a deuterium-rich component and thus is probably formed by spallogenic processes. Their studies as well as those of Merlivat et al. (1974,1976) on splits of the sample indicates that the indigenous hydrogen content of lunar basalts is below 1 $\mu g\ g^{-1}$. These results have important implications for the formation of lunar ore deposits by volcanic gas transport in the lunar interior (see, e.g., Fegley 1991).

C. Carbon

The carbon contents of lunar samples have been measured by a variety of techniques: (1) pyrolysis followed by manometry and mass spectrometry (see, e.g., Epstein and Taylor 1970); (2) pyrolysis, or combustion in O_2 followed by gas chromatography (GC) or by gas chromatography mass spectrometry (see, e.g., Burlingame et al. 1971; DesMarais et al. 1973; DesMarais 1978; Gibson and Moore 1973; Moore et al. 1970); (3) solvent extraction followed by GC-MS (see, e.g., Burlingame et al. 1971); and (4) dissolution/hydrolysis with mineral acids and/or deuterated acids followed by gas chromatography, mass spectrometry, or GC-MS (see, e.g., Cadogan et al. 1972; Chang et al. 1970; Sakai et al. 1972; Wszolek et al. 1972). The analytical results for bulk

TABLE V

Hydrogen Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Hydrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Breccias					
Epstein and Taylor 1970	1	101	101	101	101
Friedman et al. 1970	3	50	66	60	59 ± 10
Friedman et al. 1972	4	8	38	20	22 ± 15
Gibson et al. 1987	13	1	60.4	3.5	16 ± 23
Merlivat et al. 1974	2	34.5	39.1	—	37
All Breccias	23	1	101	20	28 ± 28
Basalts					
DesMarais et al. 1974	2	4	12	—	8
Epstein and Taylor 1971	1	<0.6	<0.6	<0.6	<0.6
Friedman et al. 1971	3	3	9	6.6	6.2 ± 3.5
Friedman et al. 1972	2	2	2.8	—	2.4
Gibson et al. 1987	16	1	3.8	1.8	2.0 ± 0.7
Merlivat et al. 1974	3	0.65	2.42	1.51	1.5 ± 0.9
Merlivat et al. 1976	8	1.25	2.96	1.55/1.98	2.0 ± 0.6
Stievenard et al. 1990	1	0.97	0.97	0.97	0.97
All Basalts	36	<0.6	12	2	2.7 ± 2.4

TABLE VI
Carbon Contents of Bulk Lunar Soils

Mission	No. of Soils	Carbon Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	2	116	226	155/157	162 ± 27
Apollo 12	13	23	250	120	113 ± 46
Apollo 14	12	42	186	120	117 ± 35
Apollo 15	35	2.5	175	98	94 ± 44
Apollo 16	33	31	280	140	125 ± 52
Apollo 17	39	3.5	200	130/135	124 ± 45
All Soils	134	2.5	280	120	124 ± 45

lunar soils are listed in Tables VI and VII.

The reported carbon contents range over 2 orders of magnitude, but the median and mean values are within a factor of 3 of the corresponding values for hydrogen. The data in Table VI apparently indicate that the carbon contents of bulk soils at the Apollo 11 site are slightly higher than those at other sites. This apparent difference is not statistically significant and because of the small number of samples does not bias the calculated median and mean. The highest carbon content reported for any bulk lunar soil is 280 $\mu g\ g^{-1}$ for 64421 (Moore et al. 1973). The lowest reported carbon content is 2.5 $\mu g\ g^{-1}$ for 15426 (Modzeleski et al. 1972). There are no replicates for either analysis. Four analyses of different splits of the orange soil 74220 give carbon contents of 3.5 $\mu g\ g^{-1}$ (Epstein and Taylor 1973), 5 $\mu g\ g^{-1}$ (Moore et al. 1974), 11 $\mu g\ g^{-1}$ (Chang et al. 1974*b*), and 100 $\mu g\ g^{-1}$ (Moore et al. 1974). A comparison of the results obtained by different analytical groups in Table VII suggests some possible discrepancies between the different laboratories. The generally lower values reported by Wszolek et al. (1972) are for the Apollo 15 deep core (15001–15006) and for sample 15426. These values were determined by DF hydrolysis and are lower than carbon contents determined by other groups (Epstein and Taylor 1975; DesMarais et al. 1973) using pyrolysis on different splits of the same samples. The carbon contents determined by Modzeleski et al. (1972) for Apollo 15 samples are also generally lower than analyses by other groups on other splits of the same samples. However, within the calculated 1σ uncertainties, the results of Wszolek et al. (1972) and Modzeleski et al. (1972) overlap those of other investigators. The calculated mean carbon content of bulk soils will not be significantly changed by removing these data from the computations.

Table VIII displays the carbon contents of lunar basalts and breccias. As with hydrogen, the calculated mean value is slighly lower than that for bulk soils, but overlaps the soil value within the 1σ uncertainties. The reported carbon contents for breccias range over almost a factor of 80, and span almost the entire range covered by bulk lunar soils. The maximum of 385 $\mu g\ g^{-1}$ is

TABLE VII

Comparison of Analytical Results for Carbon in Bulk Lunar Soils

Reference	No. of Analyses	Carbon Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Burlingame et al. 1970	2	168	185	—	176
Chang et al. 1970	1	157	157	157	157
Chang et al. 1974*a*	7	11	88	50	56 ± 28
Chang et al. 1974*b*	6	129	160	135/141	140 ± 11
DesMarais et al. 1973	22	52	125	81	83 ± 19
DesMarais et al. 1975	12	36	152	120/123	111 ± 35
Epstein and Taylor 1970	2	143	155	—	149
Epstein and Taylor 1971	3	30	145	125	100 ± 68
Epstein and Taylor 1972	3	43	89	60	64 ± 27
Epstein and Taylor 1973	4	3.5	141	95/97	84 ± 67
Epstein and Taylor 1975	8	51	216	101/144	123 ± 50
Friedman et al. 1970	1	188	188	188	188
Friedman et al. 1971	3	116	139	133	129 ± 14
Friedman et al. 1972	5	51	110	99	88 ± 25
Gibson and Moore 1973	4	131	184	150/154	155 ± 26
Kaplan and Petrowski 1971	9	37	140	112	105 ± 28
Kaplan et al. 1970	4	116	170	143/147	144 ± 26
Kaplan et al. 1976	10	81	154	125/127	122 ± 24
Kerridge et al. 1975*b*	14	31	180	92/126	116 ± 43
Kerridge et al. 1978*a*	18	23	163	111/119	106 ± 43
Modzeleski et al. 1972	26	2.5	100.5	35.2/35.4	41 ± 18
Moore and Lewis 1975	37	59	200	140	142 ± 31
Moore et al. 1970	2	142	226	—	184
Moore et al. 1971	22	23	250	130/135	122 ± 53
Moore et al. 1972	15	42	186	135	128 ± 36
Moore et al. 1973	56	29	280	130	128 ± 43
Moore et al. 1974	40	5	200	140	131 ± 42
Sakai et al. 1972	2	100	138	—	119
Wszolek et al. 1972	6	29	92	36/43	49 ± 25
All Soils	344	2.5	280	120	124 ± 45

higher than the maximum carbon content of $280\,\mu g\ g^{-1}$ in bulk soils. The carbon contents of basalts vary over a smaller range of ~5 to $70\,\mu g\ g^{-1}$ with a mean of $26\,\mu g\ g^{-1}$.

The carbon analyses of breccias are generally in good agreement. In the case of 14321, the different results are 21, 22, 28, 32, 43 $\mu g\ g^{-1}$ (Moore et al. 1972), $24\,\mu g\ g^{-1}$ (Epstein and Taylor 1972), and $27\,\mu g\ g^{-1}$ (Sakai et al. 1972). For 79135, the results are $150\,\mu g\ g^{-1}$ (Moore et al. 1974), and 136 and $157\,\mu g\ g^{-1}$ (Becker and Epstein 1981). However, there is also

TABLE VIII
Carbon Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Carbon Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
		Breccias			
Becker and Epstein 1981	6	99	385	138/152	188 ± 96
Epstein and Taylor 1970	1	262	262	262	262
Epstein and Taylor 1972	1	24	24	24	24
Friedman et al. 1970	3	113	218	181	171 ± 62
Friedman et al. 1972	4	22	50	45/50	42 ± 14
Kaplan et al. 1970	4	132	198	137/181	162 ± 32
Kaplan et al. 1976	2	62	76	—	69
Moore et al. 1970	2	102	230	—	166
Moore et al. 1971	2	65	120	—	93
Moore et al. 1972	19	21	225	80	94 ± 66
Moore et al. 1973	24	5	160	39/45	54 ± 42
Moore et al. 1974	6	21	150	54/85	73 ± 46
Sakai et al. 1972	1	27	27	27	27
All Breccias	75	5	385	65/70	93 ± 73
		Basalts			
DesMarais 1978	9	5.3	41.1	8.6	16 ± 12
Friedman et al. 1971	3	27	61	42	43 ± 20
Friedman et al. 1972	2	7.3	7.7	—	7.5
Kaplan and Petrowski 1971	2	14	19	—	16
Moore et al. 1970	2	64	70	—	67
Moore et al. 1972	1	35	35	35	35
Moore et al. 1973	6	10	27	13/16	16 ± 6
Moore et al. 1974	3	31	65	54	50 ± 20
All Basalts	28	5.3	70	18/19	26 ± 20

considerable disagreement between some replicate analyses, apparently due to heterogeneous carbon contents. For example, Becker and Epstein (1981) report 220 and 385 $\mu g\ g^{-1}$ for 10018, and Moore et al. (1972) report 140 and 210 $\mu g\ g^{-1}$ for 14047. Presumably this heterogeneity accounts for at least some of the scatter in the analytical results.

A comparison of the carbon analyses of basalts indicates that some real disagreements exist between different analytical groups. For example, the three basalts 15058, 70215 and 75035 have been analyzed for carbon by DesMarais (1978) and Moore et al. (1973,1974). In all cases higher carbon contents are reported by Moore and colleagues. The mean carbon content for the lunar basalts analyzed by Moore et al. (1970,1972,1973,1974), repre-

senting nearly half the total data, is $34 \pm 23\ \mu g\ g^{-1}$, while the mean carbon content for lunar basalts analyzed by all other investigators is $20 \pm 16\ \mu g\ g^{-1}$. However, both values are well within the 1σ uncertainty of the total mean, and they disagree with each other by less than the 1σ uncertainty in either.

D. Nitrogen

Table IX gives the nitrogen contents of bulk lunar soils at the different Apollo landing sites and at two of the Luna landing sites. These data were obtained by the following analytical techniques: (1) pyrolysis and/or combustion in O_2 followed by mass spectrometry (see, e.g., Becker and Clayton 1975; Kerridge et al. 1975*a*, *b*); (2) dissolution/hydrolysis with mineral acids followed by mass spectrometry (see, e.g., Chang et al. 1974*a*, *b*; Kerridge et al. 1975*a*, *b*); (3) neutron activation analysis (see, e.g., Goel and Kothari 1972); (4) pyrolysis followed by gas chromatography (see, e.g., Moore et al. 1970); (5) Kjeldahl analysis for chemically bound N (see, e.g., Müller 1972); and (6) spark source mass spectrometry (see, e.g., Morrison et al. 1970).

TABLE IX

Nitrogen Contents of Bulk Lunar Soils

Mission	No. of Soils	Nitrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	2	93	153	110	121 ± 27
Apollo 12	12	5.9	140	71.1	67 ± 34
Apollo 14	2	80	92	80/92	86
Apollo 15	19	18	135	96	81 ± 31
Apollo 16	41	4	141	90	85 ± 34
Apollo 17	35	7.3	130	73	72 ± 29
LUNA 16	1	242	242	242	242
LUNA 20	1	102	102	102	102
All Soils	111	7.3	242	84	81 ± 37

The reported nitrogen contents vary over a factor of 30 from a low of $7.3\ \mu g\ g^{-1}$ in 74220 (Chang et al. 1974*b*) to a high of $242\ \mu g\ g^{-1}$ in the Luna 16 soil (Goel et al. 1974*b*). There are no replicates by other groups for either analysis. Nitrogen, like carbon and hydrogen, does not show clear variations between the different Apollo sites. However, the one set of analyses for the Luna 16 soil show markedly higher nitrogen contents. As shown in Fig. 2f, nitrogen contents show a strong correlation with I_s/FeO.

A comparison of the nitrogen analyses by different groups is given in Table X. The lowest values were reported by Chang et al. (1974*b*) for submature and immature Apollo 17 soils including the orange soil 74220. The nitrogen contents reported by Chang et al. (1974*b*) follow a trend also exhibited by hydrogen and carbon in these soils and confirmed for these two elements by analyses done by other groups. With the exception of the Luna 16 analyses

reported by Goel et al. (1974*b*), no other group clearly stands out from the rest on the high side. However, a soil by soil comparison of analyses done by different groups generally shows that the results of Moore and colleagues tend to be higher than other analyses on the same soil. Given the agreement between the other analytical groups, especially the UCLA and Chicago laboratories, some of the results of Moore and colleagues may be anomalously high. However, their median and mean values do not appear out of line with the other values summarized in Table X and do not affect the overall mean.

TABLE X

Comparison of Nitrogen Analyses of Bulk Lunar Soils

Reference	No. of Analyses	Nitrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Becker and Clayton 1975	15	43	114	83	80 ± 27
Becker and Clayton 1977	9	22.7	70	44	45 ± 13
Becker and Clayton 1978	2	37	75	—	56
Becker et al. 1976	3	17	62	51	43 ± 27
Chang et al. 1974*a*	7	7.3	54.6	23.4	29 ± 17
Chang et al. 1974*b*	8	88	135	96/98	101 ± 15
Goel and Kothari 1972	1	80	80	80	80
Goel et al. 1974*a*	8	65	242	102/120	119 ± 54
Goel et al. 1974*b*	3	102	242	102/242	195 ± 83
Kaplan et al. 1976	10	33	110	90/91	79 ± 27
Kerridge et al. 1975*a*	29	15	125	68	72 ± 33
Kerridge et al. 1975*b*	14	4	125	68/83	74 ± 35
Kerridge et al. 1978*b*	18	5.9	74.8	53.6/54.6	47 ± 26
Kothari and Goel 1973	10	58	141	95/98	101 ± 26
Moore and Lewis 1975	40	61	137	96/97	99 ± 19
Moore et al. 1970	2	102	153	—	128
Moore et al. 1971	15	46	140	96	94 ± 28
Morrison et al. 1970	1	110	110	110	110
Morrison et al. 1971	1	40	40	40	40
Müller 1972	5	80	109	92	91 ± 12
Müller 1973	9	70	124	96	97 ± 20
Müller 1974	10	23	110	70/73	69 ± 27
Müller et al. 1976	3	93.4	110	94	99 ± 10
Wszolek et al. 1972	6	18	89	21/36	45 ± 33
All Soils	229	4	242	87	81 ± 37

The nitrogen analyses of basalts and breccias, which are summarized in Table XI, follow the trends already noted for hydrogen and carbon. The mean nitrogen content of breccias is slightly lower than that of bulk soils, but the two overlap within their mutual uncertainties. Likewise, the nitrogen contents of basalts are again markedly lower, although the nitrogen analyses display

more variability than the carbon analyses. The results for basalts apparently fall into two groups with one group being $\leq 2\,\mu g\ g^{-1}$ and the other group being higher than $\sim 20\,\mu g\ g^{-1}$. It is tempting to ascribe the higher analyses to contamination and/or other analytical problems. If this explanation is valid, then the true nitrogen content of lunar basalts is $\sim 1\,\mu g\ g^{-1}$ or less (see, e.g., DesMarais 1978).

TABLE XI

Nitrogen Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Nitrogen Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
		Breccias			
Becker and Clayton 1975	6	4.3	62	48	37 ± 23
Becker and Epstein 1981	6	72.3	114.1	74.3/83	87 ± 17
Becker et al. 1976	3	1.8	7.5	2.5	4 ± 2
Frick et al. 1988	1	114	114	114	114
Goel and Kothari 1972	8	22	136	41/55	58 ± 34
Kaplan et al. 1976	2	41	55	—	48
Kothari and Goel 1973	3	16	74	22	37 ± 26
Moore et al. 1970	2	98	125	—	112
Moore et al. 1972	2	57	130	—	94
Müller 1972	1	131	131	131	131
All Breccias	34	1.8	136	62	61 ± 40
		Basalts			
Becker and Clayton 1975	3	0.6	3.5	1	1.7 ± 1.2
Becker et al. 1976	1	0.5	0.5	0.5	0.5
DesMarais 1978	9	0.1	1.0	0.7	0.6 ± 0.4
Goel and Kothari 1972	1	21	21	21	21
Kothari and Goel 1973	1	18	18	18	18
Moore et al. 1970	2	30	116	—	73
Müller 1972	3	<10	64	—	—
Müller 1974	1	<8	<8	<8	< 8
Müller et al. 1976	7	3	127	5	39 ± 55
All Basalts	28	0.1	127	3.1	19 ± 38

E. Sulfur

Sulfur contents of lunar samples have been determined by the following methods: (1) spark source mass spectrometry (Bouchet et al. 1971; Morrison et al. 1970); (2) dissolution/hydrolysis with mineral acids followed by mass spectrometry (see, e.g., Chang et al. 1974*a*, *b*; Kaplan et al. 1970; Kerridge et al. 1975*a*, *b*; Rees and Thode 1972); (3) pyrolysis or combustion in O_2 followed

by mass spectrometry (see, e.g., DesMarais 1978); (4) X-ray fluorescence (see, e.g., Palme et al. 1978; Wänke et al. 1975); (5) fusion and SO_2 titration (see, e.g., Moore et al. 1972); and (6) combustion in O_2 followed by quantitative infrared spectroscopy on SO_2 (see, e.g., Gibson and Moore 1973). The analytical results for bulk lunar soils are listed in Tables XII and XIII, and the sulfur contents of lunar basalts and breccias are displayed in Table XIV. These data illustrate several important points.

TABLE XII
Sulfur Contents of Bulk Lunar Soils

Mission	No. of Soils	Sulfur Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	2	640	1750	770	922 ± 466
Apollo 12	14	530	1100	756	745 ± 108
Apollo 14	9	682	1149	900	865 ± 133
Apollo 15	18	337	950	654/669	638 ± 123
Apollo 16	35	320	829	592	591 ± 141
Apollo 17	37	212	1330	790/795	786 ± 261
All Soils	115	212	1750	704/706	715 ± 216

First, sulfur is much more abundant (on a mass basis) in lunar soils than H, C and N, and is the second most abundant chemically reactive volatile on an atomic basis. The S/C mass ratio is ~6, and the H/S atomic ratio is ~2. The high sulfur abundance means that extraction of other volatiles (e.g., by heating) may also be accompanied by the release of large amounts of sulfur gases if the major S-bearing mineral, troilite (FeS) decomposes during the extraction process.

Second, the sulfur content of lunar soils is much less variable than the H, C and N contents and varies by less than a factor of 8. The lowest value of $212\,\mu g\ g^{-1}$ was reported by Chang et al. (1974*a*) for the orange soil 74220 and the highest value of $1750\,\mu g\ g^{-1}$ was reported by Bouchet et al. (1971) for 10084. However, both of these extreme values may be incorrect. Two replicate analyses on 10084 by Kaplan et al. (1970) give $770\,\mu g\ g^{-1}$ sulfur, and other analyses of 74220 give higher sulfur contents in the range of $442\,\mu g\ g^{-1}$ (Chang et al. 1974*b*) up to $830\,\mu g\ g^{-1}$ (Gibson and Moore 1973; Gibson and Andrawes 1978*b*). Other sulfur determinations by Bouchet et al. (1971), who used spark source mass spectrometry, are generally higher than analyses by other groups on the same samples. Chang et al. (1974*b*) used acid hydrolysis to determine sulfur in 74220 and other Apollo 17 soils. They commented on a relationship between decreasing sulfur content and increasing ^{34}S isotopic enrichment in different aliquots of 74220 but stated that it is not due to incomplete H_2S release by acid hydrolysis of troilite. Furthermore, we note that a comparative study of acid hydrolysis and fusion techniques by Rees and Thode (1974) shows good agreement between results

TABLE XIII

Comparison of Sulfur Analyses of Bulk Lunar Soils

Reference	No. of Analyses	Sulfut Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Bouchet et al. 1971	2	1100	1750	—	1425
Chang et al. 1974*a*	9	212	1199	578	623 ± 297
Chang et al. 1974*b*	5	556	950	687	709 ± 131
Cripe and Moore 1975	3	610	640	610/640	630 ± 17
Gibson and Andrawes 1978*b*	16	410	830	555/560	753 ± 147
Gibson and Moore 1973	20	474	794	645/652	642 ± 110
Gibson and Moore 1974	12	560	1300	822/850	915 ± 244
Kaplan and Petrowski 1971	2	620	714	—	667
Kaplan et al. 1970	4	640	770	680/770	715 ± 57
Kaplan et al. 1976	10	360	710	670/680	643 ± 99
Kerridge et al. 1975*b*	13	365	829	673	617 ± 150
Kerridge et al. 1975*c*	19	320	810	534	552 ± 155
Kerridge et al. 1978	18	594	811	763/765	752 ± 53
Moore et al. 1972	5	900	1000	930	940 ± 36
Moore et al. 1974	21	710	1330	950	950 ± 174
Morrison et al. 1970	1	530	530	530	530
Palme et al. 1978	1	740	740	740	740
Rees and Thode 1972	13	549	1149	745	765 ± 145
Rees and Thode 1974	9	500	1014	620	709 ± 188
Sakai et al. 1972	2	337	920	—	629
Wänke et al. 1975	3	500	585	510	532 ± 38
All Soils	188	212	1750	704/706	715 ± 216

obtained by three different methods. Thus, there is no reason to doubt the low value obtained for 74220, except that it is unconfirmed.

Finally, a comparison of the sulfur contents of soils, basalts and breccias also shows much less variability of only a factor of 2 between these materials. Unlike the solar-wind-derived volatiles, sulfur is more abundant in basalts than in either breccias or soils.

F. Fluorine and Chlorine

Fluorine and chlorine analyses of lunar samples have been done by the following methods: (1) neutron activation analysis for F and Cl (see, e.g., Jovanovic and Reed 1973,1974); (2) halogen extraction by pyrohydrolysis followed by F analysis with a fluoride-sensitive electrode (see, e.g., Palme et al. 1978; Wänke et al. 1972); (3) neutron activation analysis for Cl (see, e.g., Wänke et al. 1972); and (4) spark source mass spectrometry for F and Cl (see, e.g., Bouchet et al. 1971; Morrison et al. 1971). Most of the analyses in the literature have been done by methods (1) to (3). The results for fluorine are given in Tables XV–XVII and the results for chlorine are given in Tables XVIII–XX.

TABLE XIV
Sulfur Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Sulfur Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
		Breccias			
Cripe and Moore 1975	8	230	2100	910	926 ± 496
Gibson and Moore 1973	1	844	844	844	844
Gibson and Moore 1974	9	720	1020	890	866 ± 98
Kaplan et al. 1970	4	1070	1120	1070/1120	1095 ± 25
Kaplan et al. 1976	2	660	870	—	765
Moore et al. 1972	8	580	1020	910/950	846 ± 163
Moore et al. 1974	4	380	1110	500/860	713 ± 290
Palme et al. 1978	4	420	2150	470/1000	1010 ± 696
Rees and Thode 1974	3	400	828	500	576 ± 224
Wänke et al. 1975	4	560	1100	660/750	768 ± 203
Wänke et al. 1976	1	236	236	236	236
Wänke et al. 1977	9	420	1040	610	668 ± 197
All Breccias	57	230	2150	810	818 ± 339
		Basalts			
Cripe and Moore 1975	1	1970	1970	1970	1970
DesMarais 1978	7	642	1850	1400	1213 ± 479
Gibson and Moore 1974	6	1580	2770	2210	2105 ± 400
Gibson et al. 1975	15	600	2770	965	1274 ± 632
Gibson et al. 1976	32	1580	2770	1850/1865	1871 ± 238
Gibson et al. 1977	18	560	1520	1105/1110	1099 ± 262
Kaplan and Petrowski 1971	1	754	754	754	754
Moore et al. 1974	3	1860	3140	2040	2347 ± 566
Rees and Thode 1974	2	1223	1581	—	1402
Wänke et al. 1975	8	485	1620	1350/1370	1132 ± 442
Wänke et al. 1976	1	715	715	715	715
All Basalts	94	485	3140	1580	1513 ± 562

Fluorine contents of bulk lunar soils range from a low of $14\ \mu g\ g^{-1}$ in 67455 (Wänke et al. 1973) to a high of $230\ \mu g\ g^{-1}$ in 74241 (Jovanovic and Reed 1973). The mean F content given in Table XV has a large uncertainty which is 66% of the mean. However, the large scatter is probably not due to analytical errors. Because of the small number of analyses there are few multiple analyses of the same sample, but there are no obvious systematic differences involving the two most extensive data sets, those of the Mainz and Argonne groups. Considering only the data of Wänke and colleagues, the

TABLE XV
Fluorine Contents of Bulk Lunar Soils

Mission	No. of Soils	Fluorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	1	66	96	86	83 ± 18
Apollo 12	2	60	63	60/61	61 ± 1
Apollo 14	2	106	145	—	126
Apollo 15	2	45	59	—	52
Apollo 16	8	14	105	39/47	48 ± 28
Apollo 17	10	17	230	58	77 ± 61
All Soils	25	14	230	59/60	68 ± 45

TABLE XVI
Comparison of Fluorine Analyses of Bulk Lunar Soils

Reference	No. of Analyses	Fluorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Bouchet et al. 1971	2	63	96	—	80
Jovanovic and Reed 1973	5	47	230	102	109 ± 65
Jovanovic and Reed 1974	8	17	61	38/52	35 ± 24
Morrison et al. 1970	1	66	66	66	66
Morrison et al. 1971	1	60	60	60	60
Wänke et al. 1972	7	45	145	61	80 ± 32
Wänke et al. 1973	5	14	210	41	64 ± 62
Wänke et al. 1974	4	31	71	58/59	55 ± 15
Wänke et al. 1975	1	39	39	39	39
All Soils	34	14	230	59/60	68 ± 45

mean F content is $70\pm40\ \mu g\ g^{-1}$ if the KREEP sample 14163 ($145\ \mu g\ g^{-1}$ F) is included and $66\pm38\ \mu g\ g^{-1}$ is 14163 is excluded. Considering only the data of Reed and colleagues, the mean F content is $63\pm59\ \mu g\ g^{-1}$. Within the uncertainties the two calculated mean F contents are identical, although the results of Wänke and colleagues have a smaller uncertainty.

The chlorine contents of lunar soils range from a low of $8\ \mu g\ g^{-1}$ in 71501 (Jovanovic and Reed 1974) to a high of $350\ \mu g\ g^{-1}$ in 10084 (Morrison et al. 1970). Again, only a small number of Cl analyses have been done and apparently no Cl data are available for Apollo 15 soils. The formal 1σ uncertainty on the mean Cl content of lunar soils is larger than the mean value itself. If the two largest Cl contents of $350\ \mu g\ g^{-1}$ (reported for 10084 by Morrison et al. [1970]) and $280\ \mu g\ g^{-1}$ (reported for 14163 by Wänke et al. [1972]) are excluded, then the mean Cl content becomes $30\pm20\ \mu g\ g^{-1}$.

The calculated F/Cl mass ratio in lunar soils in which both elements have been analyzed is 2.4 ± 1.4. This can be contrasted with the F/Cl mass ratio of

TABLE XVII
Fluorine Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Fluorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
		Breccias			
Palme et al. 1978	1	19	19	19	19
Wänke et al. 1972	2	101	119	—	110
Wänke et al. 1974	6	14	90	31/40	44 ± 25
Wänke et al. 1975	2	15	36	—	26
Wänke et al. 1976	1	15	15	15	15
Wänke et al. 1977	5	31	94	59	64 ± 23
All Breccias	17	14	119	40	54 ± 32
		Basalts			
Wänke et al. 1975	5	27	78	49	54 ± 19
Wänke et al. 1976	1	22	22	22	22
All Basalts	6	22	78	43/49	49 ± 21

TABLE XVIII
Chlorine Contents of Bulk Lunar Soils

Mission	No. of Soils	Chlorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Apollo 11	1	27	350	84	154 ± 191
Apollo 12	1	24	48	30	34 ± 14
Apollo 14	2	40	280	—	160
Apollo 15	0	—	—	—	—
Apollo 16	10	12	41	19/22	24 ± 11
Apollo 17	16	8	103	18	30 ± 24
LUNA 16	1	24	24	24	24
LUNA 20	1	26	26	26	26
All Soils	32	8	350	25	44 ± 67

~0.09 in CI chondrites. The lunar ratio is larger because Cl is depleted in lunar soils relative to CI chondrites where its abundance is about 700 $\mu g\ g^{-1}$. The F/Cl mass ratio in lunar breccias and basalts in which both elements have been analyzed is 4.0±4.8. Again, the high value is due to a Cl depletion.

G. Depth Profiles of Volatiles in the Lunar Regolith

To find the total inventories, we need only to combine the estimated volatile contents with estimates of the variations in volatile abundance with location and with depth within the regolith.

TABLE XIX

Comparison of Chlorine Analyses of Bulk Lunar Soils

Reference	No. of Analyses	Chlorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
Bouchet et al. 1971	2	48	84	—	66
Jovanovic and Reed 1973	15	13	72	26	30 ± 16
Jovanovic and Reed 1974	9	8	103	16	31 ± 28
Morrison et al. 1970	1	350	350	350	350
Morrison et al. 1971	1	30	30	30	30
Wänke et al. 1970	1	27	27	27	27
Wänke et al. 1971	1	24	24	24	24
Wänke et al. 1972	2	40	280	—	160
Wänke et al. 1973	2	12	20	—	16
Wänke et al. 1974	4	12	25	15/16	17 ± 5
Wänke et al. 1975	1	15	15	15	15
All Soils	39	8	350	25	44 ± 67

TABLE XX

Chlorine Contents of Lunar Breccias and Basalts

Reference	No. of Analyses	Chlorine Abundance $\mu g\ g^{-1}$ Min.	Max.	Median	Mean$\pm 1\sigma$
		Breccias			
Jovanovic and Reed 1981	1	195	195	195	195
Palme et al. 1978	2	18	51	—	35
Wänke et al. 1974	6	12	270	26/104	115 ± 110
Wänke et al. 1975	2	10	33	—	21
Wänke et al. 1976	1	7	7	7	7
Wänke et al. 1977	5	4	30	20	20 ± 9
All Breccias	17	4	270	26	65 ± 86
		Basalts			
Jovanovic and Reed 1980	2	1.4	5.9	—	3.7
Wänke et al. 1975	5	3.5	6	4.9	5 ± 0.9
Wänke et al. 1976	1	4	4	4	4
All Basalts	8	1.4	6	4.5/4.9	4.5 ± 1.5

The largest uncertainty in these calculations stems from our incomplete knowledge of the distribution of solar-wind-implanted volatiles with depth beneath the surface of the regolith. As the regolith is stirred ("gardened") by impacts, grains that had been on the surface are mixed to some depth. The solar-wind-rich layer could be as deep as the regolith itself, which, on average,

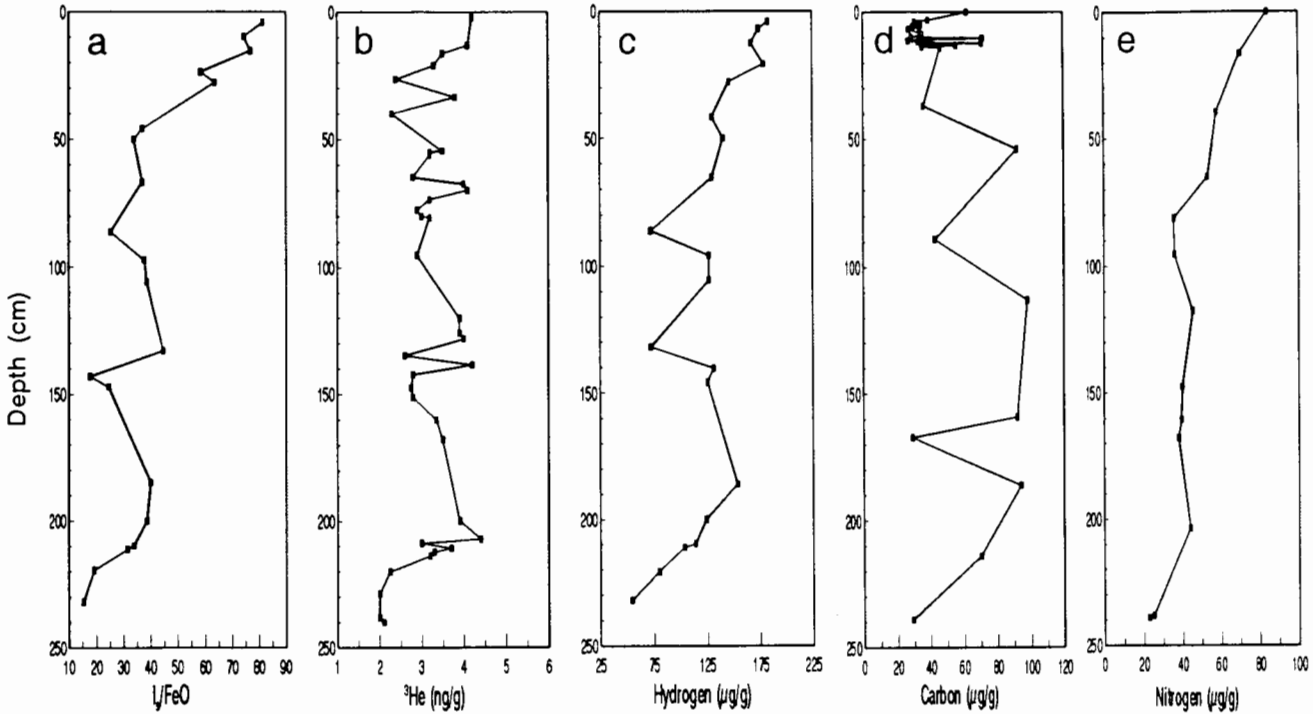


Figure 6. Profiles with depth for left: I_s/FeO; (center left) ^{3}He; (center) hydrogen; (center right) carbon and (right) nitrogen in the Apollo 15 deep drill core (samples 15001 to 15006). The I_s/FeO and H data are from Gibson and Bustin (1987); the ^{3}He data are from the compilation in Appendix A of Swindle et al. (1990); the C data are from DesMarais et al. (1973), Mozeleski et al. (1972), and Wszolek et al. (1972), and the N data are from Becker and Clayton (1977), Smith et al. (1973), and Wszolek et al. (1972). Note that the H data from Gibson and Bustin (1987) are for the sub-10 μm grain fraction.

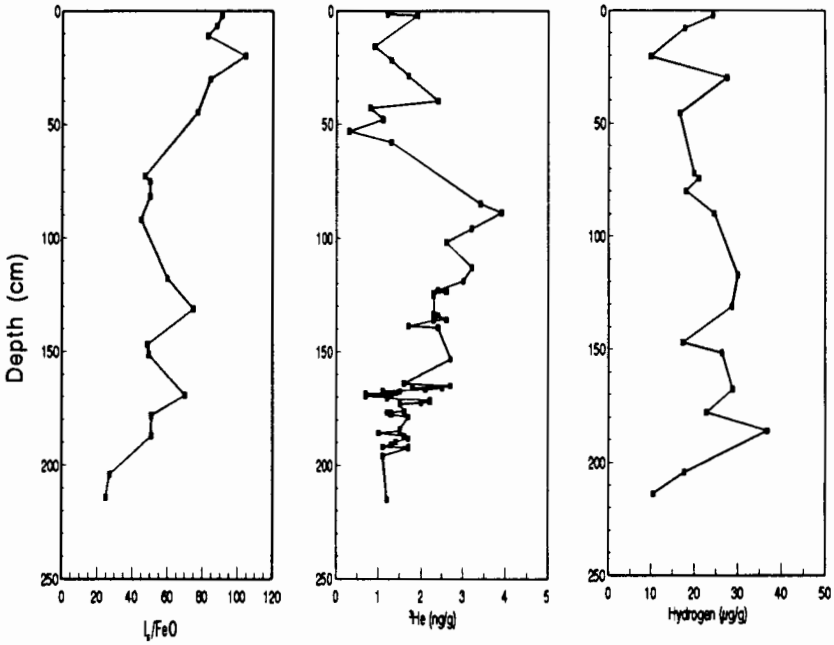


Figure 7. Profiles with depth for (left) I_s/FeO; (center) ^{3}He, and (right) hydrogen in the Apollo 16 deep drill core (samples 60001 to 60007). The I_s/FeO and H data are from Gibson et al. (1988) and the ^{3}He data are from Appendix A of Swindle et al. (1990).

is about 8 m deep in the maria and 15 m deep in the highlands (McKay et al. 1991, and references therein). On the other hand, quantitative models of regolith evolution typically predict something resembling an exponential decrease in the abundance of surface-exposed grains with depth, falling to $1/e$ of its surface value in a depth of about 3 m (see, e.g., Arnold 1975). Only three Apollo drill cores extend to greater than 1 m depth, and the deepest of these (Apollo 17) extends to only 3 m depth. The data from these drill cores (Figs. 6–8) are ambiguous, because the random variations (resulting from variations in maturity of layers of soil found at different depth) are large enough to mask the difference between a uniform abundance with depth and a substantial fall-off with depth. Most estimates of the total lunar inventory of volatiles assume a uniform concentration to a depth of 2 or 3 m and nothing deeper (see, e.g., Wittenberg et al. 1986; Haskin 1989; Taylor 1991), which tends to be closer to the exponential model considered above (i.e., a lower limit). For our estimates, we will assume a uniform concentration to 3 m depth and nothing deeper, but note that present evidence does not rule out the possibility that this accounts for only 20 to 30% of the total inventory.

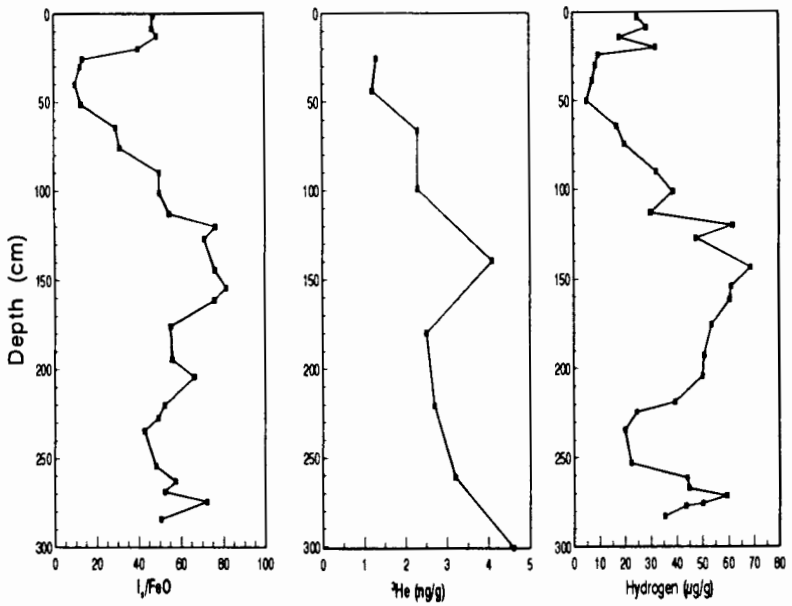


Figure 8. Profiles with depth for (a) left: I_s/FeO, (b) center: ^{3}He, and (c) right: hydrogen in the Apollo 17 deep drill core (samples 70001 to 70009). The I_s/FeO and H data are from Gibson et al. (1988) and the ^{3}He data are from Appendix A of Swindle et al. (1990).

H. Geographic Effects

In our estimates, we also ignore the possibility of latitude- and longitude-dependent effects. However, if saturation does not occur until doses reach three or more times those received at the central nearside, the total inventories could be a factor of 2 larger than our estimates. (Our estimates could also be about 10% too high, if saturation occurs at a fluence very close to that seen by the Apollo samples.)

For He and Ne, we also must somehow account for the variation in Ti content. We will use the method of Swindle et al. (1990), who averaged the Ti contents in mare and highland areas overflown by the Apollo orbiter gamma-ray spectrometers and assumed that those averages were typical of the Moon as a whole. They then converted these into ^{3}He abundances by using a plot analogous to Fig. 5a. This approach emphasizes the equatorial regions, but there is no evidence of systematic compositional variations with latitude. We will then use the average weight ratios of ^{4}He/^{3}He and Ne/^{3}He to calculate the abundances of those species.

I. Global Inventory Estimates

Our estimates are given in Table XXI. Several points should be emphasized from a resource perspective. First, none of these elements should be con-

sidered abundant. Typical concentrations are 125 μg g^{-1} or less, with the exception of sulfur. Second, the solar-wind volatiles come as a package; mining for one of them is mining for all of them. If mining of ^{3}He occurs, for every ton of ^{3}He extracted, more than 10^3 tons each of H, C, N and ^{4}He will also be extracted. Because these are valuable resources in their own right, any ^{3}He mining scheme should include processing of these gases as well. Third, these estimates are conservative; for reasons we have stated above, the actual inventories could be as much as an order of magnitude larger. Fourth, if we have underestimated the inventories, a substantial portion of those inventories must be at depths of >3 m and/or on the farside; this has obvious implications for mining scenarios.

TABLE XXI[a]

Globally Averaged Inventory of Volatiles on the Lunar Surface

Volatile	Concentration in Soil	Estimated Inventory (10^9 kg) This Work	Other Values
^{3}He	4.2 ± 3.4ng g^{-1}	0.84 ± 0.69	0.45–4.6 (Swindle et al. 1990); 1.1 (Wittenberg et al. 1989); 0.4–0.5 (Taylor 1991)
^{4}He	14.0 ± 11.3 μg g^{-1}	2890 ± 2300	
^{20}Ne	1.2 ± 0.8 μg g^{-1}	250 ± 160	
^{36}Ar	0.50 ± 0.19 μg g^{-1}	100 ± 40	
^{84}Kr	0.54 ± 0.22 ng g^{-1}	0.10 ± 0.04	
^{132}Xe	0.14 ± 0.06 ng g^{-1}	0.03 ± 0.01	
H	46 ± 16 μg g^{-1}	$9,300 \pm 3,200$	8,000 (Haskin 1989)
C	124 ± 45 μg g^{-1}	$25,200 \pm 9,100$	15,000 (Haskin 1989)
N	81 ± 37 μg g^{-1}	$16,500 \pm 7,500$	8,000 (Haskin 1989)
S	715 ± 216 μg g^{-1}	$145,000 \pm 44,000$	
F	70 ± 47 μg g^{-1}	$14,200 \pm 9,500$	
Cl[a]	44 ± 67 μg g^{-1}	$8,900 \pm 13,600$	

[a] As shown in Table XVIII, the formal 1σ uncertainty on the Cl concentration is greater than the concentration itself. If the two highest values of 350 μg g^{-1} (Morrison et al. 1970) and 280 μg g^{-1} (Wänke et al. 1972) are excluded the mean $\pm 1\sigma$ becomes 30 ± 20 μg g^{-1}. The corresponding inventory is then 6100 ± 4100 in units of 10^9 kg.

There are two other possible uncertainties, both most serious for hydrogen, that should also be mentioned. First, some unknown (but perhaps large) fraction of the hydrogen measured in lunar soils may be terrestrial contamination, so the true regolith inventory could be less than that estimated by us and others (see, e.g., Haskin 1989). This possibility is suggested by the isotopic studies of H_2 released from lunar samples (see, e.g., Epstein and Taylor 1970, and subsequent papers).

Second, as originally suggested by Watson et al. (1961), and later ex-

panded upon by Arnold (1976,1979,1987), it is possible that there are permanently shadowed regions near the lunar poles which remain cold enough to trap significant amounts of water ice. The estimated water inventory in the permanently shadowed regions is in the range of 10^{16} to 10^{17} g (Arnold 1979), or about 1 to 10 times as much hydrogen as contained as H_2 in lunar soils. The potential sources of lunar water are the solar-wind reduction of Fe^{2+} to Fe metal in the lunar regolith, the impact of water-bearing meteorites, cometary impact, and water degassed from the lunar interior. If the water from these sources could have been efficiently transported to the permanently shadowed regions at the lunar poles (covering ~0.02% of the lunar surface according to Hodges [1980]), then deposits with column densities of 10^3 g cm^{-2}, or greater, may have formed (Hodges 1991). However, a recent re-examination of the efficiency of the transport process by Hodges (1991) implies that transport is not efficient and that the amount of water ice in lunar polar deposits is probably too small to be of practical interest. Instead, Hodges (1991) proposes that the water is assimilated into the lunar regolith, which on average would give 30 μmoles g^{-1} of H_2O in lunar soils. This alternative requires efficient exchange of lunar water for terrestrial water in all lunar samples, none of which show evidence for any water other than terrestrial contamination (see Sec. V below). However, presumably this alternative can be tested with more careful analyses of existing lunar samples or by the collection of new samples under very carefully controlled conditions. Likewise, remote sensing observations from lunar polar orbit, or observations by robotic probes deployed on the lunar surface are apparently required to test unambiguously the concept of water ice deposits in the lunar polar regions.

IV. EXTRACTION OF VOLATILES FROM LUNAR SOILS

The extensive laboratory studies of volatiles in lunar samples suggest two obvious ways to extract solar-wind-derived volatiles from lunar soils: (1) heating bulk soils, and (2) grain-size sorting of bulk soils followed by volatile extraction from the finest size fractions. In the case of He and Ne, beneficiation by concentration of ilmenite could also be imagined. In this section, we examine these issues in detail to explore whether or not they may turn out to be practical. We begin by considering volatile release by heating bulk soils.

A. Thermal Extraction Schemes

Most studies of the resource potential of solar-wind gases assume that gas will be extracted by simple heating, although the exact mechanism, such as microwaves or concentrated solar energy, varies from study to study (see, e.g., Houdashelt et al. 1989). This technique clearly will work, because virtually all the solar-wind-derived gas is released by a temperature of about 700°C (see Figs. 9 and 10). However, whether or not the outcome is desirable for volatile extraction will depend to some extent on the chemical speciation of the volatiles released from the heated soil and how they can be separated from

one another. The physical properties of the evolved volatiles such as their condensation temperatures, thermal conductivities, viscosities, and so on may also have an influence on engineering designs for an extraction facility.

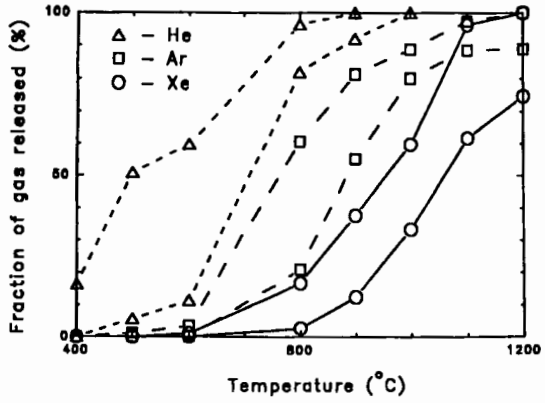


Figure 9. Release patterns for He, Ar and Xe in lunar soils heated to progressively higher temperatures (stepwise heating). Data are from Basford et al. (1973), Hohenberg et al. (1970), Pepin et al. (1970) and Srinivasan et al. (1972). Note the progressively higher release temperatures in going from He to Ar to Xe.

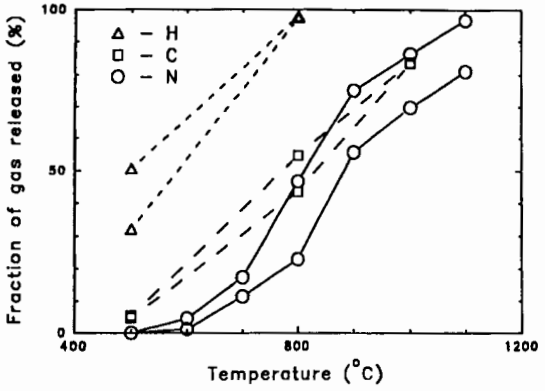


Figure 10. The release patterns for the chemically reactive volatiles H, C and N for lunar soils heated to progressively higher temperatures under vacuum. Data are from Becker (1980), Becker and Clayton (1977,1978), Norris et al. (1983) and Thiemens and Clayton (1980). Note the significantly higher temperatures required for quantitative release of C and N relative to hydrogen.

Fortunately, the analytical literature is a good guide to what to expect in this regard. Epstein and Taylor (1970) found that ~20 to 33% of the

total hydrogen in lunar soils was evolved as H_2O. However, the observed D/H ratio and oxygen isotopic composition suggests that most, if not all, of the water released from lunar soils is probably terrestrial contamination (see, e.g., Epstein and Taylor 1970,1971,1972,1973,1974,1975; Merlivat et al. 1972,1974,1976; Stievenard et al. 1990). Thus, hydrogen will probably be evolved dominantly as H_2.

The situation is a little more complex for carbon. Acid hydrolysis of lunar samples decomposes cohenite [ideally $(Fe,Ni)_3C$], which is the major carbide present in lunar samples (see, e.g., Goldstein et al. 1976), and yields CH_4 along with lesser amounts of higher hydrocarbons (see, e.g., Cadogan et al. 1972; Chang et al. 1970). Pyrolysis and inorganic gas release studies yield a mixture of $CO + CO_2 + CH_4$ (see, e.g., Chang et al. 1974*a*; DesMarais et al. 1973; Gibson and Moore 1972). Studies of the gases released from lunar rocks and soils upon crushing show that $CO > CO_2 \sim CH_4$ in lunar basalts (Gibson and Andrawes 1978*a*).

Nitrogen is released as N_2. Ernsberger (1972) reports that laser heating of individual soil particles yields approximately equal amounts of N_2 and CH_4. Pyrolysis studies of lunar soils commonly yield N_2, although NH_3 has also been reported (see, e.g., Simoneit et al. 1972). The NH_3/N_2 ratio was much less than one, and the NH_3 may be a contaminant. Crushing experiments on lunar rocks and soils release N_2 only (Gibson and Andrawes 1978*a*). Inorganic gas release studies also show only N_2 (Gibson and Moore 1972).

Sulfur is released primarily as H_2S and SO_2. Hydrolysis of lunar soils commonly gives the former, while pyrolysis commonly gives the latter (see, e.g., Chang et al. 1974*a*, *b*; DesMarais 1978). The inorganic gas release study of Gibson and Moore (1972) reported both H_2S and SO_2.

Finally, apparently no studies have been done on the speciation of F and Cl released from lunar soils upon heating or upon the release profiles. These elements will readily combine with hydrogen to form HF and HCl upon heating. The mean H, S, F and Cl contents in lunar soils correspond to an atomic ratio $H/(S + F + Cl) \sim 1.7$, which implies that all F and Cl will be released as HF and HCl, respectively.

B. Grain-Size Sorting

The hydrogen content of lunar soils as a function of grain size has been studied by DesMarais et al. (1974) and Gibson and collaborators (Bustin et al. 1984,1986; Gibson et al. 1987). Related studies of hydrogen distribution with depth in lunar soil grains have been done using F_2 etching (Becker 1980; Epstein and Taylor 1975) and nuclear techniques (Leich et al. 1973,1974). These studies generally indicate that hydrogen is indeed concentrated near grain surfaces and in finer grain-size fractions of lunar soils, but the interpretation of the results is not totally unambiguous (cf. Becker 1980).

DesMarais et al. (1974) studied the hydrogen content of grain-size fractions of five Apollo 15 and 16 soils. Not surprisingly they found more hydrogen in the finer grain-size fractions. DesMarais et al. (1974) resolved

both surface- and volume-correlated hydrogen components in their samples and claimed that the soils were not saturated with respect to hydrogen. Instead, in their view, the hydrogen content of lunar soils is controlled by a kinetic steady state in which the hydrogen implanted by the solar wind is balanced by hydrogen lost by diffusive escape and sputtering. This lack of saturation for hydrogen is similar to the lack of saturation which is observed for helium.

DesMarais et al. (1974) did not present data for the weights of the different size fractions and a complete mass balance relating the hydrogen contents of the different size fractions to the hydrogen content of the bulk sample is not possible for their samples. However, this was done by Gibson and colleagues (Bustin et al. 1984,1986; Gibson et al. 1987). Their data for the hydrogen content of different grain size fractions of five lunar soils are presented in Table XXII. The replicate analyses reported in their abstracts have been averaged together and the 1σ uncertainties were calculated using small number statistics (Wilson 1952). The bulk soil hydrogen contents are taken from our compilation of literature data.

TABLE XXII

Hydrogen Content of Lunar Soils as a Function of Grain Size[a]

Grain-Size Fraction (μm)	Hydrogen Abundance $\mu g\ g^{-1}$ ($\pm 1\sigma$)				
	10084	12070	15021	60501	71501
<20	133 ± 25	108 ± 0.5	148 ± 22	134 ± 18	154 ± 50
20–45	32 ± 4	25 ± 10	54 ± 6	46 ± 6	59 ± 21
45–75	29 ± 8	10 ± 11	25 ± 4	27 ± 20	39 ± 37
75–90	25 ± 8	7 ± 4	21 ± 2	14 ± 2	25 ± 31
90–150	18 ± 4	5 ± 6	18 ± 5	9 ± 0.5	6 ± 3
150–250	22 ± 18	5 ± 5	8 ± 5	4 ± 2	2 ± 0
250–500	17 ± 2	6 ± 6	13 ± 7	4 ± 0.4	11 ± 15
500–1000	5 ± 5	6 ± 5	9 ± 3	4 ± 3	3 ± 2
Weighted Sum	52 ± 7	32 ± 3	56 ± 5	47 ± 5	47 ± 11
Bulk Soil[b]	53 ± 11	33 ± 6	54 ± 12	34 ± 6	23 ± 5

[a] Mean $\pm 1\sigma$ of values from Bustin et al. (1984,1986) and Gibson et al. (1987).
[b] Mean $\pm 1\sigma$ of literature values for the hydrogen content of the bulk soil.

A comparison of the weighted sum of the hydrogen contents of the different grain-size fractions with the literature data on bulk soil hydrogen contents shows generally good agreement for soils 10084, 12070 and 15021. In these three cases the calculated and observed hydrogen contents are almost identical and always agree within the combined 1σ uncertainties. However, this is not the case for soils 60501 and 71501 where the weighted sums are considerably larger than the observed bulk values. This could be explained by contamination from terrestrial water vapor on the finest grain size fractions. This would then lead to higher apparent hydrogen contents in these fractions and to a weighted sum larger than the bulk soil value. However, the hydrogen

contents of the <20 μm fractions of all five soils are similar and in any case it is not clear why only two of the five samples would be affected by terrestrial water vapor.

Figure 11 presents the same data in graphical form. In this case, the percentage of the total soil mass in each grain-size fraction is compared to the percentage of total hydrogen in each grain-size fraction. If we consider only the three samples which display mass balance, 61 to 74% of the total hydrogen is in the sub 20 μm grains, and 78 to 87% of the total hydrogen is in the sub 45 μm grains. However, 22 to 26% of the total mass resides in the sub 20 μm grains and 40 to 46% of the total mass is in the sub 45 μm grains. The corresponding hydrogen enrichment factors for the sub 45 μm fractions of the three soils are 1.8 (10084), 2.2 (12070), and 1.8 (15021). In other words, the hydrogen enrichment in the fine-grained fractions is only about a factor of 2. Given this small enrichment, the economics of grain-size beneficiation may not be viable.

Likewise, Fig. 12 illustrates similar data for ^{4}He. These data, which are from R. O. Pepin (personal communication, 1991; Pepin et al. 1975) show similar trends for the He enrichment factor. Specifically, for the sub 37 μm fractions, the He enrichments are 2.1 (10084), 2.0 (12033), 2.9 (15531), 2.9 (67701) and 2.3 (70181). More results for He are shown in Table XXIII, where we summarize He results from three laboratories that analyzed both bulk samples and samples with grain size less than about 20 μm from the same soil for five or more soils. Again, the results are consistent with average enhancement factors of about 2. As the <20 μm fraction typically contains about 20% of the mass of the soil (Butler et al. 1973,1974), this means that such a grain-size separation would lose more than half the He. Whether this would be useful would depend on the detailed economics of the mining, sorting and extraction systems.

TABLE XXIII

He Enrichment in Fine-Grained Soil Particles

Laboratory	$He_{<20\mu m}/He_{bulk}$	No. of Samples	Notes[a]
Bogard	1.84 ± 0.51	10	1
Hintenberger	1.56 ± 0.45	11	2
Kirsten	2.39 ± 0.52	5	3

[a] Notes: (1) Data from Bogard et al. (1973,1974) and Bogard and Hirsch (1975). (2) Data from Hintenberger et al. (1970,1971,1974,1975) and Hintenberger and Weber (1973). One sample <30 μm and one sample <25 μm are included. (3) Data from Kirsten et al. (1972,1973).

Finally we note that because the solar-wind-derived volatiles are located near the surfaces of grains, a technique that released the gas in the outer 1 μm of every grain would extract virtually all of the gas (a small amount of gas does migrate to the interior of grains). In the laboratory, this is often done

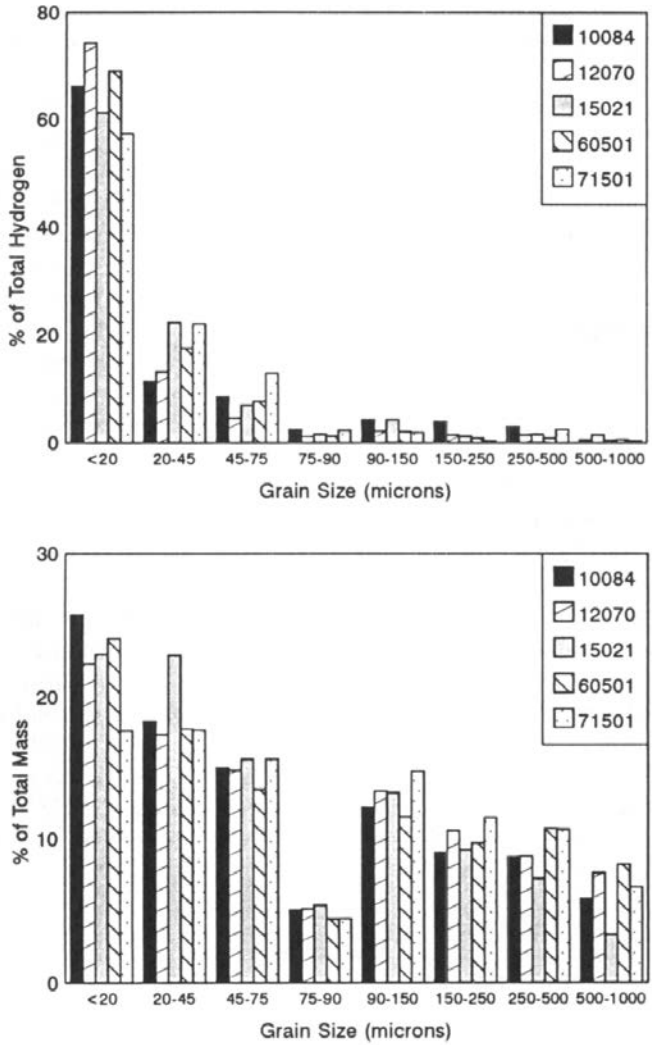


Figure 11. (top) The percentage (by mass) of total hydrogen found in different grain-size fractions of lunar soils from the five Apollo landing sites. (bottom) The percentage of the total mass in each grain-size fraction of the same soils. Note that the smallest grain-size fraction in this figure approximately corresponds to the sum of the two smallest size fractions in Fig. 12. The hydrogen content and mass as a function of grain size are from Bustin et al. (1984,1986) and Gibson et al. (1987).

by etching with mineral acids or F_2. While that might not be practical on the Moon, it might be possible to achieve similar results through the mutual abrasion of agitated particles. Such a scheme has not been investigated in detail.

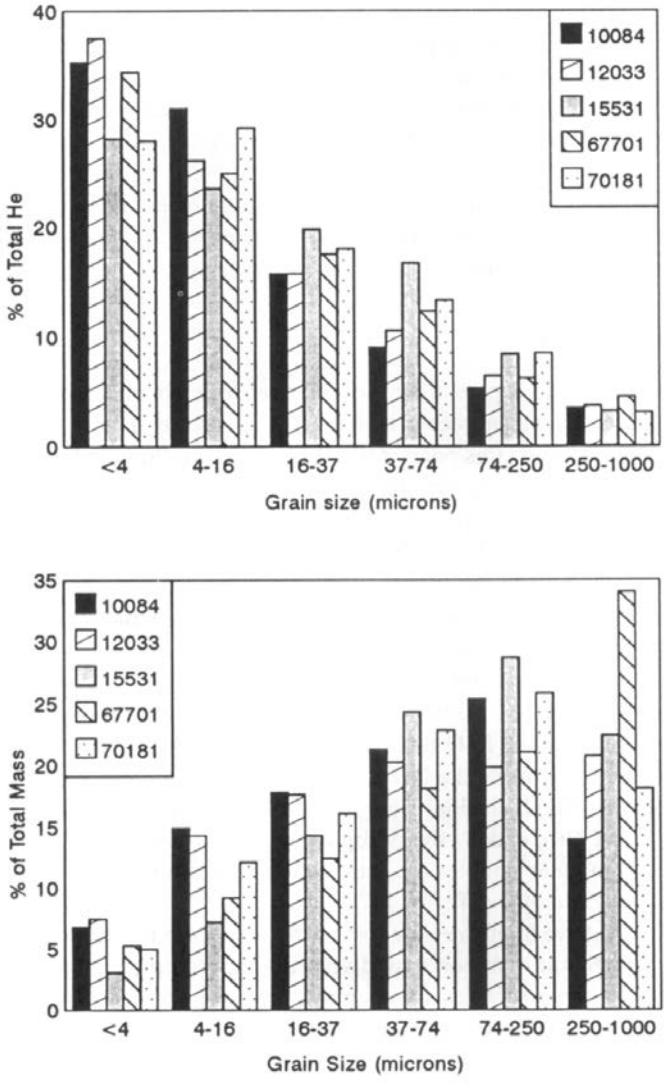


Figure 12. (top) The percentage (by mass) of total ^{4}He in different grain-size fractions of lunar soils from the five Apollo landing sites. (bottom) The percentage of the total mass in each grain-size fraction of the same soils. Note that the two smallest size fractions in this figure approximately correspond to the smallest size fraction in Fig. 11. The ^{4}He content and mass as a function of grain size are from R. O. Pepin (personal communication, 1991) and Pepin et al. (1975).

C. Ilmenite Separation

Beneficiation by ilmenite separation seems plausible for He extraction, because He is much more abundant (sometimes by more than a factor of 100) in ilmenite than in other lunar materials. However, separation of ilmenite from soil is apparently extremely difficult (see, e.g., Heiken and Vaniman 1990), and even if it could be achieved, such a separation would miss the common He- and Ti-rich compound particles, such as agglutinates and breccias (whose high He and Ti contents presumably reflect ilmenite grains that became a part of the compound particles).

V. LUNAR WATER

One of the first, and apparently one of the most firm, conclusions drawn from examination of lunar samples is their anhydrous nature. The virtually unanimous agreement upon this point is illustrated by the following list taken from Charles et al. (1971): (1) searches for hydrous material in lunar rocks have been singularly unsuccessful, with the few exceptions to the contrary being discussed in Sec. V(B) below; (2) spectral studies by Adams and McCord (1970), which are cited in Charles et al. (1971), reveal that lunar materials completely lack bands due to OH^- and H_2O; (3) electron microprobe analyses of lunar apatites [ideally $Ca_5(PO_4)_3(F,Cl,OH)$] show F and Cl in amounts high enough to indicate that OH is either absent or minor (also see Frondel 1975, and references therein); (4) major element analyses of the Apollo 11 samples revealed no detectable H_2O (see, e.g., Engel and Engel 1970); (5) the high Fe^{2+}/Fe^{3+} ratios indicate the lack of a suitable oxidizing agent such as H_2O; (6) no aqueous fluid inclusions were found in the Apollo 11 samples (Roedder and Weiblen 1970); and (7) the isotopic composition of the trace amounts of water found in lunar samples are consistent with contamination by terrestrial water vapor (Epstein and Taylor 1970).

In Sec. V.A we briefly review some of this evidence in the context of lunar resource utilization. We begin by discussing the isotopic evidence showing that water in lunar samples may simply be due to terrestrial water vapor contamination and then describe the small amount of mineralogical evidence which implies the presence of water in the lunar interior at some point in the past.

A. Isotopic Evidence

The H_2 and H_2O contents and isotopic compositions of lunar samples have been studied by three groups: (1) Epstein and Taylor (1970,1971,1972,1973, 1974,1975); (2) Friedman and colleagues (Friedman et al. 1970,1971,1972); and (3) by French workers (Merlivat et al. 1972,1974,1976; Stievenard et al. 1990). Their data for H_2 contents are presented in Tables III–V and their data for water contents and isotopic composition are summarized in Table XXIV.

The highest water contents, about 14 μmoles g^{-1} are found in lunar breccias and soils, while lower contents of about 2 μmoles g^{-1} are found in

lunar basalts. The mean contents of all three groups overlap due to the large uncertainties on the means for breccias and soils, but it is clear that basalts contain the least water. However, the isotopic results suggest that even this small amount of water is dominated by terrestrial contamination.

The observed D/H ratios in the samples analyzed span a wide range from 21 ppm in breccia 10060 ($\delta D = -865$ per mil) to 150 ppm ($\delta D = -37$ per mil) in basalt 15555. However, the mean D/H values observed in the basalts, breccias, and soils only range from 85 to 136 ppm and all three means overlap within the 1σ uncertainties (see Table XXIV).

As stressed by Epstein and Taylor in their papers, the isotopic composition of water extracted from lunar samples is generally indistinguishable from that of terrestrial water. This point is vividly illustrated in Fig. 13. This figure shows that the D/H ratio, and where determined the $^{16}O/^{18}O$ ratio, in lunar samples is the same as that of atmospheric water vapor in Pasadena, Calif. Given this similarity, both the Caltech group and the French workers have concluded that there is no unambiguous evidence that the water extracted from lunar samples originated on the Moon instead of being merely terrestrial contamination. (But see the papers by Friedman and colleagues for an opposing viewpoint.)

However, as noted recently by Hodges (1991), assuming that the water produced by the solar-wind reduction of Fe^{2+} to Fe metal is retained by the Moon, then the mean water content of lunar soils should be about 15 μmoles g^{-1}, or about the same as that observed for lunar soils and breccias. If the probable water input due to cometary and meteoritic impacts is also considered, then this figure becomes about 30 μmoles g^{-1}. Thus, the observed water contents are not inconsistent with independent theoretical predictions.

B. Mineralogical Evidence for Water

The mineralogical evidence for water falls into two categories: (1) amphiboles and micas found in lunar samples; and (2) the presence of β-FeOOH, the mineral akaganéite, in the rusty rock 66095 and various other lunar samples. We consider the occurrences of amphiboles and micas first and rely on Frondel (1975) who summarizes the relevant observations.

Amphiboles and Micas. Gay et al. (1970) found an amphibole in a vug in a sample of 10058, a medium-grained basalt. The electron microprobe analysis corresponds to a low-alumina Na-amphibole known as magnesioarfvedsonite with the general formula $(Na,K,Ca)_3\ (Mg,Mn,Fe)_3\ ([OH]?,F)_2$. The amphibole contains 1.2±0.3 mass % F and gave a low total of 97.1%. The X-ray diffraction data for the crystal are also consistent with the identification as magnesioarfvedsonite. Charles et al. (1971) found loose grains of an amphibole in a package with a fragment of the porphyritic basalt 12021. The electron microprobe analysis shows low amounts of F (0.4%) and Cl (0.2%) and a low total of 97.5%. The microprobe analysis is consistent with an aluminotschermakite which has the general formula $(Ca,Na,K)_{2.5}\ (Mg,Fe,Mn)_4$

TABLE XXIV

Water Content and Isotopic Composition in Lunar Samples

Sample Number	Sample Type	Water μmoles g^{-1}	D/H ppm	δD^a (per mil)	Reference
12051	Basalt	0.45	142	−88	Friedman et al. 1971
15555	Basalt	0.49	150	−37	Friedman et al. 1972
70215	Basalt	2.7	130	−165	Merlivat et al. 1974
70215[b]	Basalt	3.9±2.7	135±9	−133±58	Merlivat et al. 1976
70215[c]	Basalt	3.3	140	~ −100	Stievenard et al. 1990
75035	Basalt	2.48	141.4	−92	Merlivat et al. 1974
75035	Basalt	2.8	116.5	−252	Merlivat et al. 1974
Mean		2.3±1.3	136±11	−127±71	
10046	Breccia	25.3	38	−756	Friedman et al. 1970
10046	Breccia	20.6	66	−576	Friedman et al. 1970
10060	Breccia	8.4	21	−865	Friedman et al. 1970
10061	Breccia	8.9	92	−409	Epstein and Taylor 1970
14305	Breccia	8.8	144	−76	Friedman et al. 1972
15299	Breccia	12.7	147.7	−52	Merlivat et al. 1974
Mean		14.1±7.2	85±53	−454±340	
10084	Soil	11.07	116	−255	Epstein and Taylor 1970
10084	Soil	11.58	114	−268	Epstein and Taylor 1970
12030	Soil	12.4	123	−210	Friedman et al. 1971
12033	Soil	7.5	142	−88	Epstein and Taylor 1971
12042	Soil	13.5	134	−140	Epstein and Taylor 1971
12070	Soil	11.8	120	−230	Epstein and Taylor 1971

14240	Soil	17.4	121	−223	Epstein and Taylor 1972
14422	Soil	10.4	104	−332	Epstein and Taylor 1972
15001	Soil	25.1	132	−152	Epstein and Taylor 1975
15004	Soil	25.1	126	−191	Epstein and Taylor 1975
15006	Soil	30.9	125	−198	Epstein and Taylor 1975
15021[d]	Soil	15.1	127	−185	Epstein and Taylor 1973
15100	Soil	23.7	44.6	−714	Friedman et al. 1972
15301	Soil	7.8	121	−223	Epstein and Taylor 1972
15600	Soil	2.6	146.5	−59	Merlivat et al. 1974
15600	Soil	3.8	140.9	−95	Merlivat et al. 1974
62221[d]	Soil	18.5	130	−165	Epstein and Taylor 1973
64421	Soil	15.6	134	−140	Epstein and Taylor 1973
65513[e]	Soil	~20	—	—	Epstein and Taylor 1973
66095	Soil	40.0	142	−88	Epstein and Taylor 1974
68501	Soil	5.2	137.9	−115	Merlivat et al. 1974
72501	Soil	5.0	142	−88	Merlivat et al. 1974
72501	Soil	4.6	143	−82	Merlivat et al. 1974
74220	Soil	4.6	139	−108	Epstein and Taylor 1973
74240	Soil	11.0	117	−248	Epstein and Taylor 1975
78501	Soil	5.8	144.5	−72	Merlivat et al. 1974
Mean		13.8±9.2	127±20	−185±129	

[a] The δD values were all recalculated using D/H = 155.76 ppm for Standard Mean Ocean Water (SMOW) as done by Merlivat et al. (1974,1976). [b] Mean of eight values determined for 70215 (Merlivat et al. 1976). [c] Calculated from data of Stievenard et al. (1990) by mass balance. [d] Approximate values for D/H and δD because Epstein and Taylor (1973) combined the high temperature fractions of H_2 and H_2O for 15021 and 62221. [e] The high temperature H_2O fraction for 65513 was lost by Epstein and Taylor (1973) and no δD value was given.

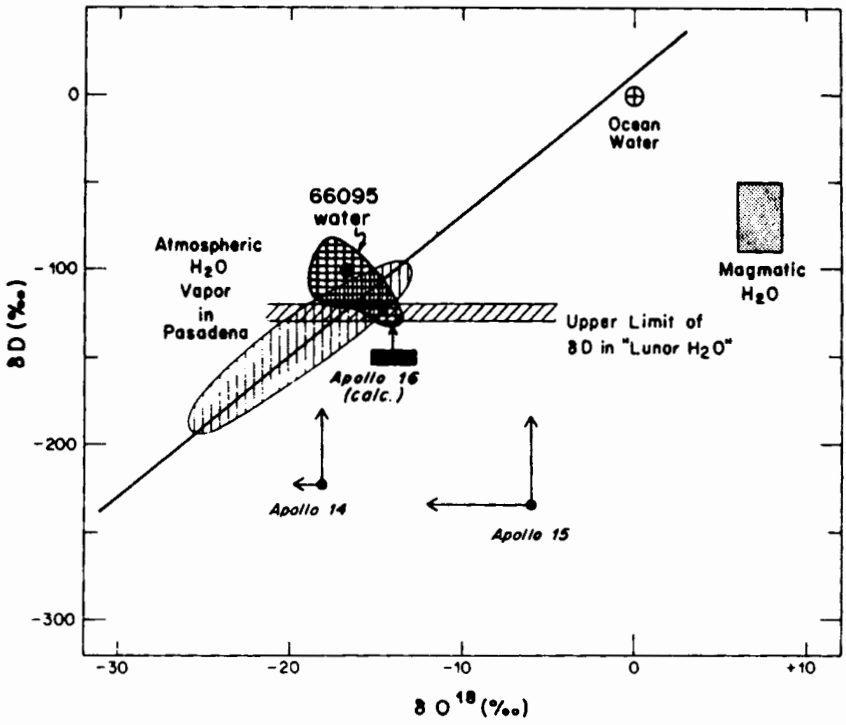


Figure 13. A comparison of the hydrogen and oxygen isotopic composition of water released from lunar soils with the H and O isotopic composition of terrestrial water. The Apollo samples plotted are 14321, 15301, 64221 and 66095. The O and H isotopic composition of these samples was determined by Epstein and Taylor (1973,1974). The cross-hatched horizontal band is their estimate for the upper limit of D/H in "lunar" water. The diagonal line is the locus of terrestrial meteoric waters. The arrows show the directions in which the isotopic data should be shifted to account for O and H isotopic exchange during sample processing. The isotopic similarity of "lunar" water to terrestrial water suggests that all water recovered from lunar samples is a terrestrial contaminant (figure taken from Epstein and Taylor 1974).

$Al_{1.5}([OH]?,F,Cl)_2Al_2Si_6O_{22}$.

Mason et al. (1972) reported a few grains of hornblende in a thin section of breccia 14319,13 and in fines from 14163. The electron microprobe analysis corresponds to a titanian pargasite or a kaersutite with the general formula $(Ca_{1.88}\ Na_{0.79}\ K_{0.35})\ (Mg_{1.84}Mn_{0.05}Fe_{2.47}\ Al_{0.18}\ Ti_{0.45})Al_2Si_6O_{23}$ and the total is 100.5%. In this case, the microprobe total does not show any evidence for a mass deficit which could be attributed to OH or water, but hornblendes are water-bearing minerals (see, e.g., Deer et al. 1963).

Gay et al. (1970) also reported a few small crystals of mica in fines from 10084. The electron microprobe analysis corresponds to that of a biotite

[general formula K_2 $(OH,F)_4$ $(Mg,Fe,Al)_8$ $(Si,Al)_8$ O_{20}] and has a low total of 94.9%. The X-ray diffraction pattern is also consistent with a mica. However, as Frondel (1975) notes, the possibility that this material is a contaminant cannot be ruled out.

Drever et al. (1971) also reported flakes of an unknown mica in fines from 10084. In this case, only a qualitative probe analysis was possible because of the small size of the flakes. The X-ray diffraction pattern is consistent with various dioctahedral phyllosilicates such as muscovite or kaolinite. Drever et al. (1971) also reported rare μm-sized flakes of a mica on the surface of 10017, a fine-grained basalt. The *b* parameter in an X-ray pattern was consistent with biotite, chlorite, or a serpentine.

Although the identification of amphiboles in trace amounts seems firm, the inference that these minerals contain OH rests solely on mass deficits in microprobe analyses. Furthermore, the presence of trace amounts of micas is less firm and may be due to contamination. Thus, the available evidence by itself is not a compelling argument for water in the lunar interior.

"Rust" on Lunar Rocks. Mineralogical examination of Apollo 14 samples showed the presence of rust-like alteration on samples from every station and on breccia 66095 in particular (see, e.g., El Goresy et al. 1973; Taylor et al. 1973). This alteration product was originally believed to be goethite α-FeOOH (see, e.g., El Goresy et al. 1973), but was later shown to be akaganéite β-FeOOH by Taylor et al. (1974). This phase was interpreted by some as being lunar in origin (see, e.g., El Goresy et al. 1973), but by others as being merely a reaction product between lawrencite $FeCl_2$ and terrestrial water vapor, either in the Apollo module or on the Earth (see, e.g., Taylor et al. 1974).

The evidence in favor of an origin by reaction with terrestrial water vapor, either in the spacecraft, or upon return to Earth, is as follows. Taylor et al. (1974) showed that lawrencite, the probable precursor, is rapidly oxidized and hydrated to form akaganéite at 25° C and 40% relative humidity. Epstein and Taylor (1974) showed that the O and H isotopic composition of the akaganéite in breccia 66095 is identical to that of atmospheric water vapor (e.g., in Pasadena, Calif.) and that the amounts of water released and the isotopic composition of the water released from a terrestrial goethite and a synthetic akaganéite are (with the exception of the $\delta^{18}O$ values in the synthetic sample) also virtually identical to that of the lunar sample. Taken together, these two studies considerably weaken the case for a lunar origin of the akaganéite in 66095 and strongly suggest a terrestrial origin.

VI. LUNAR VOLCANIC GASES

Evidence for lunar volcanic gases comes from the presence of vesicles and vugs in lunar rock samples (e.g., LSPET 1969; Taylor et al. 1991) and from the volatile-rich surface coatings on the Apollo 15 green and Apollo 17 orange glasses (see, e.g., Meyer et al. 1975). Detailed examination of these surface coatings (see, e.g., Butler and Meyer 1976; Chou et al. 1975; Goldberg et

al. 1976; Wasson et al. 1976) shows that the coatings are probably condensates from a volcanic gas. The chemistry of this gas phase is of interest for several reasons: (1) to place constraints on the water abundance in the lunar interior at the time that volcanism occurred; (2) to identify the major gaseous species of volatile and ore-forming metals; and (3) to model the transport mechanisms which may have fractionated and concentrated metals into lunar "ore deposits." Fegley (1991,1992) used mass balance arguments and thermodynamic calculations to constrain the chemistry of lunar volcanic gases. He modeled lunar volcanic gases as modifications of terrestrial volcanic gases with two important differences. First, lunar volcanic gases were probably more reducing than terrestrial volcanic gases. Oxygen fugacity measurements on lunar samples (see, e.g., Sato 1976,1979; Sato et al. 1973) show that lunar rocks are generally about 4 orders of magnitude more reduced than terrestrial volcanic rocks and lie at or below the iron-wüstite (IW) buffer (e.g., see Fig. 14). Previous theoretical models of lunar volcanic gases have also assumed that they are more reducing than their terrestrial counterparts (see, e.g., Naughton et al. 1972; Wellman 1970). Second, the absence of indigenous water (and thus hydrogen) on the Moon means that lunar volcanic gases will be much drier than terrestrial volcanic gases which are generally rich in steam. As reviewed above in Secs. III.B and V, there is no firm evidence for indigenous water in lunar samples and the amounts of H_2 in lunar basalts is $\leq 0.5\,\mu$moles g^{-1} . In the case of the orange soil 74220, the H_2 contents reported by Epstein and Taylor (1973) and Chang et al. (1974*b*) are 0.1 to 0.3 μmoles g^{-1} . Epstein and Taylor (1973) also found 4.6 μmoles g^{-1} H_2O with a $\delta D = -117$ per mil in 74220, but the isotopic composition is consistent with that of atmospheric water vapor in Pasadena, Calif. It is therefore most plausibly explained by contamination.

Fegley (1991,1992) noted that these two factors will lead to first-order differences between the chemistry of lunar and terrestrial volcanic gases. He also pointed out that changes in the ratios of sulfur and the halogens, although also important, will be less significant for the overall chemistry of the gas.

A representative model composition studied by Fegley (1991,1992) is displayed in Table XXV. This was derived by starting with a measured composition for a terrestrial volcanic gas from Hawaii and removing water (i.e., removing H and the corresponding stoichiometric amount of O needed for water). The oxidation state was also adjusted by setting the O/C atomic ratio to 1.2. The Br and I abundances were specified by assuming CI chondritic Cl/Br and Cl/I ratios. Ideal gas chemical equilibrium calculations were then done from 500 to 2000 K at pressures from 0.01 to 100 bar. This pressure range was chosen because it is appropriate for volcanic gases originating at different depths within the Moon (see, e.g., Wellman 1970). The thermodynamic data used in the calculations were taken from standard compilations such as the JANAF Tables (Chase et al. 1985) and Gurvich et al. (1989).

In addition to considering the chemistry of the elements listed in Table XXV, the major rock-forming elements (Ca, Mg, Si, Al, Ti, Fe), volatile

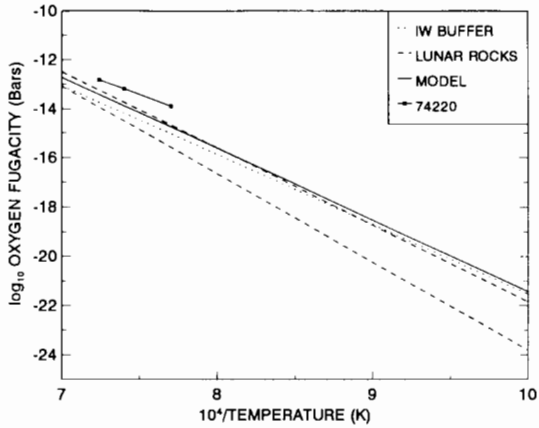


Figure 14. The measured oxygen fugacities as a function of temperature for lunar samples are compared with the iron-wüstite (IW) buffer and with the oxygen fugacities calculated for the lunar volcanic gas model of Fegley (1992). The oxygen fugacity of the orange soil 74220 is plotted separately because it is higher (e.g., more oxidizing) than most lunar samples (figure modified from Fegley 1992).

elements (P, Na, K, Zn, Cu), and ore-forming metals (Ni, Cr, Mn, Co, V) were also included in the calculations. Although these elements are of interest, these abundances in lunar volcanic gases are not at all well constrained. However, studies of terrestrial volcanic gases and sublimates from these gases (see, e.g., Symonds et al. 1987; White and Waring 1963) show that elements other than those listed in Table XXV generally have negligible abundances relative to those of the major constituents of terrestrial volcanic gases. Furthermore, as noted by Fegley (1991), theoretical models of the vaporization of the rock-forming elements from silicate magmas show negligible vapor pressures for compounds formed by these elements (Fegley and Cameron 1987). The more volatile elements (P, Na, K, Zn, Cu) will have higher vapor pressures, but even so the abundances of their gases over silicate magmas are expected to be less than those of gases formed by S, F, Cl and C. Therefore, the chemistry of the additional elements considered was calculated by giving them low abundances, for example, mole fractions of 10^{-9} or less. As a result, the chemistry of these elements can be calculated without affecting the overall mass balance of the system.

Some of the results of the calculations are displayed in Figs. 14–18. Figure 14 shows that the calculated oxygen fugacity (f_{O_2}) of the model lunar volcanic gas is basically the same as that as IW and is a good match to the measured f_{O_2} values for lunar samples. Figures 15–17 show the major Cl, F and S gases as a function of temperature at a total pressure of 1 bar. In contrast to terrestrial volcanic gases where steam dominates, the calculated H_2O abundance is totally insignificant and is not graphed because its abundance is below 0.1 ppb by volume. The low abundance of water vapor is a result of

TABLE XXV
Lunar Volcanic Gas Model

Element	Abundance (%)
H	0.1
O	46.0
C	38.4
Cl	8.5
S	5.1
F	1.7
N	0.2
Br	190 ppm
I	14 ppm

removing H from the system. However, an important consequence of the low H elemental abundance is that the most abundant H-bearing gases in this case, and in other plausible models of lunar volcanic gases, are HCl and HF. As shown in Fig. 17, H_2S is also totally insignificant and is not graphed because its abundance is also below 0.1 ppb by volume.

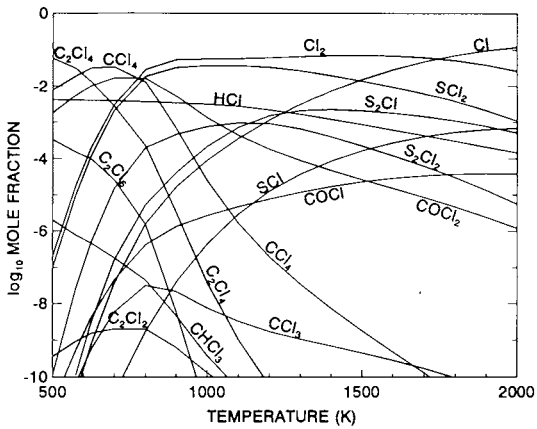


Figure 15. The calculated abundances of the major Cl-bearing compounds in the lunar volcanic gas model of Fegley (1992) are plotted as a function of temperature for a total pressure of 1 bar. Note that the chlorofluorocarbon (CFC) gases constitute a significant fraction of total Cl compounds. Also note that several other gases such as Cl_2, CCl_4 and C_2Cl_4 are generally more abundant than HCl over the entire temperature range considered (figure modified from Fegley [1992] with the inclusion of more Cl-bearing gases in the thermodynamic database).

Although HCl and HF are the two major H-bearing gases they are not the major Cl- and F-bearing gases. This is because the H/Cl and H/F atomic ratios of ~0.01 and 0.06, respectively, are too low to bind chemically all Cl and F in the form of hydrogen halides. Instead, Cl and F form other gases

such as Cl_2, CCl_4, C_2Cl_4, COF_2, COClF, CF_4 and the chlorofluorocarbons $CClF_3$, CCl_2F_2, and CCl_3F. The results in Fig. 15 show that the major Cl gases are Cl and Cl_2 (at $T \geq 800$ K) and CCl_4 and C_2Cl_4 (at $T < 800$ K). The results in Fig. 16 show that the major F gases are COF_2 (above 900 K), and COClF and CF_4 at lower temperatures. However, the chlorofluorocarbon gases together make up about 10% of the total Cl abundance and several tens of % of the total fluorine abundance. This is a dramatic change from terrestrial volcanic gases where calculations show that chlorofluorocarbons typically have volume mixing ratios in the range of 10^{-28} to 10^{-42} over the same temperature range (500 to 2000 K).

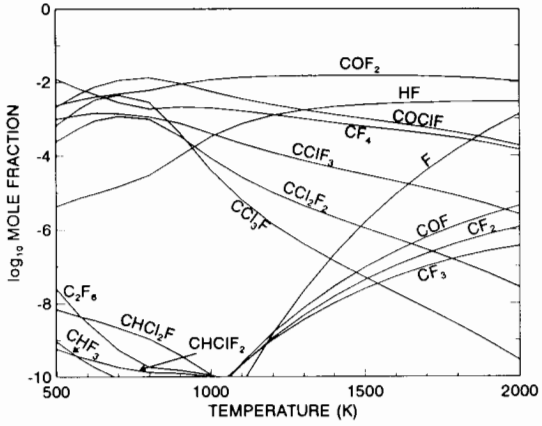


Figure 16. The calculated abundances of the major F-bearing compounds in the lunar volcanic gas model of Fegley (1992) are plotted as a function of temperature for a total pressure of 1 bar. Note that the chlorofluorocarbon (CFC) gases constitute a significant fraction of total F compounds. Also note that several other gases such as COF_2, COClF, CF_4 and CCl_3F are generally more abundant than HF over the entire temperature range considered (figure modified from Fegley [1992] with the inclusion of more F-bearing gases in the thermodynamic database).

The qualitatively different halogen chemistry predicted for lunar volcanic gases can be understood by considering how halogen chemistry depends on the abundances of H, Cl and F. Volcanic gases with the bulk abundance ratio $H/(F + Cl) > 1$ will have HF and HCl as the dominant halogen gases, as in terrestrial volcanic gases. However, volcanic gases with $H/(F + Cl) < 1$ will have HF and HCl as the dominant H gases, but not as the dominant halogen gases. In the intermediate case, volcanic gases with $H/(F + Cl) = 1$ will display more complex chemistry involving HF, HCl and other halogen gases.

These considerations are of interest for lunar resources because the predicted large abundances of the chlorofluorocarbons are a direct consequence of the anhydrous and reducing conditions assumed to prevail in the lunar interior. The chlorofluorocarbons can thus be viewed as probes of conditions

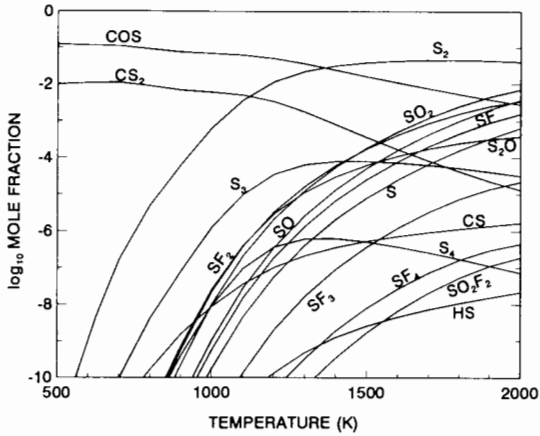


Figure 17. The calculated abundances of the major S-bearing gases in the same model are plotted as a function of temperature for a total pressure of 1 bar. Note the absence of H_2S, which has an insignificant abundance at all temperatures considered (figure from Fegley 1992).

in the lunar interior, and if any remained trapped in lunar glasses or rocks, their abundances in principle provide important constraints on compositional conditions at the time the gases were generated.

Finally, Fig. 18 illustrates the chemistry of two of the metals, Cu and Fe, which were also included in the thermodynamic calculations. Fegley (1991) showed that with very few exceptions the chlorides and fluorides were the major species for the rock-forming elements, volatile elements, and ore-forming metals included in his calculations. Similar results were reported later by Colson (1992). The results in this figure illustrate the same point. Copper chemistry in lunar volcanic gases is predicted to be dominated by CuCl over a wide temperature range, down to about 800 K where the trimer $(CuCl)_3$ takes over. Unpublished calculations by Fegley (1992) also predict similar behavior for Cu in terrestrial volcanic gases. These predictions are in accord with the observations by Murata (1960) of CuCl in volcanic gases.

In the case of Fe, two gases FeF_3 and $FeCl_3$ are the dominant Fe gases over the 500 to 2000 K temperature range, with small amounts of $FeCl_2$ also being formed at temperatures below 1000 K. As noted in Fegley (1991) other transition metals display qualitatively similar chemistry.

Finally we note that the results presented in Fig. 18 and the more extensive results discussed in Fegley (1991) and Colson (1992) are complemented by the proposal of Gillett (1990) that chloride magmas may have fractionated and concentrated metals in the lunar interior.

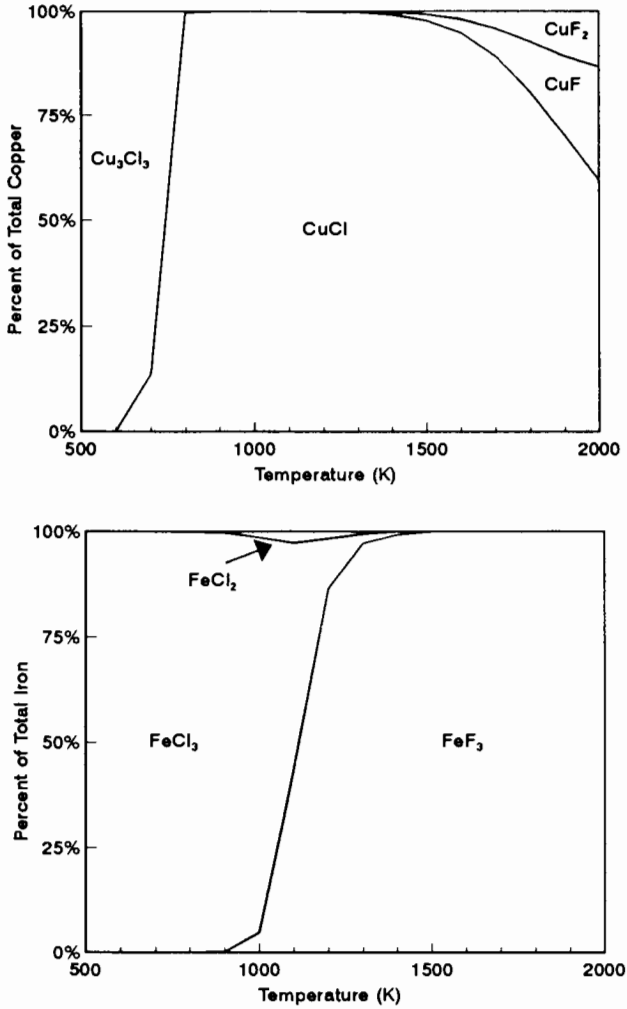


Figure 18. (top) The gas phase chemistry of copper as a function of temperature for the model lunar volcanic gas composition illustrated in Figs. 14–17. Note that in this graph the abundance scale expresses the percent of total Cu which is in each gas. Copper is a relatively volatile ore-forming metal and CuCl(g) has been observed in terrestrial volcanic gases (Murata 1960). (bottom) The same for iron, except that the abundance scale expresses the percent of total Fe which is in each gas. Fe-bearing minerals are found in sublimates collected from volcanic gases and Fe is enriched in the aerosol + vapor of volcanic gases relative to the magma (see, e.g., Symonds et al. 1987).

VII. SUMMARY

In this section we summarize what we regard as the major remaining unanswered questions with regard to the utilization of lunar volatiles as resources. We also present our recommendations for future work specifically directed towards the utilization of lunar volatiles as resources.

A. Remaining Unanswered Questions

The major remaining unanswered questions of interest from a resources perspective include the following:

1. What is the average distribution of solar-wind-implanted volatiles with depth? Different reasonable models lead to order-of-magnitude differences in calculated total inventories, by far the largest uncertainty for most volatiles.
2. Are exposed surfaces of Apollo soils saturated with any or all of the solar-wind-implanted volatiles? If not, limb and farside locations become much more promising for resources.
3. Is there water trapped near the lunar poles? If the amount is significant, this would be by far the most valuable source of H on the Moon.
4. What is the actual abundance of hydrogen and of water in lunar materials? The existing laboratory data for hydrogen indicate a mean abundance of about 50 μg g^{-1} in bulk lunar soils, but this may be high due to contamination by terrestrial water. In many cases, no isotopic data were taken when hydrogen abundances were determined, and thus potentially important information about the degree (if any) of terrestrial contamination is unavailable. In the case of water, the actual water content of lunar materials is still unknown, because all the water detected could be explained by terrestrial contamination.

B. Recommendations for Future Work

The following studies are of interest from a resources perspective:

1. Determination of the abundance of solar-wind-implanted volatiles at depths of >3 m for several sites, either by using drill cores, trenches, or penetrators. As this recommendation implies, either returned samples or *in-situ* observations could be used to get the required data.
2. Determination of the feasibility of various extraction techniques for solar-wind-implanted volatiles. In particular, extraction by acid etching (preferably in a closed system involving recycling of the aqueous phase) or abrasion should be studied from an engineering perspective, and compared with thermal extraction processes. If lunar fluorine extraction schemes are feasible, then etching by F_2 may also be a possibility.
3. Determination of the water abundance (if any) in the lunar polar regions. A polar orbiting mission with a gamma-ray spectrometer would be able

to detect economically important deposits (see, e.g., Feldman et al. 1991; Metzger and Drake 1990).

4. Determination of the actual hydrogen and water contents of different types of lunar materials with an emphasis on lunar soils and breccias. Microanalytical techniques such as the ion microprobe and improved stepwise heating techniques, e.g., as employed by Stievenard et al. (1990), may be very useful for these studies. The combined determination of abundances and isotopic compositions is highly recommended.

REFERENCES

Arnold, J. R. 1975. Monte Carlo simulation of turnover processes in the lunar regolith. *Proc. Lunar Sci. Conf.* 6:2375–2396.

Arnold, J. R. 1976. Water on the moon. In *Lunar Utilization: Abstracts of Papers Presented at a Special Session of the Seventh Annual Lunar Science Conf.* pp. 61–65.

Arnold, J. R. 1979. Ice in the lunar polar regions. *J. Geophys. Res.* 84:5659–5668.

Arnold, J. R. 1987. Ice at the lunar poles revisited. *Lunar Planet. Sci.* XVIII:29–30 (abstract)

Basford, J. R. 1974. K-Ar analysis of Apollo 11 fines 10084. *Proc. Lunar Sci. Conf.* 5:1375–1388.

Basford, J. R., Dragon, J. C., Pepin, R. O., Coscio, M. R., Jr., and Murthy, V. R. 1973. Krypton and xenon in lunar fines. *Proc. Lunar Sci. Conf.* 4:1915–1955.

Becker, R. H. 1980. Light elements in lunar soils revisited: Carbon, nitrogen, hydrogen and helium. *Proc. Lunar Planet. Sci. Conf.* 11:1743–1761.

Becker, R. H., and Clayton, R. N. 1975. Nitrogen abundances and isotopic compositions in lunar samples. *Proc. Lunar Sci. Conf.* 6:2131–2149.

Becker, R. H., and Clayton, R. N. 1977. Nitrogen isotopes in lunar soils as a measure of cosmic-ray exposure and regolith history. *Proc. Lunar Sci. Conf.* 8:3685–3704.

Becker, R. H., and Clayton, R. N. 1978. Nitrogen isotope systematics of two Apollo 12 soils. *Proc. Lunar Planet. Sci. Conf.* 9:1619–1627.

Becker, R. H., and Epstein, S. 1981. Carbon isotopic ratios in some low-δ^{15}N lunar breccias. *Proc. Lunar Planet. Sci.* 12:289–293.

Becker, R. H., Clayton, R. N., and Mayeda, T. K. 1976. Characterization of lunar nitrogen components. *Proc. Lunar Sci. Conf.* 7:441–458.

Bogard, D. D., and Hirsch, W. C. 1975. Noble gas studies on grain size separates of Apollo 15 and 16 deep drill cores. *Proc. Lunar Sci. Conf.* 6:2057–2083.

Bogard, D. D., Nyquist, L. E., Hirsch, W. C., and Moore, R. D. 1973. Trapped solar and cosmogenic noble gas abundances in Apollo 15 and 16 deep drill samples. *Earth Planet. Sci. Lett.* 21:52–69.

Bogard, D. D., Hirsch, W. C., and Nyquist, L. E. 1974. Noble gases in Apollo 17 fines: Mass fractionation effects in trapped Xe and Kr. *Proc. Lunar Sci. Conf.* 5:1975–2003.

Bouchet, M., Kaplan, G., Voudon, A., and Bertoletti, M. J. 1971. Spark mass spectrometric analysis of major and minor elements in six lunar samples. *Proc. Lunar Sci. Conf.* 2:1247–1252.

Burlingame, A. L., Calvin, M., Han, J., Henderson, W., Reed, W., and Simoneit, B. R. 1970. Study of carbon compounds in Apollo 11 lunar samples. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1779–1791.

Bustin, R., Kotra, R. K., Gibson, E. K., Nace, G. A., and McKay, D. S. 1984. Hydrogen abundances in lunar soils. *Lunar Planet. Sci.* XV:112–113 (abstract).

Bustin, R., Carr, R. H., and Gibson, E. K., Jr. 1986. The hydrogen content of lunar soils. *Lunar Planet. Sci.* XVII:95–96 (abstract).

Butler, J. C., and King, E. A., Jr. 1974. Analysis of the grain size-frequency distributions of lunar fines. *Proc. Lunar Sci. Conf.* 5:829–841.

Butler, P., Jr., and Meyer, C., Jr. 1976. Sulfur prevails in coatings on glass droplets: Apollo 15 green and brown glasses and Apollo 17 orange and black (devitrified) glass. *Proc. Lunar Sci. Conf.* 7:1561–1581.

Butler, J. C., Greene, G. M., and King, E. A., Jr. 1973. Grain size frequency distributions and modal analyses of Apollo 16 fines. *Proc. Lunar Sci. Conf.* 4:267–278.

Cadogan, P. H., Eglinton, G., Firth, J. N. M., Maxwell, J. R., Mays, B. J., and Pillinger, C. T. 1972. Survey of lunar carbon compounds: II. The carbon chemistry of Apollo 11, 12, 14, and 15 samples. *Proc. Lunar Sci. Conf.* 3:2069–2090.

Chang, S., Smith, J. W., Kaplan, I., Lawless, J., Kvenvolden, K. A., and Ponnamperuma, C. 1970. Carbon compounds in lunar fines from Mare Tranquillitatis—IV. Evidence for oxides and carbides. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1857–1869.

Chang, S., Lawless, J., Romiez, M., Kaplan, I. R., Petrowski, C., Sakai, H., and Smith, J. W. 1974*a*. Carbon, nitrogen, and sulfur in lunar fines 15012 and 15013: abundances, distributions, and isotopic compositions. *Geochim. Cosmochim. Acta* 38:853–872.

Chang, S., Lennon, K., and Gibson, E. K., Jr. 1974*b*. Abundances of C, N, H, He, and S in Apollo 17 soils from stations 3 and 4: Implications for solar wind exposure ages and regolith evolution. *Proc. Lunar Sci. Conf.* 5:1785–1800.

Charette, M. P., Soderblum, L. A., Adams, J. B., Gaffey, M. J., and McCord, T. B. 1976. Age-color relationships in the lunar highlands. *Proc. Lunar Sci. Conf.* 7:2579–2592.

Charles, R. W., Hewitt, D. A., and Wones, D. R. 1971. H_2O in lunar processes: The stability of hydrous phases in lunar samples 10058 and 12013. *Proc. Lunar Sci. Conf.* 2:645–664.

Chase, M. W., Jr., Davies, C. A., Downey, J. R., Jr., Frurip, D. J., McDonald, R. A., and Syverud, A. N. 1985. *JANAF Thermochemical Tables* (New York: American Chemical Soc. and American Inst. of Physics).

Chou, C. L., Boynton, W. V., Sundberg, L. L., and Wasson, J. T. 1975. Volatiles on the surface of Apollo 15 green glass and trace-element distributions among Apollo 15 soils. *Proc. Lunar Sci. Conf.* 6:1701–1727.

Colson, R. O. 1992. Mineralization on the Moon?: Theoretical considerations of Apollo 16 "rusty rocks," sulfide replacement in 67016, and surface-correlated volatiles on lunar volcanic glass. *Proc. Lunar Planet. Sci.* 22:427–436.

Cripe, J. D., and Moore, C. B. 1975. Total sulfur contents of Apollo 15, 16, and 17 samples. *Lunar Sci.* VI:167–168 (abstract).

Davis, P. A., Jr. 1980. Iron and titanium distribution on the Moon from orbital gamma ray spectrometry with implications for crustal evolutionary models. *J. Geophys. Res.* 85:3209–3224.

Deer, W. A., Howie, R. A., and Zussman, J. 1963. *Rock-Forming Minerals*, vols. 1–5 (London: Longmans, Green).

DesMarais, D. J. 1978. Carbon, nitrogen and sulfur in Apollo 15, 16, and 17 rocks. *Proc. Lunar Planet. Sci. Conf.* 9:2451–2467.

DesMarais, D. J., Hayes, J. M., and Meinschein, W. G. 1973. The distribution in lunar soil of carbon released by pyrolysis. *Proc. Lunar Sci. Conf.* 4:1543–1558.

DesMarais, D. J., Hayes, J. M., and Meinschein, W. G. 1974. The distribution in lunar soil of hydrogen released by pyrolysis. *Proc. Lunar Sci. Conf.* 5:1811–1822.

DesMarais, D. J., Basu, A., Hayes, J. M., and Meinschein, W. G. 1975. Evolution of carbon isotopes, agglutinates, and the lunar regolith. *Proc. Lunar Sci. Conf.* 6:2353–2373.

Drever, J. I., Fitzgerald, R. W., Liang, S. S., and Arrhenius, G. 1970. Phyllosilicates in Apollo 11 samples. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 341–345.

Eberhardt, P., Geiss, J., Graf, H., Grögler, N., Krähenbühl, U., Schwaller, H., Schwarzmüller, J., and Stettler, A. 1970. Trapped solar wind noble gases, exposure age and K/Ar-age in Apollo 11 lunar fine material. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1037–1070.

El Goresy, A., Ramdohr, P., Pavicevic, M., Medenbach, O., Müller, O., and Gentner, W. 1973. Zinc, lead, chlorine, and FeOOH–bearing assemblages in the Apollo 16 sample 66095: Origin by impact of a comet or a carbonaceous chondrite. *Earth Planet. Sci. Let.* 18:411–419.

Engel, A. E. J., and Engel, C. G. 1970. Lunar rock compositions and some interpretations *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1081–1084.

Epstein, S., and Taylor, H. P., Jr. 1970. The concentration and isotopic composition of hydrogen, carbon and silicon in Apollo 11 lunar rocks and minerals. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1085–1096.

Epstein, S., and Taylor, H. P., Jr. 1971. O^{18}/O^{16} Si^{30}/Si^{28}D/H, and C^{13}/C^{12} ratios in lunar samples. *Proc. Lunar Sci. Conf.* 2:1421–1441.

Epstein, S., and Taylor, H. P., Jr. 1972. O^{18}/O^{16} $Si^{30}/Si^{28}C^{13}/C^{12}$, and D/H studies of Apollo 14 and 15 samples. *Proc. Lunar Sci. Conf.* 3:1429–1454.

Epstein, S., and Taylor, H. P., Jr. 1973. The isotopic composition and concentration of water, hydrogen, and carbon in some Apollo 15 and 16 soils and in the Apollo 17 orange soil. *Proc. Lunar Sci. Conf.* 4:1559–1575.

Epstein, S., and Taylor, H. P., Jr. 1974. D/H and $^{18}O/^{16}O$ ratios of H_2O in the "rusty" breccia 66095 and the origin of "lunar" water. *Proc. Lunar Sci. Conf.* 5:1839–1854.

Epstein, S., and Taylor, H. P., Jr. 1975. Investigation of the carbon, hydrogen, oxygen, and silicon isotope and concentration relationships on the grain surfaces of a variety of lunar soils and in some Apollo 15 and 16 core samples. *Proc. Lunar Sci. Conf.* 6:1771–1798.

Ernsberger, F. M. 1972. Analysis of single particles of lunar dust for dissolved gases. *Proc. Lunar Sci. Conf.* 3:2025–2027.

Fegley, B., Jr. 1991. Thermodynamic models of the chemistry of lunar volcanic gases. *Geophys. Res. Lett.* 18:2073–2076.

Fegley, B., Jr. 1992. Lunar volcanic gases: The predicted presence of chlorofluorocarbon gases. *Lunar Planet. Sci.* XXIII:347–348 (abstract).

Fegley, B., Jr., and Cameron, A. G. W. 1987. A vaporization model for iron/silicate fractionation in the Mercury protoplanet. *Earth Planet. Sci. Lett.* 82:207–222.

Feldman, W. C., Reedy, R. C., and McKay, D. S. 1991. Lunar neutron leakage fluxes as a function of composition and hydrogen content. *Geophys. Res. Lett.* 18:2157–2160.

Filleux, C., Tombrello, T. A., and Burnett, D. S. 1977. Direct measurement of surface carbon concentration. *Proc. Lunar Sci. Conf.* 8:3755–3772.

Filleux, C., Spear, R. H., Tombrello, T. A., and Burnett, D. S. 1978. Direct measurement of surface carbon concentrations for lunar soil breccias. *Proc. Lunar Planet. Sci.* 9:1599–1617.

Frick, U., Becker, R. H., and Pepin, R. O. 1988. Solar wind record in the lunar regolith:

Nitrogen and noble gases. *Proc. Lunar Planet. Sci.* 18:87–120.
Friedman, I., Gleason, J. D., and Hardcastle, K. G. 1970. Water, hydrogen, deuterium, carbon and C^{13} content of selected lunar material. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1103–1109.
Friedman, I., O'Neil, J. R., Gleason, J. D., and Hardcastle, K. 1971. The carbon and hydrogen content and isotopic composition of some Apollo 12 materials. *Proc. Lunar Sci. Conf.* 2:1407–1415.
Friedman, I., Hardcastle, K. G., and Gleason, J. D. 1972. Isotopic composition of carbon and hydrogen in some Apollo 14 and 15 samples, In *The Apollo 15 Lunar Samples*, eds. J. W. Chamberlain and C. Watkins (Houston: Lunar and Planetary Inst.), pp. 302–306 (abstract).
Frondel, J.W. 1975. *Lunar Mineralogy* (New York: Wiley).
Funkhouser, J. G., Schaeffer, O. A., Bogard, D. D., and Zähringer, J. 1970. Gas analysis of the lunar surface. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1111–1116.
Futagami, T., Ozima, M., and Nakamura, Y. 1990. Helium ion implantation into minerals. *Earth Planet. Sci. Let.* 101:63–67.
Gay, P., Bancroft, G. M., and Brown, M. G. 1970. Diffraction and Mössbauer studies of minerals from lunar rocks and soils. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 481–497.
Gibson, E. K., Jr., and Andrawes, F. F. 1978*a*. Nature of the gases released from lunar rocks and soils upon crushing. *Proc. Lunar Planet. Sci. Conf.* 9:2433–2450.
Gibson, E. K., Jr., and Andrawes, F. F. 1978*b*. Sulfur abundances in the 74001/74002 drive tube core from Shorty Crater, Apollo 17. *Proc. Lunar Planet. Sci. Conf.* 9:2011–2017.
Gibson, E. K., Jr., and Bustin, R. 1987. Hydrogen abundances vs depth in the lunar regolith: Results from an Apollo 15 double drive tube and deep drill core. *Lunar Planet. Sci.* XVIII 324–325 (abstract).
Gibson, E. K., Jr., and Moore, G. W. 1972. Inorganic gas release and thermal analysis study of Apollo 14 and 15 soils. *Proc. Lunar Sci. Conf.* 3:2029–2040.
Gibson, E. K., Jr., and Moore, G. W. 1973. Carbon and sulfur distributions and abundances in lunar fines. *Proc. Lunar Sci. Conf.* 4:1577–1586.
Gibson, E. K., Jr., and Moore, G. W. 1974. Sulfur abundances and distributions in the valley of Taurus-Littrow. *Proc. Lunar Sci. Conf.* 5:1823–1837.
Gibson, E. K., Jr., Chang, S., Lennon, K., Moore, G. W., and Pearce, G. W. 1975. Sulfur abundances and distributions in mare basalts and their source magmas. *Proc. Lunar Sci. Conf.* 6:1287–1301.
Gibson, E. K., Jr., Usselman, T. M., and Morris, R. V. 1976. Sulfur in the Apollo 17 basalts and their source regions. *Proc. Lunar Sci. Conf.* 7:1491–1505.
Gibson, E. K., Jr., Brett, R., and Andrawes, F. 1977. Sulfur in lunar mare basalts as a function of bulk composition. *Proc. Lunar Sci. Conf.* 8:1417–1428.
Gibson, E. K., Jr., Bustin, R., Skaugset, A., Carr, R. H., Wentworth, S. J., and McKay, D. S. 1987. Hydrogen distributions in lunar materials. *Lunar Planet. Sci.* XVIII:326–327 (abstract).
Gibson, E. K., Jr., Bustin, R., and Mannion, P. 1988. Hydrogen abundances in Apollo 16 and 17 deep drill core and 79001/79002 core samples. *Lunar Planet. Sci.* XIX:387–388 (abstract).
Gillett, S. L. 1990. Chloride lavas: A possible magmatic differentiation product under anhydrous conditions? *Lunar Planet. Sci.* XXI:417–418 (abstract).
Goel, P. S., and Kothari, B. K. 1972. Total nitrogen contents of some Apollo 14 samples by neutron activation analysis. *Proc. Lunar Sci. Conf.* 3:2041–2050.
Goel, P. S., Shukla, P. N., Kothari, B. K., and Garg, A. N. 1974*a*. Solar wind as a source of nitrogen in lunar fines. *Lunar Sci.* V:270–272 (abstract).
Goel, P. S., Shukla, P. N., and Kothari, B. K. 1974*b*. Solar wind nitrogen in the lunar

regolith samples from Luna 16 and Luna 20 sites. In *Further Advances in Lunar Research: Luna 16 and Luna 20 Samples*, eds. N. Bhandari and M. N. Rao (New Delhi: Indian National Science Academy), pp. 72–78.

Goldberg, R. H., Burnett, D. S., and Tombrello, T. A. 1975. Fluorine surface films on lunar samples: Evidence for both lunar and terrestrial origins. *Proc. Lunar Sci. Conf.* 6:2189–2200.

Goldberg, R. H., Tombrello, T. A., and Burnett, D. S. 1976. Fluorine as a constituent in lunar magmatic gases. *Proc. Lunar Sci. Conf.* 7:1597–1613.

Goldstein, J. I., Hewins, R. H., and Romig, A. D., Jr. 1976. Carbides in lunar soils and rocks. *Proc. Lunar Sci. Conf.* 7:807–818.

Gurvich, L. V., Veyts, I. V., and Alcock, C. B., eds. 1989. *Thermodynamic Properties of Individual Substances*, 4th ed. (New York: Hemisphere Publ.).

Haskin, L. A. 1989. The Moon as a practical source of hydrogen and other volatile elements. *Lunar Planet. Sci.* XX:387–388 (abstract).

Haskin, L. A., and Warren, P. H. 1991. Chemistry. In *Lunar Sourcebook: A User's Guide to the Moon* eds. G. H. Heiken, D. T. Vaniman and B. M. French (New York: Cambridge Univ. Press), pp. 357–474.

Heiken, G. H., and Vaniman, D. T. 1990. Characterization of lunar ilmenite resources. *Proc. Lunar Planet. Sci. Conf.* 20:239–247.

Heymann, D., Yaniv, A., and Lakatos, S. 1972. Inert gases from Apollo 12, 14, and 15 fines. *Proc. Lunar Sci. Conf.* 3:1857–1863.

Hintenberger, H., and Weber, H. W. 1973. Trapped rare gases in lunar fines and breccias. *Proc. Lunar Sci. Conf.* 4:2003–2019.

Hintenberger, H., Weber, H. W., Voshage, H., Wänke, H., Begemann, F., and Wlotzka, F. 1970. Concentrations and isotopic abundances of the rare gases, hydrogen and nitrogen in Apollo 11 lunar matter. *Proc. Apollo 11 Lunar Conf.* 2:1269–1282.

Hintenberger, H., Weber, H. W., and Takaoka, N. 1971. Concentrations and isotopic abundances of the rare gases in lunar matter. *Proc. Lunar Sci. Conf.* 2:1607–1625.

Hintenberger, H., Weber, H. W., and Schultz, L. 1974. Solar, spallogenic, and radiogenic rare gases in Apollo 17 soils and breccias. *Proc. Lunar Sci. Conf.* 5:2005–2022.

Hintenberger, H., Schultz, L., and Weber, H. W. 1975. A comparison of noble gases in lunar fines and soil breccias: Implications for the origin of soil breccias. *Proc. Lunar Sci. Conf.* 6:2261–2270.

Hodges, R. R., Jr. 1980. Lunar cold traps and their influence on argon-40. *Proc. Lunar Sci. Conf.* 11:2463–2477.

Hodges, R. R., Jr. 1991. Exospheric transport restrictions on water ice in lunar polar traps. *Geophys. Res. Lett.* 18:2113–2116.

Hohenberg, C. M., Davis, P. K., Kaiser, W. A., Lewis, R. S., and Reynolds, J. H. 1970. Trapped and cosmogenic rare gases from stepwise heating of Apollo 11 samples. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1283–1309.

Hood, L. L., and Schubert, G. 1980. Lunar magnetic anomalies and surface optical properties. *Science* 208:49–51.

Hood, L. L., and Williams, C. R. 1989. The lunar swirls: Distribution and possible origins. *Proc. Lunar Planet. Sci. Conf.* 19:99–113.

Houdashelt, M. L., Bustin, R., and Gibson, E. K., Jr. 1989. Hydrogen extraction from lunar soil: Methods applicable to a lunar processing facility. *Lunar Planet. Sci.* XX:424–425 (abstract).

Hübner, W., Kirsten, T., and Kiko, J. 1975. Rare gases in Apollo 17 soils with emphasis on analysis of size and mineral fractions of soil 74241. *Proc. Lunar Sci. Conf.* 6:2009–2026.

Johnson, J. R., Larson, S. M., and Singer, R. B. 1991. Remote sensing of potential lunar resources: 1. Near-side compositional properties. *J. Geophys. Res.* 96:18861–

18882.
Jordan, J. L. 1989. Prediction of the He distribution at the lunar surface. In *Space Mining and Manufacturing: Proc. of Annual Invitational Symp. of the UA/NASA SERC*, Oct. 24–26, Tucson, Ariz., pp. VII-38–VII-50.
Jovanovic, S., and Reed, G. W., Jr. 1973. Volatile trace elements and the characterization of the Cayley Formation and the primitive lunar crust. *Proc. Lunar Sci. Conf.* 4:1313–1324.
Jovanovic, S., and Reed, G. W., Jr. 1974. Labile and nonlabile element relationships among Apollo 17 samples. *Proc. Lunar Sci. Conf.* 5:1685–1701.
Jovanovic, S., and Reed, G. W., Jr. 1980. Cl, P_2O_5, U and Br associated with mineral separates from a low and a high Ti mare basalt. *Proc. Lunar Planet. Sci. Conf.* 11:125–134.
Jovanovic, S., and Reed, G. W., Jr. 1981. Aspects of the history of 66095 based on trace elements in clasts and whole rock. *Proc. Lunar Planet. Sci. Conf.* 12:295–304.
Kaplan, I. R., and Petrowski, C. 1971. Carbon and sulfur isotope studies on Apollo 12 lunar samples. *Proc. Lunar Sci. Conf.* 2:1397–1406.
Kaplan, I. R., Kerridge, J. F., and Petrowski, C. 1976. Light element geochemistry of the Apollo 15 site. *Proc. Lunar Sci. Conf.* 7:481–492.
Kaplan, I. R., Smith, J. W., and Ruth, E. 1970. Carbon and sulfur concentration and isotopic composition in Apollo 11 lunar samples. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1317–1329.
Kerridge, J. F., Kaplan, I. R., and Petrowski, C. 1975*a*. Nitrogen in the lunar regolith: Solar origin and effects. *Lunar Sci.* VI:469–471 (abstract).
Kerridge, J. F., Kaplan, I. R., Petrowski, C., and Chang, S. 1975*b*. Light element geochemistry of the Apollo 16 site. *Geochim. Cosmochim. Acta* 39:137–162.
Kerridge, J. F., Kaplan, I. R., and Petrowski, C. 1975*c*. Evidence for meteoritic sulfur in the lunar regolith. *Proc. Lunar Sci. Conf.* 6:2151–2162.
Kerridge, J. F., Kaplan, I. R., and Petrowski, C. 1978*a*. Carbon isotope systematics in the Apollo 16 regolith. *Lunar Planet. Sci.* IX:618–620 (abstract).
Kerridge, J. F., Kaplan, I. R., Kung, C. C., Winter, D. A., Friedman, D. L., and DesMarais, D. J. 1978*b*. Light element geochemistry of the Apollo 12 site. *Geochim. Cosmochim. Acta* 42:391–402.
Kirsten, T., Müller, O., Steinbrunn, F., and Zähringer, J. 1970. Study of distribution and variations of rare gases in lunar material by a microprobe technique. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1331–1343.
Kirsten, T., Steinbrunn, F., and Zähringer, J. 1971. Location and variation of trapped rare gases in Apollo 12 lunar samples. *Proc. Lunar Sci. Conf.* 2:1651–1699.
Kirsten, T., Deubner, J., Horn, P., Kaneoka, I., Kiko, J., Schaeffer, O. A., and Thio, S. K. 1972. The rare gas record of Apollo 14 and 15 samples. *Proc. Lunar Sci. Conf.* 3:1865–1889.
Kirsten, T., Horn, P., and Kiko, J. 1973. ^{39}Ar-^{40}Ar dating and rare gas analysis of Apollo 16 rocks and soils. *Proc. Lunar Sci. Conf.* 4:1757–1784.
Kothari, B. K., and Goel, P. S. 1973. Nitrogen in lunar samples. *Proc. Lunar Sci. Conf.* 4:1587–1596.
Leich, D. A., Tombrello, T. A., and Burnett, D. S. 1973. The depth distribution of hydrogen and fluorine in lunar samples. *Proc. Lunar Sci. Conf.* 4:1597–1612.
Leich, D. A., Goldberg, R. H., Burnett, D. S., and Tombrello, T. A. 1974. Hydrogen and fluorine in the surfaces of lunar samples. *Proc. Lunar Sci. Conf.* 5:1869–1884.
LSPET (Lunar Sample Preliminary Examination Team). 1969. Preliminary examination of lunar samples from Apollo 11. *Science* 165:1211–1227.
Mason, B., Melson, W. G., and Nelen, J. 1972. Spinel and hornblende in Apollo 14 fines. *Lunar Sci.* III:512–514 (abstract).
McKay, D. S., Heiken, G., Basu, A., Blanford, G., Simon, S., Reedy, R., French, B. M.,

and Papike, J. 1991. The lunar regolith. In *Lunar Sourcebook: A User's Guide to the Moon* eds. G. Heiken, D. Vaniman and B. French (New York: Cambridge Univ. Press), pp. 285–356.

Merlivat, L., Nief, G., and Roth, E. 1972. Deuterium content of lunar material. *Proc. Lunar Sci. Conf.* 3:1473–1477.

Merlivat, L., Lelu, M., Nief, G., and Roth, E. 1974. Deuterium, hydrogen, and water content of lunar material. *Proc. Lunar Sci. Conf.* 5:1885–1895.

Merlivat, L., Lelu, M., Nief, G., and Roth, E. 1976. Spallation deuterium in rock 70215. *Proc. Lunar Sci. Conf.* 7:649–658.

Metzger, A. E., and Drake, D. M. 1991. Identification of lunar rock types and search for polar ice by gamma ray spectroscopy. *J. Geophys. Res.* 95:449–460.

Meyer, C., Jr., McKay, D. S., Anderson, D. H., and Butler, P., Jr. 1975. The source of sublimates on the Apollo 15 green and Apollo 17 orange glass samples. *Proc. Lunar Sci. Conf.* 6:1673–1699.

Modzeleski, J. E., Modzeleski, V. E., Nagy, L. A., Nagy, B., Hamilton, P. B., McEwan, W. S., and Urey, H. C. 1972. Carbon compounds in Apollo 15 lunar samples. In *The Apollo 15 Lunar Samples*, eds. J. W. Chamberlain and C. Watkins (Houston: Lunar and Planetary Inst.), pp. 311–315.

Moore, C. B., Gibson, E. K., Larimer, J. W., Lewis, C. F., and Nichiporuk, W. 1970. Total carbon and nitrogen abundances in Apollo 11 lunar samples and selected achondrites and basalts. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1375–1382.

Moore, C. B., and Lewis, C. F. 1975. Total nitrogen contents of Apollo 15, 16, and 17 lunar fines samples. *Lunar Sci.* VI:569–571 (abstract).

Moore, C. B., Lewis, C. F., Larimer, J. W., Delles, F. M., Gooley, R. C., Nichiporuk, W., and Gibson, E. K., Jr. 1971. Total carbon and nitrogen abundances in Apollo 12 lunar samples. *Proc. Lunar Sci. Conf.* 2:1343–1350.

Moore, C. B., Lewis, C. F., Cripe, J., Delles, F. M., and Kelly, W. R. 1972. Total carbon, nitrogen, and sulfur in Apollo 14 lunar samples. *Proc. Lunar Sci. Conf.* 3:2051–2058.

Moore, C. B., Lewis, C. F., and Gibson, E. K., Jr. 1973. Total carbon contents of Apollo 15 and 16 lunar samples. *Proc. Lunar Sci. Conf.* 4:1613–1623.

Moore, C. B., Lewis, C. F., and Cripe, J. D. 1974. Total carbon and sulfur contents of Apollo 17 lunar samples. *Proc. Lunar Sci. Conf.* 2:1897–1906.

Morris, R. V. 1978. The surface exposure (maturity) of lunar soils: Some concepts and I_s/FeO compilation. *Proc. Lunar Planet. Sci. Conf.* 9:2287–2297.

Morrison, G. H., Gerard, J. T., Kashuba, A. T., Gangadharam, E. V., Rothenberg, A. M., Potter, N. M., and Miller, G. B. 1970. Elemental abundances of lunar soils and rocks. *Proc. Apollo 11 Lunar Sci. Conf.* 2:1383–1392.

Morrison, G. H., Gerard, J. T., Potter, N. M., Gangadharam, E. V., Rothenberg, A. M., and Burdo, R. A. 1971. Elemental abundances of lunar soils and rocks from Apollo 12. *Proc. Lunar Sci. Conf.* 2:1169–1185.

Müller, O. 1972. Chemically bound nitrogen abundances in lunar samples, and active gases released by heating at lower temperatures (250 to 500°C). *Proc. Lunar Sci. Conf.* 3:2059–2068.

Müller, O. 1973. Chemically bound nitrogen contents of Apollo 16 and Apollo 15 lunar fines. *Proc. Lunar Sci. Conf.* 4:1625–1634.

Müller, O. 1974. Solar wind nitrogen and indigenous nitrogen in Apollo 17 lunar samples. *Proc. Lunar Sci. Conf.* 5:1907–1918.

Müller, O., Grallath, E., and Tölg, G. 1976. Nitrogen in lunar igneous rocks. *Proc. Lunar Sci. Conf.* 7:1615–1622.

Murata, K. 1960. Occurrence of CuCl emissions in volcanic flames. *Amer. J. Sci.* 258:769–782.

Naughton, J. J., Hammond, D. A., Margolis, S. V., and Muenow, D. W. 1972. The

nature and effect of the volatile cloud produced by volcanic and impact events on the moon as derived from a terrestrial volcanic model. *Proc. Lunar Sci. Conf.* 3:2015–2024.

Norris, S. J., Swart, P. K., Wright, I. P., Grady, M. M., and Pillinger, C. T. 1983. A search for correlatable, isotopically light carbon and nitrogen components in lunar soils and breccias. *Proc. Lunar Planet. Sci. Conf.* 13, *J. Geophys. Res.* 88:B200–B210.

Palme, H., Baddenhausen, H., Blum, K., Cendales, M., Dreibus, G., Hofmeister, H., Kruse, H., Palme, C., Spettel, B., Vilcsek, E., and Wänke, H. 1978. New data on lunar samples and achondrites and a comparison of the least fractionated samples from the earth, the moon, and the eucrite parent body. *Proc. Lunar Planet. Sci. Conf.* 9:25–57.

Pepin, R. O., Dragon, J. C., Johnson, N. L., Bates, A., Coscio, M. R., Jr., and Murthy, V. R. 1975. Rare gases and Ca, Sr, and Ba in Apollo 17 drill-core fines. *Proc. Lunar Sci. Conf.* 6:2027–2055.

Pepin, R. O., Nyquist, L. E., Phinney, D., and Black, D. C. 1970. Rare gases in Apollo 11 lunar material. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1443–1454.

Pieters, C. M. 1978. Mare basalt types on the front side of the moon: A summary of spectral reflectance data. *Proc. Lunar Planet. Sci. Conf.* 9:2825–2849.

Rees, C. E., and Thode, H. G. 1972. Sulphur concentrations and isotope ratios in lunar samples. *Proc. Lunar Sci. Conf.* 3:1479–1485.

Rees, C. E., and Thode, H. G. 1974. Sulfur concentrations and isotope ratios in Apollo 16 and 17 samples. *Proc. Lunar Sci. Conf.* 5:1963–1973.

Roedder, E., and Weiblen, P. W. 1970. Lunar petrology of silicate melt inclusions in Apollo 11 rocks. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 801–837.

Sakai, H., Petrowski, C., Goldhaber, M. B., and Kaplan, I. R. 1972. Distribution of carbon, sulfur and nitrogen in Apollo 14 and 15 material. *Lunar Sci.* III:672–674 (abstract).

Sato, M. 1976. Oxygen fugacity and other thermochemical parameters of Apollo 17 high-Ti basalts and their implications on the reduction mechanism. *Proc. Lunar Sci. Conf.* 7:1323–1344.

Sato, M. 1979. The driving mechanism of lunar pyroclastic eruptions inferred from the oxygen fugacity behavior of Apollo 17 orange glass. *Proc. Lunar Planet. Sci. Conf.* 10:311–325.

Sato, M., Hickling, N. L., and McLane, J. E. 1973. Oxygen fugacity values of Apollo 12, 14, and 15 lunar samples and reduced state of lunar magmas. *Proc. Lunar Sci. Conf.* 4:1061–1079.

Signer, P., Bauer, H., Derksen, U., Etique, P., Funk, H., Horn, P., and Wieler, R. 1977. Helium, neon, and argon records of lunar soil evolution. *Proc. Lunar Sci. Conf.* 8:3657–3683.

Simoneit, B. R., Wszolek, P. C., and Burlingame, A. L. 1972. Apollo 15 lunar samples: LM exhaust products in the SESC 15013. In *The Apollo 15 Lunar Samples,* eds. J. W. Chamberlain and C. Watkins (Houston: Lunar and Planetary Inst.), pp. 286–290 (abstract).

Smith, J. W., Kaplan, I. R., and Petrowski, C. 1973. Carbon, nitrogen, sulfur, helium, hydrogen, and metallic iron in Apollo 15 drill stem fines. *Proc. Lunar Sci. Conf.* 4:1651–1656.

Srinivasan, B., Hennecke, E. W., Sinclair, D. E., and Manuel, O. K. 1972. A comparison of noble gases released from lunar fines (#15601,64) with noble gases in meteorites and in the earth. *Proc. Lunar Sci. Conf.* 3:1927–1945.

Stievenard, M., Jouzel, J., and Robert, F. 1990. The isotopic signature of hydrogen in lunar basalts: Spallogenic or primordial? *Lunar Planet. Sci.* XXI:1204–1205 (abstract).

Swindle, T. D., Glass, C. E., and Poulton, M. M. 1990. *Mining Lunar Soils for ^{3}He.* UA/NASA Space Engineering Research Center TM-90/1 (Tucson: UA/NASA SERC).

Swindle, T. D., Burkland, M. K., Johnson, J. R., Larson, S. M., Morris, R. V., Rizk, B., and Singer, R. B. 1992. Systematic variations in solar wind fluence with lunar location: Implications for abundances of solar-wind-implanted volatiles, *Lunar Planet. Sci.* XXIII:1395–1396 (abstract).

Symonds, R. B., Rose, W. I., Reed, M. H., Lichte, F. E., and Finnegan, D. L. 1987. Volatilization, transport and sublimation of metallic and non-metallic elements in high temperature gases at Merapi Volcano, Indonesia. *Geochim. Cosmochim. Acta* 51:2083–2101.

Taylor, G. J., Warren, P., Ryder, G., Delano, J., Pieters, C., and Lofgren, G. 1991. Lunar Rocks. In *Lunar Sourcebook: A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (New York: Cambridge Univ. Press), pp. 183–284.

Taylor, L. A. 1991. Helium abundances on the Moon: Assumptions and estimates. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 40.

Taylor, L. A., Mao, H. K., and Bell, P. M. 1973. "Rust" in the Apollo 16 rocks. *Proc. Lunar Sci. Conf.* 4:829–839.

Taylor, L. A., Mao, H. K., and Bell, P. M. 1974. β-FeOOH, akaganéite in lunar rocks. *Proc. Lunar Sci. Conf.* 5:743–748.

Thiemens, M. H., and Clayton, R. N. 1980. Solar and cosmogenic nitrogen in the Apollo 17 deep drill core. *Proc. Lunar Planet. Sci. Conf.* 11:1435–1451.

Wänke, H., Rieder, R., Baddenhausen, H., Spettel, B., Teschke, F., Quijano-Rico, M., and Balacesu, A. 1970. Major and trace elements in lunar material. *Proc. Apollo 11 Lunar Sci. Conf.*, pp. 1719–1727.

Wänke, H., Wlotzka, F., Baddenhausen, H., Balacescu, A., Spettel, B., Teschke, F., Jagoutz, E., Kruse, H., Quijano-Rico, M., and Rieder, R. 1971. Apollo 12 samples: Chemical composition and its relation to sample locations and exposure ages, the two component origin of the various soil samples and studies on lunar metallic particles. *Proc. Lunar Sci. Conf.* 2:1187–1208.

Wänke, H., Baddenhausen, H., Balacescu, A., Teschke, F., Spettel, B., Dreibus, G., Palme, H., Quijano-Rico, M., Kruse, H., Wlotzka, F., and Begemann, F. 1972. Multielement analyses of lunar samples and some implications of the results. *Proc. Lunar Sci. Conf.* 3:1251–1268.

Wänke, H., Baddenhausen, H., Dreibus, G., Jagoutz, E., Kruse, H., Palme, H., Spettel, B., and Teschke, F. 1973. Multielement analyses of Apollo 15, 16, and 17 samples and the bulk composition of the moon. *Proc. Lunar Sci. Conf.* 4:1461–1481.

Wänke, H., Palme, H., Baddenhausen, H., Dreibus, G., Jagoutz, E., Kruse, H., Spettel, B., Teschke, F., and Thacker, R. 1974. Chemistry of Apollo 16 and 17 samples: Bulk composition, late stage accumulation and early differentiation of the moon. *Proc. Lunar Sci. Conf.* 5:1307–1335.

Wänke, H., Palme, H., Baddenhausen, H., Dreibus, G., Jagoutz, E., Kruse, H., Palme, C., Spettel, B., Teschke, F., and Thacker, R. 1975. New data on the chemistry of lunar samples: Primary matter in the lunar highlands and the bulk composition of the moon. *Proc. Lunar Sci. Conf.* 6:1313–1340.

Wänke, H., Palme, H., Kruse, H., Baddenhausen, H., Cendales, M., Dreibus, G., Hofmeister, H., Jagoutz, E., Palme, C., Spettel, B., and Thacker, R. 1976. Chemistry of lunar highland rocks: A refined evaluation of the composition of the primary matter. *Proc. Lunar Sci. Conf.* 7:3479–3499.

Wänke, H., Baddenhausen, H., Blum, K., Cendales, M., Dreibus, G., Hofmeister, H., Kruse, H., Jagoutz, E., Palme, C., Spettel, B., Thacker, R., and Vilcsek, E. 1977.

On the chemistry of lunar samples and achondrites. Primary matter in the lunar highlands: A re-evaluation. *Proc. Lunar Sci. Conf.* 8:2191–2213.

Wasson, J. T., Boynton, W. V., Kallemeyn, G. W., Sundberg, L. L., and Wai, C. M. 1976. Volatile compounds released during lunar lava fountaining. *Proc. Lunar Sci. Conf.* 7:1583–1595.

Watson, K., Murray, B. C., and Brown, H. 1961. The behavior of volatiles on the lunar surface. *J. Geophys. Res.* 66:3033–3045.

Wellman, T. R. 1970. Gaseous species in equilibrium with the Apollo 11 holocrystalline rocks during their crystallization. *Nature* 225:716–717.

White, D. E., and Waring, G. A. 1963. Volcanic emanations. In *Data of Geochemistry,* 6th ed., ed. M. Fleischer (Washington, D. C.: U. S. G. P. O.), pp. K1–K27.

Wieler, R., Etique, Ph., Signer, P., and Poupeau, G. 1980. Record of the solar corpuscular radiation in minerals from lunar soils: A comparative study of noble gases and tracks. *Proc. Lunar Planet. Sci. Conf.* 11:1369–1393.

Wilson, E. B., Jr. 1952. *An Introduction to Scientific Research* (New York: McGraw-Hill).

Wittenberg, L. J., Santarius, J. F., and Kulcinski, G. L. 1986. Lunar source of ^{3}He for commercial fusion power. *Fusion Tech.* 10:167–178.

Wszolek, P. C., Jackson, R. F., and Burlingame, A. L. 1972. Carbon chemistry of the Apollo 15 deep drill stem and a glass-rich sample related to the uniformity of the regolith and lunar surface processes. In *The Apollo 15 Lunar Samples,* eds. J. W. Chamberlain and C. Watkins (Houston: Lunar and Planetary Inst.), pp. 324–328.

LUNAR BASE SITING

ROBERT L. STAEHLE, JAMES D. BURKE and
GERALD C. SNYDER
Jet Propulsion Laboratory

RICHARD DOWLING
World Space Foundation

and

PAUL D. SPUDIS
Lunar and Planetary Institute

There are widely dispersed lunar sites of interest for known and potential resources, selenology and observatories. Discriminating characteristics include certain geologic and topographic features, local mineralogy and petrology, solar illumination, view of Earth and the celestial sphere, and soil engineering properties. Space vehicle arrival and departure trajectories favor equatorial and polar sites. Over time, base sites will be developed serving different purposes. Information may be the initial lunar "resource," in the form of observational and *in-situ* research. Resource-driven sites may see the fastest growth during early decades of lunar development, but selection of initial sites is likely to be driven by suitability for a combination of activities. Only equatorial locations offer nearly all-sky views for astronomy, while most of the far side offers radio isolation. A base in Mare Smythii with subsidiary outposts is favorable for a variety of purposes, and preserves broad resource flexibility. Discovery of accessible volatiles, in the form of polar permafrost, subsurface gas reservoirs, or comet impact remnants, would dramatically increase the attractiveness of such a site from a logistical support and selenological point of view. Amid much speculation, no reliable evidence of such volatiles exists. With the availability of near-constant sunlight for power generation, and permanently shadowed areas at cryogenic temperatures, polar sites require substantially less Earth-launched mass and lower equipment complexity for an initial base. Polar sites are scientifically less interesting. Reliable evidence exists for areas of certain mineral concentrations, such as ilmenite, which could form a feedstock for some proposed resource extraction schemes. In addition to being a source for oxygen and iron, ilmenite harbors higher concentrations of solar-wind-implanted H, C, N and He. New data from a lunar polar orbiter are essential for the most informed site selection. Data from the first Galileo flyby have already revealed previously unknown features and will aid surface mineralogical characterization.

I. INTRODUCTION

Of nineteen lunar surface sites explored to date, a diversity of features and

characteristics have been examined. If the first lunar "resource" is information, then the utility of a locale for the *in-situ* and observational sciences will rank high. Early site selection will be governed by safety, economy, and immediate utility of the resources already known. Later site selections will depend on new knowledge of all types of resources. Present discussion of base sites is driven most strongly by the scientific community with consideration to engineering feasibility and eventual resource utilization (see, e.g., Duke et al. 1984; Selzer 1989; Morrison 1990; Spudis et al. 1989). Lunar geology, lunar geophysics and other disciplines concerning the Moon and its environs (selenology) dominate one branch of scientific utilization, while use of the Moon as a platform for astronomy, space physics, Earth and solar observations dominates the other branch. Overlying any discussion of site selection for these uses are the suitability of local terrain, viewing of the Sun, Earth and heavens, availability of energy and heat rejection paths, and local material resources (Fig. 1).

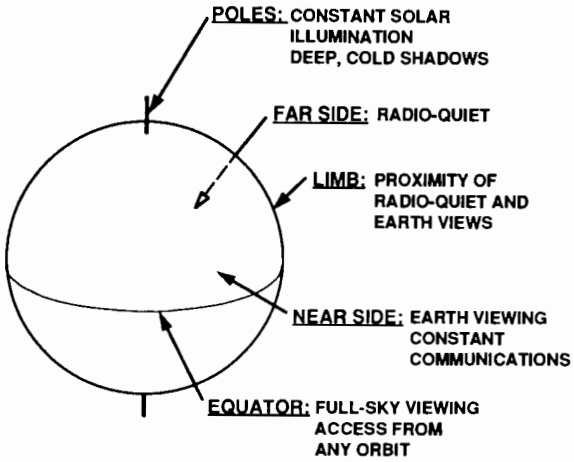


Figure 1. Different parts of the Moon suitable for bases set up for different purposes. Visibility of the Earth, Sun, and celestial sphere, and shielding from the Earth and Sun, are all in a sense "resources" of importance to different potential users of a first lunar base.

One of the earliest discussions of lunar base location was by R. H. Goddard (Goddard 1920), in which he noted, "The best location on the Moon would be at the north or south pole with the [propellant] liquefier in a crater, from which the water of crystallization may not have evaporated, and with the [solar] power plant on a summit constantly exposed to the Sun. Adequate protection should, of course, be made against meteors, by covering the essential parts of the apparatus with rock." (Fig. 2) Many would still credit him with a valid conclusion, even though geologists will offer different explanations if volatiles are found at the poles. The possibility of polar ice deposits was "rediscovered" in 1960 (Watson et al. 1961). Recent groundbased radar

indications of the possibility of ice near Mercury's poles (Harmon and Slade 1991; Slade et al. 1991) weakens some arguments against the possibility of lunar polar ice by suggesting that solar wind erosion may be less important than proposed (Lanzerotti et al. 1981) in limiting ice buildup.

Early locales with diverse materials are likely to outrank locales with the highest concentration of a single desired substance. The exception may be any site with a concentration of H or C in some form, such as ices or subsurface gas reservoirs. Scarcity of reducing agents has come to be the dominant limitation in most discussions of lunar resource utilization. Geologists want to sample and record a diversity of terrain representing the major geologic phases of the Moon's formation and evolution. Mare and highland sampling at many sites is considered essential, with age diversity important.

Energy is another resource, certainly for surface operations, and perhaps even for export (see, e.g., Criswell 1984). If nuclear power is unavailable at the required levels, energy storage equipment for the 14-day night is nontrivial. High crater rims and peaks near both poles may offer near-constant solar illumination, and modest towers at these locations certainly will, but a thorough lighting survey has yet to be conducted to pin down the best locations. Any location on the Moon would do for a partial gravity test facility for life sciences investigations (Griffin, personal communication). For photosensitive organisms, polar locations could offer piped-in sunlight on any diurnal cycle researchers might choose.

Flight mechanics into and out of a base site can be an important consideration. Equatorial and polar sites are favored for their near-constant accessibility. Slopes and terrain features can be resources themselves. Slopes offer favored illumination and shadowing, while craters offer natural depressions for astronomical instruments, barriers to lander exhaust-driven debris, reactor shields and other uses. Elevation differences also figure into some energy storage schemes.

Surface mobility will influence site selection by dictating the range of accessibility from a core base site. Subsidiary sites can serve a variety of specialized purposes, such as mining where different ores are accessible. Sensitive astronomical instruments will need to be away from frequent surface activities. If a main base is located near the limb as viewed from Earth, a subsidiary site at about 101° east or west longitude affords sufficient radio isolation from Earth at the limits of the Moon's east-west libration, after accounting for diffraction.

If we had to choose a site today and be certain of a workable, if not at all optimal locale, Apollo 15/Hadley Rille would be a reasonable choice. But we can already see superior sites, though we do not know precisely where it is safe to put the base's first landers down. Virtually all investigators agree on the wisdom of a lunar polar orbiter with suitable composition-measuring instruments plus imaging. Surface rovers may be advisable at "finalist" sites, while teleoperated rovers will surely play an important role in exploration from any base site. Precursor missions could even be used to build a cache of

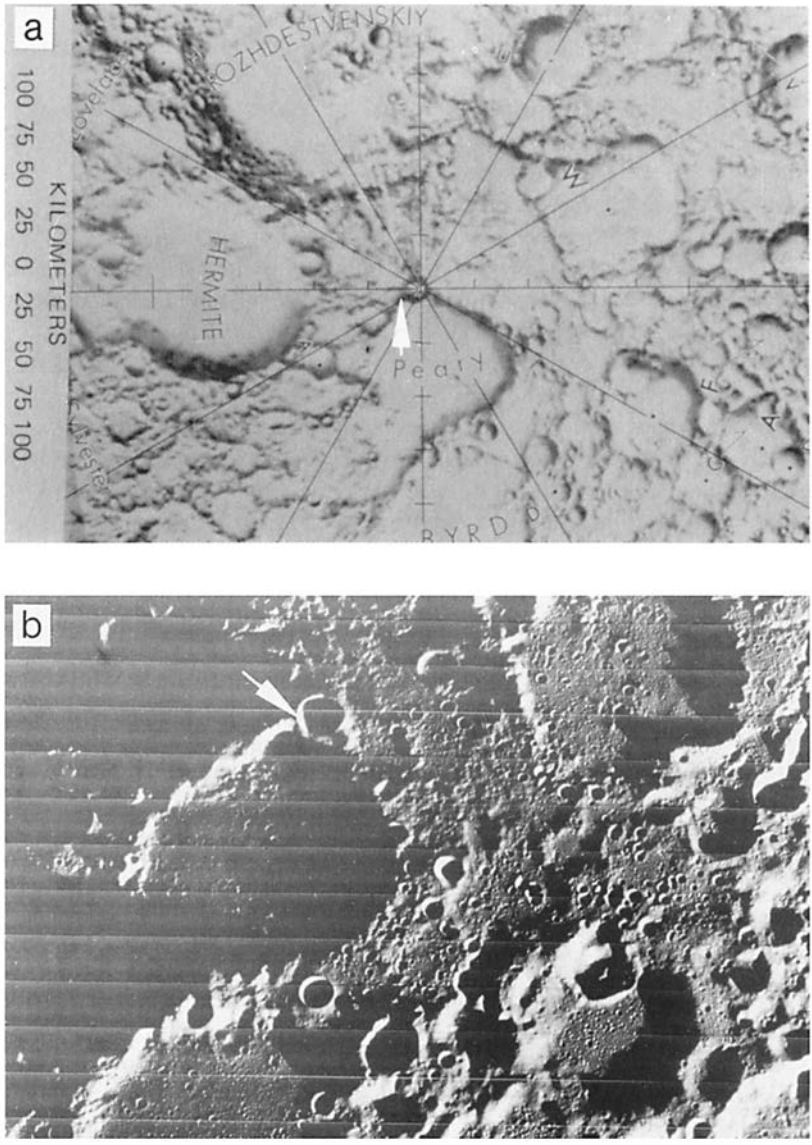


Figure 2. This north polar location, at the intersection of the large crater Peary and smaller Polaris, could satisfy R. Goddard's criteria for a base site. Top half of the map is the lunar far side, with the east limb to the right. Map and high-resolution Lunar Orbiter IV photo courtesy of U. S. Geological Survey.

some useful product, such as oxygen, for use by later human explorers.

As important as further lunar reconnaissance is, terrestrial development and testing of alternative resource extraction processes is essential. Operation

of one or more lunar base analogs would offer invaluable experience at a fraction of the cost of making mistakes on the Moon. Determination of the most workable and economical resource extraction processes will influence any resource-driven site selection.

II. PRESENT UNDERSTANDING OF THE MOON

The last unmapped region of the Moon, near the south pole, was photographed during the 1990 December 8 Galileo flyby, but there is a great deal more that would be helpful to know in selecting base sites. From Ranger through to open up mission constraints to afford better scientific opportunities. Apollo 11 was sent to a flat mare region for safety. The Apollo 17 site was selected for its geological diversity within a small area. It was believed that vertical sections could be sampled by collecting the talus fallen from steep slopes, presaging the possible employment of similar slopes as a resource, such as for excavation of shielding with minimum use of energy.

The last three Apollos carried a set of orbital instruments designed to map the surface at fine resolution and infer its composition, but near-equatorial orbits limited their coverage to less than 20% of the Moon. Crude geologic maps of the entire surface have been constructed from the best available data of all types (Wilhelms 1987). Other physical and compositional information, to the extent we know it, has been summarized by a variety of authors (see, e.g., Heiken et al. 1991).

Information needed for selecting the best base sites depends on the objectives of these bases. However, some kinds of data are required for nearly any base. Local topography is an obvious need, and most investigators agree that, except for the immediate vicinity of the Apollo sites, present information is inadequate. Even without elevation data, positions of features are typically uncertain by 1 to 3 km on the near side, by 3 to 6 km poleward of 65° latitude, and by 3 to 15 km on the far side (Davies, M., personal communication).

An orbiting laser altimeter and a metric camera system offer the preferred means for improving lunar topographic maps. Knowing topographic obstacles is essential for safe approach from and departure to orbit, as well as for designing solar power and thermal radiation installations for a specific site. Spatial resolution of 1 m or to certify landing sites.

The next most important new information probably concerns the subsurface mechanical properties, to a depth of at least a meter, that affect digging, foundation-building and other preparations at any specific site. Where mining is contemplated, such information is important to a greater depth. Though the lunar surface has been somewhat homogenized by impacts, it does vary in state of compaction, grain size distribution, size of embedded rocks and other mechanical properties.

While compositional properties may be less important than local topography and soil mechanics during the earliest lunar operations, composition will dominate once resource development begins. Compositional information is

therefore highly desireable even before choosing the first base site. Multispectral remote sensing from orbit provides needed regional data, after which surface traverses are best for detailing the most promising locales. Long range rovers teleoperated from Earth, carrying imaging, geochemical and geophysical instruments, would be suitable for both scientific and resource site reconnaissance. Use of these rovers could continue during base buildup.

Depending on a short list of candidate sites, different kinds of local information may be useful for selecting a final site. For a polar site, an orbiter with altimetry and metric imaging could perform a survey of varying surface lighting conditions for siting solar power generators, radiators and instruments. In the event an orbiter detects indications of volatiles near one of the poles (see, e.g. Watson et al. 1961; Lanzerotti et al. 1981; Staehle 1983), surface exploration may be required for precise location. In a similar fashion, it could prove useful to explore volcanic areas, such as the region around Aristarchus, for possible vents and associated mineralization, lava tubes which could make natural base shelters (Horz 1985), and other physical and compositional features.

III. MATERIAL RESOURCES

First consideration of material resources is given to a site's ability to support local operations. At least a meter, and preferably deeper regolith is desirable for burying initial habitation structures to suitable depth for long-term cosmic ray and solar flare shielding. Two meters of loose material protecting inhabited structures from all directions, achieved by a combination of trenching and burying, is considered adequate. Mechanical properties should offer easy excavation. Ilmenite-rich mare soil provides slightly superior radiation protection for a given thickness than lower-density highland material, but this is not likely to be a decisive advantage in base construction.

Second consideration is given to reducing need for costly importation of terrestrial material for functions easily replaced by lunar material. Perhaps the simplest processed lunar material is cast or sintered basalt (see, e.g., Green 1961; Kopecky and Voldan 1965; Dalton and Hohmann 1972; Garvey and Magoffin 1991). Based on testing of close terrestrial analogs, mare basalts appear to be of suitable composition to be melted, poured into forms, and cooled into bricks and more complex structural forms, or spun into insulating rock wool (University of Houston 1990), as has been done in a few terrestrial industries for decades. Melting and sintering temperatures are about 200°C less for basalts than for typical plagioclase-rich highland materials, and therefore require less process heat. Microwave heating could be an alternative to direct solar heating, and could require ilmenite-bearing feedstock for electromagnetic coupling (Meek et al. 1985). Materials for production of some metals, solar cells, cement (based on CaO), concrete, etc. may be more easily extracted from highlands although concentrates from mare materials will be adequate. Plagioclase-rich highland materials produce a higher-strength,

more transparent glass, but anorthite can be beneficiated from most soils for specific processes. For simple building materials, a mare site is superior but highland materials will work.

Volatiles in lunar samples have been shown to originate from solar-wind implantation (Haskin and Warren 1991). Concentrations of H, C and N, the most valuable for life support and propellant, range from approximately 10 to 100 μg g^{-1} in lunar soils and regolith breccias. Because these elements implant over time on the surface of mineral grains, their mass concentrations are highest on smaller grains in older soils. Concentrations are much lower in solid igneous rocks. Retention on ilmenite grains is preferential to other common minerals. It is not clear that the bulk availability of solar wind-implanted H, C or N is sufficient for practical production quantities of propellant. Other possible sources of volatile compounds include cometary impacts. H_2O, CO_2, CH_4, H_2S, NH_3 or other volatiles are unlikely to last long near the impact points, but could collect in polar cold traps.

Simple heating of lunar soils to 700°C will liberate most of the volatiles, with heating above 1050°C required to obtain most of the rest. While large quantities of soil are required, a simple solar-driven process can yield gas pressures up to 10 atm on material sufficient at least for habitat leakage make-up and life support process losses. Assuming a bulk soil density of 1600 kg m^{-3} for the top 3 m (Carrier et al. 1991), a yield of 100+ g m^{-3} each of H, C and N could be expected in a variety of simple compounds. Young crater rims and ejecta blankets are probably deficient in implanted volatiles; other areas with sufficient regolith depth (probably most of the Moon) are likely to be satisfactory, though there may be a preference for ilmenite-enriched regions.

Specialized ore bodies could take several forms. First, "ore" should be defined as a natural concentration of a useful substance to a level and in a form which makes its extraction economical. Most mineral concentrations remain to be discovered. Even on Earth, ore bodies are seldom discovered and never confirmed without *in-situ* sampling. At this point we can only suggest a few kinds of lunar materials which might prove important to base location. A surely primitive list could read, in descending order of importance: mare basalt regolith, ilmenite ($FeTiO_3$, often associated with FeO), metallic Fe, pyroclastic glasses with semi-volatiles (S, Pb, Hg, Zn, Te, etc.), high Al highland material, and KREEP.

Ilmenite has been discussed as a feedstock for oxygen production by chemical reduction (see, e.g., Triffet at al. 1991), for its higher solar wind volatiles content, and for the potential to beneficiate it from soil using relatively simple electrostatic techniques (Agosto 1985). However, no one has yet demonstrated that naturally occurring lunar ilmenite can be adequately separated from accompanying substances to form a suitable cost-effective feedstock. Therefore, ilmenite availability as a major siting criterion could be a trap. Early use of ilmenite is less often described in terms of a source of Fe or Ti. Ilmenite is especially abundant (up to 20% by volume) in some

Apollo 11 and Apollo 17 mare basalts. Ilmenite is most often associated with high-Ti basalts (Shevchenko 1992; Johnson et al. 1991) in maria. Reduced metallic Fe and Fe-Ni phases are ubiquitous in soil at fractional weight percents, apparently the product of meteoroid impacts, igneous crystallization and a reducing environment (Lewis et al. 1988; Papike et al. 1991). While not considered an important early base-siting criterion, availability of reduced metals such as Fe could become important later. Older terrains, with deeper regolith, presumably have more metallics, i.e., Fe and Ni, which may be easily beneficiated magnetically.

For oxygen extraction, magma electrolysis, high temperature pyrolysis (Senior 1989) and fluoride processing are somewhat site independent, though process energies may vary. Ilmenite reduction and pyroclastic glass processing require site-variable feedstocks.

A conclusion of the 1990 April Johnson Space Center lunar base workshop participants bears repeating: "...We conclude that from the point of view of resource utilization, a viable strategy would be to select a high titanium mare site, perhaps on or near a pyroclastic area, and near a highland area so that calcium-rich feedstock would also be available" (Morrison 1990).

There are not today any credible economic exports from the Moon to the Earth's surface. Exporting propellants to low Earth orbit could be viable on a mass basis (Carroll et al. 1983) (i.e., on a per-ton basis, less than a ton of equipment need be launched into Earth orbit to retrieve a ton of oxygen from the Moon), but the market size and launch costs do not make this economically attractive now. As with the terrestrial frontier, the economics are expected to change with time as transportation and operating costs decrease, markets enlarge and scarcities develop. Such changes in economics could drive lunar base-siting considerations several decades hence. One example, lunar ^{3}He, has been proposed as a possible thermonuclear fuel source (Wittenberg et al. 1987). If this proves attractive in the future, this isotope can be found concentrated along with solar-wind-implanted H, C and N in fine soil grains.

IV. LUNAR GEOSCIENCES

The lunar geosciences seek to ascertain the origin, history and evolution of the Moon. As on Earth and Mars, different regions afford examples of different processes of varying ages. Lunar geoscience investigations will be among the prime tasks at early lunar bases (see, e.g., Spudis et al. 1989). While much has been learned to date, iterative field study is required to understand the Moon's geologic processes at increasing levels of detail.

Geologists most want field access to a variety of different geologic provinces and features. Human intelligence and experience is required, either on-site or through telepresence, backed up by field measurements and the ability to analyze samples in a variety of ways in lunar and terrestrial laboratories. Some examples of important landforms are central peaks of large craters where complex outcrops occur, megablocks of brecciated highland crust that

may occur as ejecta and as exposures within crater walls, crater and basin ejecta deposits, sinuous rilles, wrinkle ridges and other volcanic features, and highland-mare contact regions. The Apollo 17 site at Taurus-Littrow was chosen for the one mission carrying a geologist because of its access to a variety of geologic features within the ~10 km range of the Lunar Roving Vehicle (Schmitt 1975).

Any features of present activity, such as gas vents (Johansen and Kirk 1982), would be of particular interest, as would polar volatiles deposits.

Wide nets of geophysical instrumentation are also important. Seismic and magnetic stations can be used to study the internal structure of the Moon, and would be deployed from a base in conjunction with remotely deployed stations around the globe.

V. ASTRONOMY FROM THE MOON

The site-selection criteria for astronomy and astrophysics, space physics, and even part of planetary science differ greatly from those the lunar sciences, in that they include not only what is under foot, but also what is overhead. What is shared with other areas is diversity—for instance, some disciplines wish to view Earth, while to others Earth is a source of electromagnetic noise. Site evaluation criteria apply to terrain (and other local aspects of the environment) and celestial accessibility. This discussion assumes there is a limitation to a single site. Arguments against a polar site would be much weaker if both poles could be used for observations.

By far the dominant criterion for the celestial sciences is sky accessibility (Figs. 3, 4). From either pole half the sky can never be seen. From near the equator virtually all the sky is available at some time during a month, and the part of the sky obscured by the Sun becomes available after a month or two. Unfortunately, the issue of how much science is lost as a function of latitude is complicated enough that there is not a single criterion available. Some studies require observations of a single individual object, such as Supernova 1987A. For these, the location of the observatory site is make-or-break. A site too far north would block the view. Other studies with a very small number of targets would be seriously compromised if one or two were not available but could still be performed. Such targets might be the nearest stars of a given type within our own Galaxy, or the nearest external galaxies. Still other studies have enough targets that they suffer only moderately if part of the sky is unavailable. Finally, some studies have so many targets that any moderate-sized area of the sky would suffice. In some cases increasing either the instrument sensitivity or aperture (or some combination) can result in more target objects becoming available, offsetting the disadvantages of a higher latitude site for some investigations.

However, many top priority astronomical questions have a small number of useful targets and would suffer significantly if the observatory were placed far from the equator.

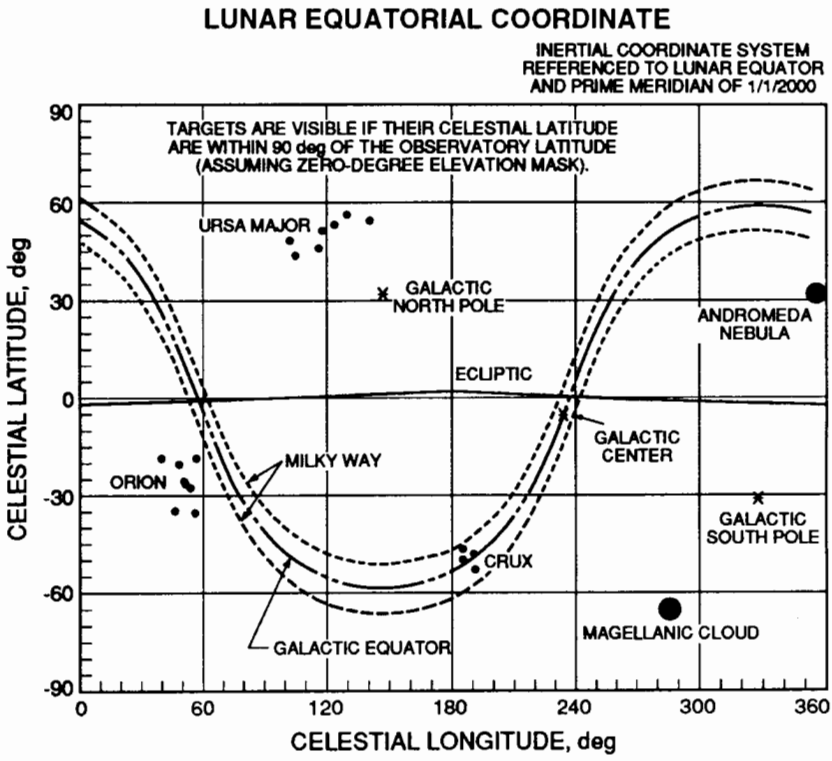


Figure 3. Schematic view of the sky referenced to lunar equator (figure courtesy of JPL).

Another observing parameter dependent on lunar latitude is the (u,v)-plane coverage for interferometer arrays. Interferometric imaging combines information from many different baselines to simulate a large, filled aperture. The longest baselines give information about the finest detail in the object, and the shortest baselines give data about the large-scale aspects of the object. Unless an array has a very large number of elements, it relies on the changing geometry as the target rises and sets to generate the different (projected) baselines. The shortest baselines are achieved near rise and set for any given target. From a polar site, objects in the sky do not rise or set, but rather just circle around the pole. Therefore interferometric imaging is much enhanced near the equator.

Observatory longitude is also of interest, at least to some investigations, because of the Earth. The one which cares the most is the proposed very low frequency (VLF 1-30 MHz) interferometer array which is best placed a few hundred kilometers around onto the far side of the Moon, so that the instrument will be protected from natural and man-made radio noise, taking into account diffraction effects and worst-case lunar libration.

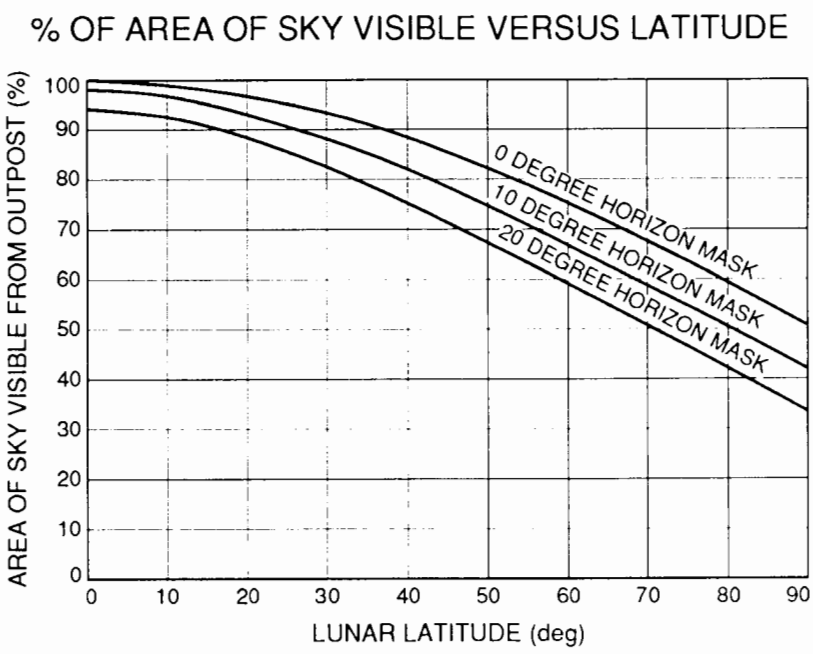


Figure 4. Fraction of sky visible as a function of lunar observatory latitude. Horizon masks refer to the minimum acceptable elevation above the ideal horizon at which observations can be made. Actual limitations will vary with local site and topography, viewing direction and instrument characteristics (figure courtesy of JPL).

For other experiments, greatly eased communications makes near-side siting preferable. This is true even though Earth is also a bright source of visible light and thermal infrared. Because of this radiation, sensitive instruments can be required to have an Earth-avoidance angle of several degrees, and the telescope guidance algorithms can be simplified if Earth appears near the horizon rather than near the zenith. This is not a high-priority criterion. Where an external Earth-occulting disk is needed, either a site near the limb (Earth near the horizon) or a site near the sub-Earth point (Earth near zenith) may be simpler to accommodate than intermediate geometries.

Thermal environment is also latitude dependent. Cooling is more difficult near the equator. Poleward of about 70°, portions of some crater interiors never receive direct sunlight. With surface temperatures perhaps as low as 90 K, they would constitute an ideal environment for infrared and other instruments which need to be as cold as possible. There is believed to be a particle haze associated with the terminator within a few meters of the surface, caused by electrostatic suspension of particles. Because the terminator is always present in the polar regions, this haze and terminator-correlated (but very tenuous)

gas clouds could influence site selection for instruments sensitive to scattered light (Burke 1985). However, low surface temperatures near the poles might suppress these phenomena compared with lower latitudes.

Despite seeming to have a fixation with the sky, astronomers care about terrain, also. For stand-alone instruments soil mechanics over the instrument footprint are what matters. For an interferometer array, both the sites of the individual telescopes and the lines of sight between them are important. The site for a lunar telescope needs to be reasonably level and stable. Levelness within a few degrees is sufficient, depending on the emplacement scheme and pointing mechanisms. Stability needs to be to a fraction of the resolution of the instrument: from about 10^{-1} arcsec for a stand-alone telescope to 10^{-6} arcsec for an optical interferometer. Ease of meeting the line-of-sight requirement depends, among other things, on the size of the array. Over baselines up to $\sim$1 km, curvature of the Moon is a fraction of a meter, but at 10 km curvature approaches 30 m, requiring either tall towers or several intermediate relay devices, unless a bowl-shaped site can be found which approximately cancels the mean lunar curvature. Suitable craters are almost certainly available.

Gravity-wave detectors or telescopes have requirements similar to optical astronomy, with particular sensitivity to stability and seismic isolation. For experiments sensitive to gamma rays, high KREEP concentrations could prove a problem, with their radioactive K, U and Th components.

A final consideration is nearness to other base activities. Availability of power and data handling capabilities of the outpost is a benefit of proximity. Alternatively, base activity will generate some phenomena which the observatories are going to the Moon to avoid. These include dust from local activities, rocket launches and landings; volatiles from materials processing, rockets, and habitat leakage; light; radio transmissions; and vibration. This last is particularly important to optical interferometers, which need to control positions of optical surfaces to within 1 nm. Active control of structural supports is one countermeasure, but distance may be more economical to implement.

VI. ENGINEERING, LOGISTICS, BASE OPERATIONS AND RELATED COSTS

Absolute cost of the first lunar base will depend on its objectives, design, program schedule, flight vehicles employed, etc. At least three aspects of base location can drive early base cost (Table I): availability of sunlight for solar-driven power systems, communications with Earth, and dispersion of base facilities. All costs are heavily influenced by transportation, which, using Saturn V, Energia or Shuttle-C class launch vehicles could range from \$30 to $\$200 \times 10^6$/trip delivered to the lunar surface.

This high cost of transportation is among the strongest incentives to utilize lunar resources early. Perhaps the simplest example is using soil as a radiation

TABLE I
Rough Order of Magnitude Added Cost ($) for Solar Power and Communications at Different Types of Initial Lunar-Base Location

	Polar	Nearside	Limb	Farside
Power	Lowest	10^9 +	10^9 +	10^9 +
Communications	10^6–10^7	Lowest	10^6–10^8	10^7–10^9

shield. The cost of transporting shielding would be prohibitive and the local product is an easily adapted and adequate substitute.

Minimum Earth-import mass for an initial polar base could be roughly 30 t, including equipment for reserve power at subsistence levels for periods of 15 hr (Dowling et al. 1992). Reserve energy need only be stored for eclipse periods, small local terrain obscurations, and equipment breakdowns. For non-polar sites, regenerative fuel cell systems to maintain operations through the lunar night can require adding 25% or more to initial base mass. Nuclear power may be a viable alternative, but development costs are high and political acceptability questionable. Another alternative is beaming power by laser from Earth, where an adaptive optics beam director could illuminate a spot under 100 m across on the Moon where photovoltaics convert laser power to electricity (J. D. G. Rather, personal communication). At higher latitudes, arrays might be arranged down slopes, or spread out in a north-south ellipse to capture the circular beam.

Thermal management will be critical at any base site, with polar sites offering the easiest heat rejection. At lower latitudes, natural features or artificial shields may improve difficult mid-day problems. Some energy storage schemes involve solar heating of masses of lunar rock which liberate energy to a conversion device while cooling through the night (Landis 1989).

Depending on their function, early bases may require anything from intermittent communications of 100 Kbps to continuous high-rate data over 100 Mbps. If Earth is in continuous view, even the latter requirement need not be a major cost driver. Relays are required from far side and polar locations. For sites near the limb, repeaters on high crater rims may be easiest. Communications satellites in lunar or librating Earth-Moon orbits are another possibility, and would be favored for any site well onto the far side. Relay satellites, multiple relay towers, and/or long surface lines could add anywhere from $\$10^7$ to $\$10^9$.

If different facilities at a base must be dispersed, then costs must be accounted to move between the facilities and provide them with power, communications and other services. Local topography will influence these costs, as will distance between facilities. To offer both radio-quiet and Earth-viewing facilities, a single limb "base" might involve separations of as much as 300 km. Dispersion of facilities beyond a roughly kilometer-sized site could add $\$10^7$ to $\$10^9$ or more to an initial base cost.

Polar sites are accessible only from polar orbits. Equatorial sites can be reached from any orbit, but frequent access to a given equatorial site is available only from equatorial orbit. Polar orbits afford access to all sites, but with frequent access only near the poles themselves. An equatorial orbit can only access equatorial sites. The Moon's slow speed of rotation, unlike the Earth, does not significantly influence launch performance to lunar orbit. Assuming the same navigational accuracy at each site, the cost of transporting a given payload is minimally dependent on site location.

Absent any discovery of polar volatiles, if the objective of a base were primarily to set up operations on the Moon and learn from the environment, with the sciences taking second priority, then one of the poles would be a good site. Having to choose, most astronomers would prefer the southern sky, though the terrain near the south pole appears more rugged for local navigation (U. S. Geological Survey 1981).

VII. INFRASTRUCTURE DEVELOPMENT

In terrestrial real estate, location is the most important consideration in the value of land. Over time, the same principle will apply to the Moon. Certain locations will favor specific mineral resources, most so far undiscovered. The poles favor solar energy production, while equatorial sites and the poles are most accessible from orbit. Lunar topography dictates natural choke points for surface transportation, such as Vallis Alpes between Mare Imbrium and Mare Frigoris, and the pass beneath Fresnel Promontory between Mare Imbrium and Mare Serenitatis. Certain natural features favor energy storage schemes involving pressurized gasses, fluids and slopes down which material can move to generate power. A Mare Tranquillitatis site, on the equator at 33° 10′.42 east longitude, is proposed for optimum location of a mass driver launching lunar material to a mass catcher near the L_2 libration point. From this location, trajectories are achromatic, meaning that small errors in along-track velocity at the "muzzle" of the mass driver result in minimal dispersion of payloads' arrival location near L_2 (Heppenheimer 1977).

As the Moon is developed, interconnections will form between regions of local surpluses and deficiencies toward global commerce. Territorial declarations may impede or stimulate infrastructure development, as natural, economic and political factors meld to influence the character of future human habitation. While exciting to ponder, these considerations remain far off from selection of the first few base sites.

VIII. SPECIFIC SITES OF INTEREST

Specific sites have been proposed for lunar bases at least since Goddard's 1920 treatise.

The following example sites, summarized from the NASA/Johnson Space Center Lunar Outpost Site Selection Workshop held 1990 August 13–14 offer a selection dominated by today's science-oriented considerations:

Mare Smythii. 1.7N, 85.8E (Fig. 5). Equatorial limb mare-highland contact. Very high Ti (ilmenite) mare basalts and relatively thin, young regolith (which could reduce gas entrapment). High Al highland soils at basin rim. Thick, mature highlands regolith (possibly good gas entrapment, but without extra retention afforded by ilmenite). Very high Ti pyroclastic glasses abundant (semi-volatiles entrapment). KREEP very rare (low grade deposits at Balmer 600 km away).

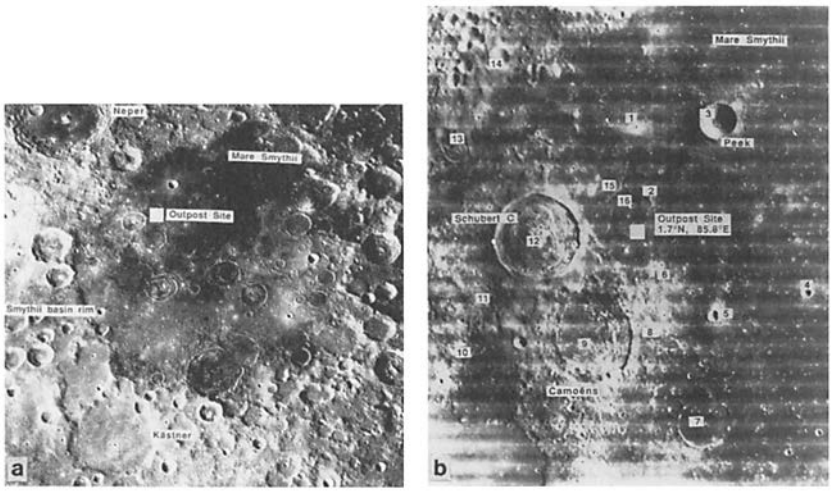


Figure 5. Candidate Mare Smythii site and region. Nearby sites of geologic interest are flagged (courtesy of NASA Johnson Space Center).

Riccioli. 2.5S, 83W (Fig. 6). Near-side limb 140 km crater with mare fill. Unusual volcanic features (speculative vents?). Unknown Ti content in mare floor of crater. H, ^{3}He, other volatiles and Fe, Mg, Ti in mare basalts, possibly in crater ejecta. Dark halo craters possible subsurface gas traps. Aluminum in feldspathic highland rocks associated with Hevelius formation.

Mare Ingenii. 35S, 165E. Far-side high-latitude mare site. Ti concentration unknown. Moderate ilmenite but mature regolith (3.5 Gyr age) for gas trapping. Highlands nearby with abundant anorthite. Radio quiet.

Mare Tranquillitatis. 3.87N, 38.39E. Near-side equatorial mare site. Spectrally "blue" mare basalts. High titanium (ilmenite) mare basalt. Wide range of basalt types and ages with probable wide maturity range.

Aristarchus. 23N, 48W (Fig. 7). Near-side high-latitude mare-highland site. Volcanic constructs. Mare basalts and highland materials, anorthite. Nearby pyroclastic deposit of Fe-rich glass. Large volumes of mature regolith

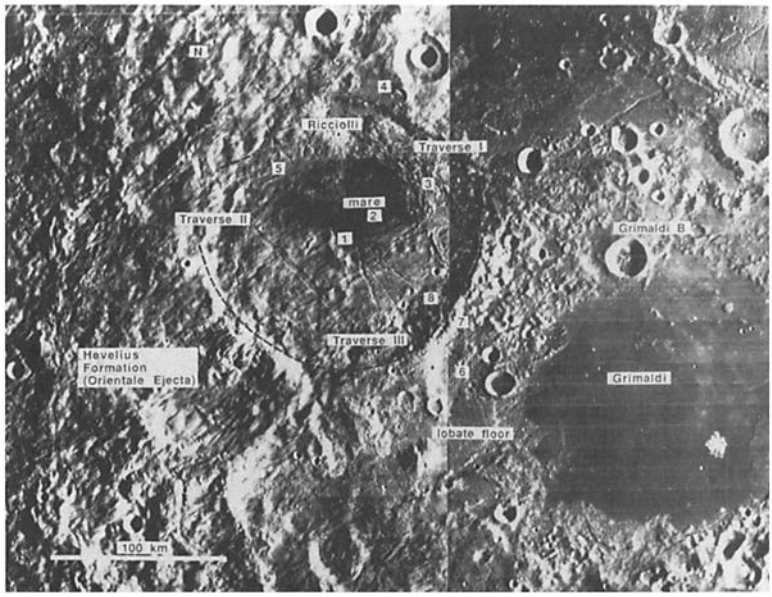


Figure 6. Riccioli crater area showing mare floor and nearby highland (figure courtesy of NASA Johnson Space Center).

for gas entrapment. TiO_2 abundance. Pyroclastics enriched in certain surface-correlated elements (Cu, Pb, Cd, Te, etc.) extreme KREEP-rich or granitic ore bodies possible. Lunar transients (subsurface gas reservoirs?) suggested by Earth observers.

South Pole/Amundsen. 88S, 60E (Fig. 8). Near-side, polar highlands site. Thick, mature highlands regolith-Al probably high, 8 to 10%? Trace element concentrations in KREEP moderate? No pyroclastics identified. Possible polar ice traps. Two potential full-time solar power illumination sites identified. No mare basalts, ilmenite unlikely.

IX. CONCLUSIONS

With roughly the area of Africa, the Moon shows a great diversity of surface features of interest to different scientists and potential users. It should be little surprise that no single candidate base site is best for all uses. Location of the first base sites will depend on the dominant users at the time of selection. Limited financial resources suggest that a single site is most likely for the first base, rather than two or more simultaneous "initial" sites. Therefore, some users will be delayed in their strongest preferences. Nearly any user will be able to find some value in any site, so the first base is likely to serve a combination of users in a loose cluster of specialized and general purpose facilities with separations up to 300 km.

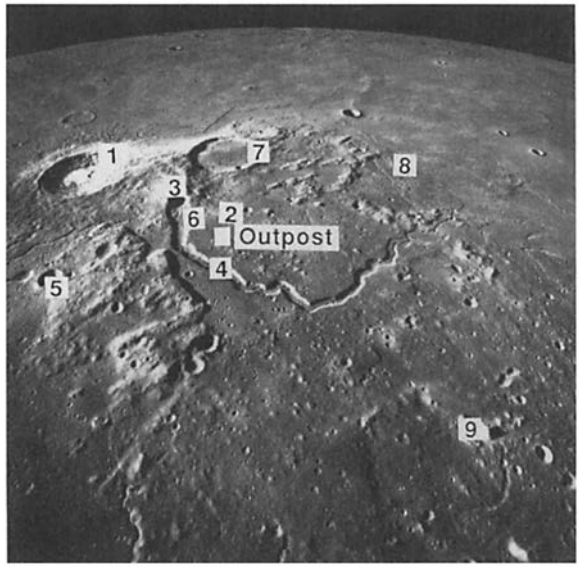


Figure 7. Candidate Aristarchus Plateau base site, showing a wealth of nearby volcanic features (figure courtesy of NASA Johnson Space Center).

With no credible near-term material export resources, the first resource utilization is likely to be in support of surface facilities, with very simple "processing" such as piling up material, sintered bricks and making simple castings. For simple building materials, mare sites are superior, but highland sites will do. Ilmenite-bearing mare soils favor a broad range of candidate processes for harvesting H, C and N, oxygen extraction and other applications, but there are other processes which are not site specific, or for which certain highlands materials appear superior. Once a site is chosen, material resource utilization schemes should be adapted to the site and materials close at hand.

Over time, added base sites will be chosen to suit evolving needs. Some bases will expand, adapt and evolve, eventually becoming seeds for small communities with expanding resource needs. New discoveries, especially of resources, will strongly influence site selection for new facilities.

The most important discovery which could influence base site selection would be a concentrated volatiles reservoir, such as postulated at the poles. Global surface chemical and mineralogical composition mapping is also of high importance, and such maps, along with a polar ice search, could be mounted with any of several proposed lunar polar orbiters (see, e.g., Phillips 1986).

Today's thinking is likely to appear primitive when viewed from the future, but knowing what we do now, we can say that the equatorial limb sites, Mare Smythii and the crater Riccioli, appear the best selection for science and have a broad range of materials within about 100 km. Polar sites are the

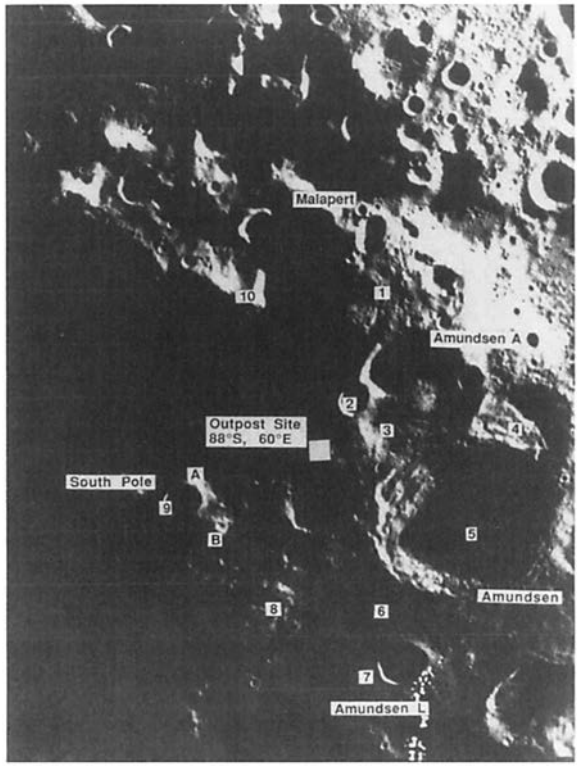


Figure 8. South polar region shows rugged terrain and a candidate base site (courtesy NASA Johnson Space Center).

best for operations, and could afford substantial savings in base establishment costs, but are much less favorable for astronomy and lunar surface science.

Acknowledgments. Portions of the work reported here were performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration. Helpful comments were made by B. N. Griffin and an anonymous reviewer. R. Staehle is also affiliated with the World Space Foundation, and R. Dowling is also affiliated with Space/Media.

REFERENCES

Agosto, W. N. 1985. Electrostatic concentration of lunar soil minerals. In *Lunar Bases*

and Space Activities of the 21st Century, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 453–464.

Burke, J. D. 1985. Merits of a lunar polar base location. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 77–84.

Carrier, W. D., III, Olhoeft, G. R., and Mendell, W. 1991. Physical properties of the lunar surface: Bulk density and porosity. In *Lunar Sourcebook: A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (Cambridge: Cambridge Univ. Press), pp. 483–494.

Carroll, W. F., Steurer, W. H., Frisbee, R. H., and Jones, R. M. 1983. Should we make products on the Moon? *Astronautics and Aeronautics* June, pp. 80–85.

Criswell, D. R. 1984. Lunar power system. In *Lunar Bases and Space Activities of the 21st Century*, ed. M. B. Duke (Houston: Johnson Space Center), p. 117 (abstract).

Dalton, C., and Hohmann, E., eds. 1972. *Conceptual Design of a Lunar Colony* (Houston: Univ. of Houston), pp. 355–362, 407–422.

Dowling, R., Staehle, R. L., and Svitek, T. 1992. A lunar polar expedition. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 1, pp. 175–182.

Duke, M. B., Mendell, W. W., and Keaton, P. W., eds. 1984. *Report of the Lunar Base Working Group* (Los Alamos: Los Alamos National Laboratory).

Garvey, J., and Magoffin, M. 1991. Lunar processing using solar energy—a research project status report. In *Space Manufacturing 8, Proc. of the Tenth Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 143–149.

Goddard, R. H. 1920. In *Papers of Robert H. Goddard, Volume 1*, eds. E. C. Goddard and G. E. Pendray (New York: McGraw-Hill, 1970), pp. 413–430.

Green, J. 1961. *The Geology of a Lunar Base*, Report SID 61-358 (Downey, Calif.: North American Aviation, Space and Information Systems Div.).

Harmon, J. K., and Slade, M. A. 1991. An S-band radar anomaly at the north pole of Mercury. In *Bull. Amer. Astron. Soc.* 23:1197 (abstract).

Haskin, L., and Warren, P. 1991. Lunar chemistry: Solar-wind-implanted elements. In *Lunar Sourcebook: A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (Cambridge: Cambridge Univ. Press), pp. 436–448.

Heiken, G. H., Vaniman, D. T., and French, B. M., eds. 1991. *Lunar Sourcebook: A User's Guide to the Moon* (Cambridge: Cambridge Univ. Press).

Heppenheimer, T. A. 1977. Trajectory dynamics in the Earth-Moon system. In *Space Manufacturing Facilities II*, ed. J. Grey (New York: AIAA), pp. 65–79.

Horz, F. 1985. Lava tubes: Potential shelters for habitats. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 405–412.

Johansen, L., and Kirk, M. L. 1982. Lunar transient phenomena. In *Foundation Astronautics Notebook—5* (South Pasadena, Calif.: World Space Foundation).

Johnson, J. R., Larson, S. M., and Singer, R. B. 1991. Remote sensing of potential lunar resources—I: Near-side compositional properties. *J. Geophys. Res.* 96:18861–1882.

Kopecky, L. and Voldan, J. 1965. *The Cast Basalt Industry*, Annals of the New York Academy of Sciences, vol. 123, pp. 1086–1105.

Landis, G. A. 1989. *Solar Power for the Lunar Night*, NASA TM-102127.

Lanzerotti, L. J., Brown, W. L., and Johnson, R. E. 1981. Ice in the polar regions of the Moon. *J. Geophys. Res.* 86:3949–3950.

Lewis, J. S., Jones, T. D., and Farrand, W. H. 1988. Carbonyl extraction of lunar and asteroidal materials. In *Engineering, Construction, and Operations in Space: Proc. Space 88*, eds. S. W. Johnson and J. P. Wetzel (New York: American Soc.

of Civil Engineers), pp. 111–122.
Meek, T. T., Vaniman, D. T., Cocks, F. H., and Wright, R. A. 1985. Microwave processing of lunar materials. In *Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell (Houston: Lunar and Planetary Inst.), pp. 479–485, esp. p. 483.
Morrison, D. A., ed. 1990. *Developing a Site Selection Strategy for a Lunar Outpost: Science Criteria for Site Selection. Conclusions of a Workshop*, April 2–3, Houston, Solar System Exploration Div., Johnson Space Center).
Papike, J., Taylor, L., and Simon, S. 1991. Lunar minerals: Native Fe. In *Lunar Sourcebook: A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (Cambridge: Cambridge Univ. Press), pp. 151–154.
Phillips, R., ed. 1986. *Contributions of a Lunar Geoscience Observer (LGO) Mission to Fundamental Questions in Lunar Science* (Dallas: Southern Methodist Univ.).
Schmitt, H. H. 1975. The great voyages of exploration. In *Apollo Expeditions to the Moon*, ed. E. M. Cortright, NASA SP-350, pp. 265–288.
Selzer, J. C. 1989. *A Lunar Base Bibliography*, JSC Doc. No. 22873 (Houston: NASA Johnson Space Center).
Senior, C. 1989. Production of Oxygen and Other Products by Pyrolysis of Lunar Materials. Final Report, Contract NAS9-18102; Report PSI-2084/TR-948 (Andover, Mass.: PSI Technology Co.).
Shevchenko, V. 1992. The choice of the location of the lunar base. In *The Second Conf. on Lunar Bases and Space Activities of the 21st Century*, ed. W. W. Mendell, NASA CP-3166, vol. 1, pp. 155–161.
Slade, M., Butler, B., Muhleman, D., and Jurgens, R. 1991. Mercury Goldstone/VLA radar: Part I. *Bull. Amer. Astron. Soc.*, p. 1197 (abstract).
Spudis, P. D., Taylor, G. J., and Hood, L. L. 1989. *Geological Requirements for Lunar Base Operations*. Project Report Contract No. T-7848 P (Houston: NASA Johnson Space Center).
Staehle, R. L. 1983. Finding "paydirt" on the Moon and asteroids. *Astronautics and Aeronautics* November, pp. 44–49.
Triffet, T., Ramohalli, K., and Lewis, J. 1991. *UA/NASA SERC Annual Progress Report* (Tucson: Univ. of Arizona).
U. S. Geological Survey 1981. *Map Showing Relief and Surface Markings of the Lunar Polar Regions*, U. S. G. S. Map I 1326-B.
University of Houston College of Architecture. 1990. In-situ resource utilization in the design of advanced lunar facilities. In *Proc. 6th Annual Summer Conf. NASA/USRA Design Program*, NASA CR-187041 (Alexandria, Va.: NASA/NTIS), pp. 79–87.
Watson, K., Murray, B. C., and Brown, H. 1961. On the possible presence of ice on the Moon. *J. Geophys. Res.* 66:1598–1600.
Wilhelms, D. E. 1987. *The Geologic History of the Moon.* U. S. Geolological Survey Prof. Paper 1348.
Wittenberg, L. J., Santarius, J. F., and Kulcinski, G. L. 1987. Lunar source of ^{3}He for commercial fusion power. *Fusion Tech.* 10:167–78.

PART III

Near-Earth Objects

ON THE SEARCH FOR NEAR-EARTH ASTEROIDS

JACK DRUMMOND and DAVID RABINOWITZ
University of Arizona

and

MARTIN HOFFMANN
Observatorium Hoher List der Universitäts-Sternwarte Bonn

The distribution of known asteroids in Earth-approaching orbits is a biased sample, but does lead to predictions where objects may be found. To date, searches for near-Earth asteroids have concentrated on the opposition direction, but it is obvious that as systems push to fainter limiting magnitudes more asteroids will be discovered closer to the Sun. These will include Earth-approaching asteroids observed away from the Earth and objects interior to the orbit of the Earth that will never pass through opposition. Least-squares fits to the density distribution of known near-Earth asteroids are presented as preliminary functions to be debiased in order to obtain a true distribution. Some of the biases are reviewed. It appears, especially if searches are made closer to the Sun, that with some reasonable assumptions, the biases that exist may be described, leading to a better understanding of the dynamics that result in the observed distribution of Earth-approaching asteroids. CCDs promise to supplant photography as the preferred search method.

Until recently, near-Earth asteroids (NEAs) were only discovered accidentally, but as programs deliberately designed to detect close asteroids have reached full operational capability the number discovered has increased dramatically. Although of immense popular appeal because of their inferred apocalyptical portent, the NEAs are also of interest to astronomers and meteoriticists, for they are the smallest asteroids that can be studied from the Earth and have a direct relation to the meteorites available for laboratory study. Furthermore, among the Earth-approaching asteroids are extinct or dormant comets, missing links that could elucidate the relationship between asteroids and comets. As potential resources for human exploitation, NEAs to date remain largely terra incognita. Therefore, both quantitatively and qualitatively, that is, in their numbers and in specific cases, near-Earth asteroids deserve special study and specialized searches.

In this chapter, we do not recount the detailed history of NEA searches, but instead concentrate on the distribution and biases of observed NEAs. In the first section, we develop a detailed distribution density function describing where in the sky NEAs can be found from a geocentric point of view, based on the known, not extrapolated, population, and comment on its implication

for searches. Next (Sec. II), we discuss some of the biases that exist in this observationally determined set of NEAs. A brief outline of the future direction of NEA searches with CCDs is presented in Sec. III, followed by some conclusions on the topics in Sec. IV.

I. ORBIT DISTRIBUTIONS

There are two approaches to searching for the generally small (~1 km) asteroids that glide swiftly by the Earth. The first, and most commonly used, is to search around the opposition point where the asteroids will be near a minimum in phase angle, and therefore the brightest, and will be moving the fastest, more than a degree per day. Even slow, distant objects are brightest and moving fastest near their opposition. The second, and to date untried, approach, is to search the apex and antapex of the Earth's way, tangent to the Earth's orbit, where the asteroids that will be overtaken by the Earth or will overtake the Earth, respectively, can be seen moving the slowest. A third method, introduced here in the last paragraph of this section, is to search a particular volume of space from anywhere in the Earth's orbit.

To illustrate the distribution of NEAs, and where to search for them, Figs. 1–8 are provided. First, Fig. 1 shows the distribution of orbital elements, compared two at a time, for all 157 numbered and unnumbered asteroids (supplied by J. Scotti) with perihelion distances <1.3 AU (the working definition for NEAs) through the end of 1990 and the three meteorites with known orbits calculated from multiply photographed falls. The strongest correlation (r^2 = 0.21) appears between q and e. This is partly an observational bias, but partly reflects a real correlation in the distribution. It is doubtful, for instance, that there are as many low-eccentricity asteroids near Mercury or Venus as there are near the Earth because of dynamical considerations, and the low e, low q region in Fig. 1 would remain depleted even if the observed distribution were debiased.

The lower three panels in Fig. 1, where the orientation of the orbits are compared to q, e and i, illustrate no real correlation. Figure 2, then, can be produced to indicate the observational distribution of NEAs. Choosing a night at random (April 4), 360 1-km diameter S asteroids, with albedos of 0.186 (Tedesco et al. 1989), $H = 17.5$, were placed in every orbit at one degree intervals of mean anomaly, and the apparent magnitudes calculated using a Lumme and Bowell phase slope of 0.23 (Lagerkvist and Magnusson 1990). Figures 2a, 2b and 2c show the positions of the asteroids to limiting magnitudes of 15, 17.5 and 20, respectively, in ecliptic coordinates centered on the opposition point. Alternatively, Fig. 2 can be viewed as representing a range in sizes of objects. If the middle panel represents 1-km objects at apparent magnitude 17.5 and brighter, then 2a and 2c show objects $\sqrt{10}$ times smaller and larger, respectively. Some interesting considerations become apparent from Fig. 2. First note that for bright limiting magnitudes or for smaller objects (Fig. 2a), the best place to search is over all the sky centered

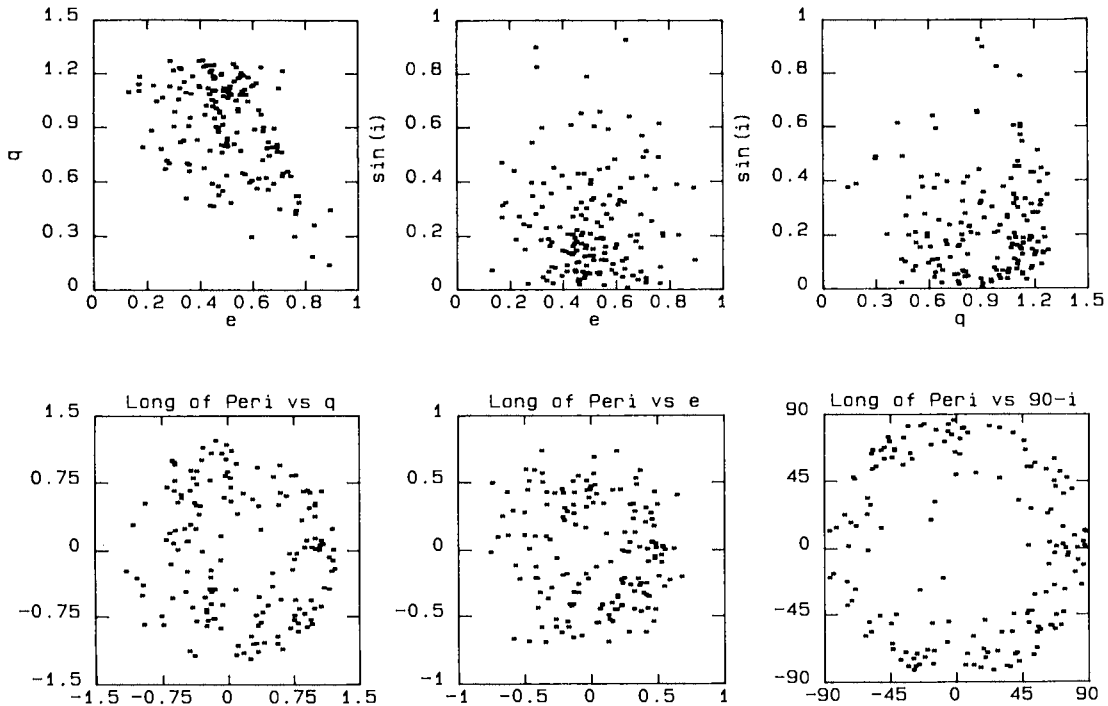
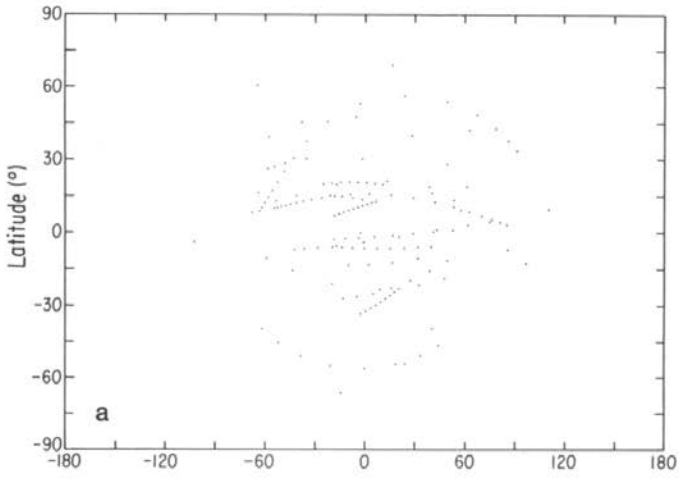
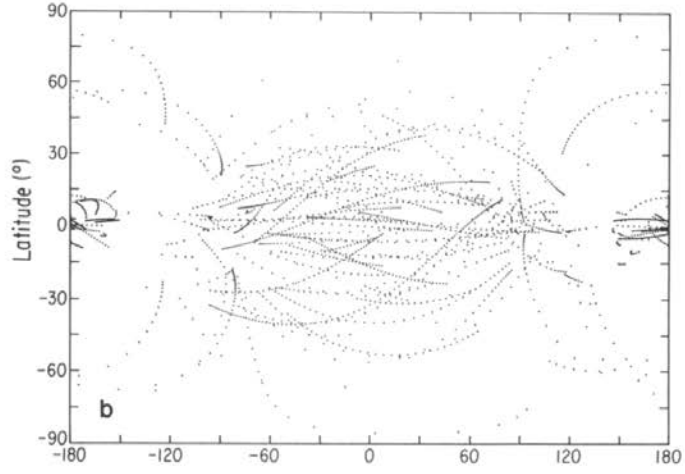


Figure 1. The distribution of orbital elements of 157 NEAs and three meteorites with known orbits. Pairs of orbital elements (or functions of elements) are compared two at a time. For the bottom panel involving the longitude of perihelion, the longitude is measured counterclockwise from the right, while the other parameter is measured from the center outward according to the scale (figure from Drummond 1991, but extended through 1990XJ).


90
60
30
0
-30
-60
-90
Latitude (°)
a
-180
-120
-60
0
60
120
180


90
60
30
0
-30
-60
-90
Latitude (°)
b
-180
-120
-60
0
60
120
180

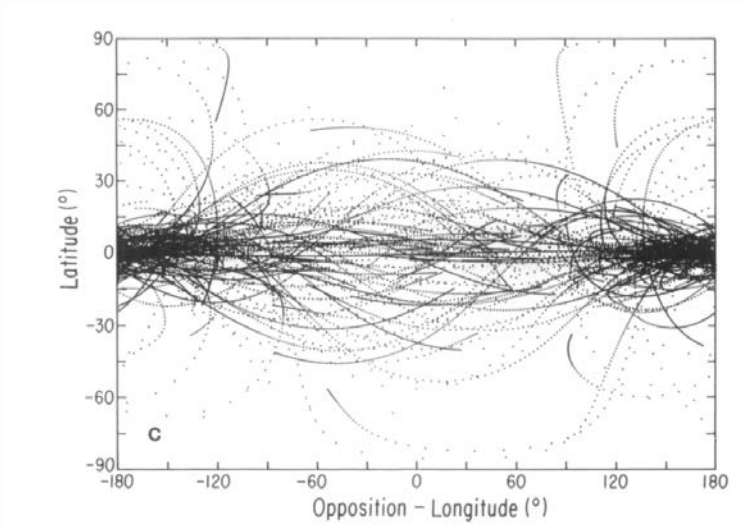


Figure 2. The distribution of NEAs on a random night, plotted in ecliptic coordinates with respect to the longitude of the opposition point (evening is to the right, morning to the left). 360 1-km S asteroids are placed in the 160 orbits of known objects at one degree intervals of mean anomaly (thus taking into account dwell time), and those brighter than apparent V magnitude 15 (panel a), 17.5 (panel b) and 20 (panel c), are plotted. Note that a system with a bright limiting magnitude detects 1 km objects in the known orbits only within 90° of the opposition point, whereas 1 km objects can be detected almost anywhere along the ecliptic, and especially near the Sun, by a system with a limiting magnitude of 20. Alternatively, for a fixed limiting magnitude of 17.5, the three panels, a, b and c, show an upper diameter limit of 0.32, 1 and 3.2 km. Thus a system with a limiting magnitude of 17.5 would detect 3.2 km objects all along the ecliptic, with increased concentrations near the Sun (panel c). As another example, for a system with a limiting magnitude of 22.5, panels a to c show the positions of asteroids in known orbits smaller than diameters of 32, 100 and 320 m.

on the opposition direction. However, for larger asteroids or for systems with faint limiting magnitudes (Fig. 2c), the best place to search (except for one glaring problem) would be around the Sun where many asteroids are at superior conjunction and bright because of a minimum in phase angle. The intermediate case (Fig. 2b) is the most instructive, for it shows where not to look for asteroids. A definite minimum in concentration occurs at 45° from the Sun, where the phase angles and distances conspire to render asteroids faint in these parts of their orbits. Nevertheless, even in Fig. 2b a concentration near the Sun can be seen. Around opposition, away from the Sun, and scattered all over the sky, but not particularly concentrated toward the ecliptic, are the asteroids that are passing through perihelion just outside the orbit of the Earth. The apex and antapex of the Earth's way appear as radiants for many of the orbits, causing a slight but noticeable increase in the density of points 90° from the Sun in Fig. 2b.

Figure 3a shows the probability density P for all 57,600 points in the 160 orbits, (360 points per orbit), in 3600 equal-area cells of 0.2 square degrees, and Fig. 3b shows a least-squares fit using the first few relevant terms in a spherical harmonic expansion according to

$$P = \exp\{a_0 + a_1 \cos(\lambda)\cos(\beta) + a_2[\cos(\lambda)\cos(\beta)]^2 + a_3 \cos^2(\beta)\} \quad (1)$$

where λ and β are the ecliptic longitude of the opposition point minus longitude of the asteroid, and the ecliptic latitude for each point, respectively, and

$$a_0 = -15.162,\ a_1 = -0.7288,\ a_2 = .06852,\ a_3 = 9.6818. \quad (2)$$

Such a fit describes the probability of finding an orbit point in a cell to no worse than 0.3% of the total observed distribution, with a standard error over all the sky 10 times smaller.

Taking the brightness of the asteroids into consideration, a probability distribution can be formulated as a function of magnitude V as well as position with respect to the opposition point. Qualitatively, the brightest 2% of the asteroids (brighter than $V = 17.5$ for 1 km objects, brighter than 22.5 for 100 m objects, etc.) can be found centered around the opposition point. There is an abrupt transition at this magnitude, where most of the remaining objects can be found centered on the Sun. The following mathematical probability distribution performs poorly in accounting for this transition, but does well for the brightest 2% and the faintest 50% of the observed orbit points. Even in the worst case in the transition from opposition to Sun-centered objects, the deviation between the model and the observed distribution is 0.8%, and overall the model describes the probabilty of finding a 1 km object at a given position and magnitude to within a standard error of 0.04%. This probability per square degree per magnitude for 1 km S objects is given by

$$P = \exp\{a_0 + a_1 \cos(\lambda)\cos(\beta) + a_2[\cos(\lambda)\cos(\beta)]^2 + a_3 \cos^2(\beta) + a_4 V + a_5 V^2 + a_6 V \cos(\lambda)\cos(\beta)\} \quad (3)$$

where $a_0 = -44.132$, $a_1 = 1.900$, $a_2 = -0.319$, $a_3 = 3.950$, $a_4 = 3.176$, $a_5 = -0.0772$, $a_6 = -0.0977$. For a given size distribution, it is only necessary to integrate Eq. (2) and multiply by a size distribution function to obtain the total probability for finding objects in a given magnitude and size range for a given area of the sky.

If the heliocentric XYZ coordinates of all 57,600 points are converted to $R\theta$ coordinates, thus putting every point on the meridian from the Earth's point of view, the distribution of orbit points within 0.5 AU of the Earth's orbit is shown in Fig. 4a, the distribution in the RZ plane. The resulting probability distribution is shown in cells of 0.015625 square AUs in Fig. 4b. R is the distance from the Sun to the projection of the point onto the ecliptic plane

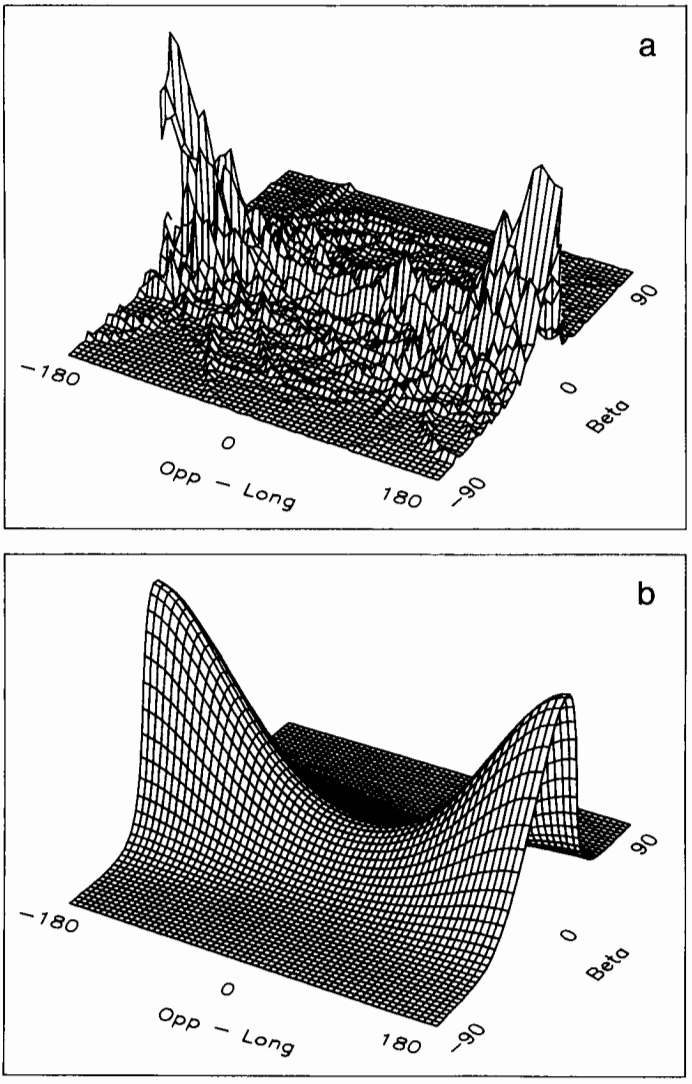
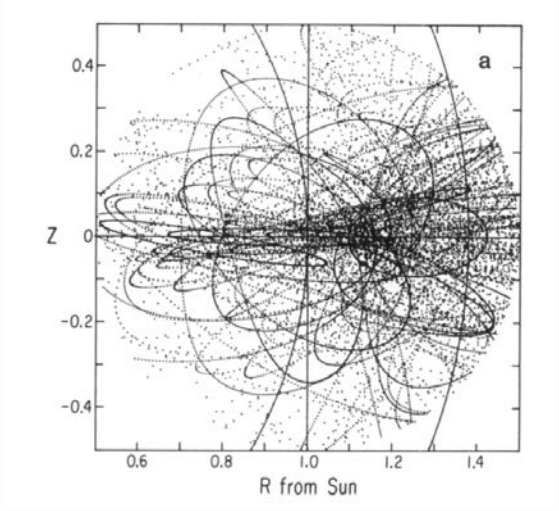
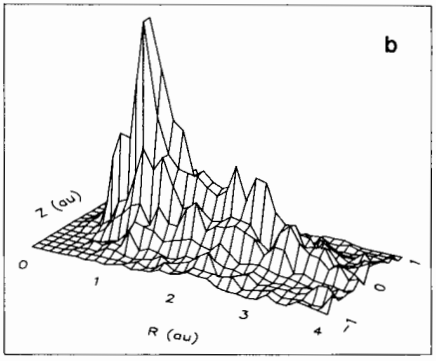
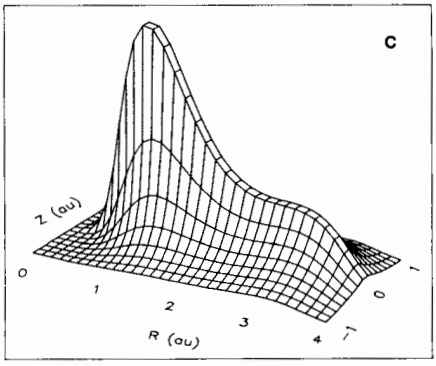


Figure 3. (a) The probability or density distribution of all 57600 orbit points (160×360) of known orbits of NEAs and (three) meteorites, plotted as in Fig. 2, but in equal area cells of 0.2 square degrees. (b) The theoretical distribution according to Eq. (1).


a
0.4
0.2
Z 0
-0.2
-0.4
0.6
0.8
1.0
1.2
1.4
R from Sun


b
Z (au)
0
1
2
3
4
R (au)
0
1


c
Z (au)
0
1
2
3
4
R (au)
0
1

($R^2 = X^2 + Y^2$) and θ is the heliocentric ecliptic latitude [$\theta = \arctan(Z/R)$]. Figure 4c shows the least-squares fit according to

$$P = \exp[a_0 + a_1|\theta| + a_2 R_\odot + a_3 R_\odot^2 + a_4 R_\odot^3 + a_5 R_\odot^4] \qquad (4)$$

where P is the probability per AU^2, $R_\odot^2 = X^2 + Y^2 + Z^2$ is the heliocentric distance squared, and $a_0 = -7.068$, $a_1 = -6.421$, $a_2 = 16.184$, $a_3 = -12.420$, $a_4 = 3.776$, $a_5 = -0.4049$. The worst deviation is 0.6% and the standard error is 20 times smaller. This probability distribution and Eqs. (1) and (2) are the functions which must be debiased to extract the true distributions of the NEA population. Figure 4, for instance, shows that the observed probability peaks strongly just outside the orbit of the Earth, between 1.125 and 1.25 AU, and within 0.0625 AU of the ecliptic plane. Orbits that pass through this region stand the greatest possibility of being discovered. (For information on debiased population estimates see Shoemaker et al. [1990] and the Chapter by Davis et al.)

Another illustration of the visibility of NEAs is given in Fig. 5, where asteroids are distributed at random in orbits of real NEAs and main-belt asteroids (but with randomized arguments of perihelion and longitudes of ascending node) and according to a distribution of sizes e^H, where H is the absolute magnitude. The range of H for main-belt objects is 3 to 23 and for NEAs 10 to 30; only objects appearing brighter than visual magnitude 20.5 (using a phase function of 0.1) are plotted. The probability function for the NEA component of this distribution should correspond to the function that would be obtained by multiplying Eq. (2) by e^H and integrating from $V = -\infty$ to 20.5 and from $H = 10$ to 30. Notice that the background of main-belt asteroids concentrate toward the opposition point, while the NEAs scatter more evenly about the sky.

Rates of apparent motion can be used to discriminate NEAs from main-

Figure 4. The probability or density distribution of the 57,600 orbit points within 0.5 AU of the Earth, plotted in RZ coordinates. (a) All of the individual orbit points are plotted, with positive Z being above the ecliptic plane. Note the concentration of dots between 1 and 1.2 AU, in and just above the ecliptic plane, indicating that most orbits pass through this region or, much more likely, that there is a discovery bias favoring objects that pass through this region. The two curved lines through the field are at distances of 1 and 1.38 AU from the Sun (the perihelion distance of Mars). (b) The relative density of points in cells of 0.015625 AU^2. Again note the concentration of points near 1 AU in the Ecliptic plane, but also the presence of secondary peaks in the asteroid belt between 2 and 3 AU. (c) The theoretical distribution according to Eq. (3).

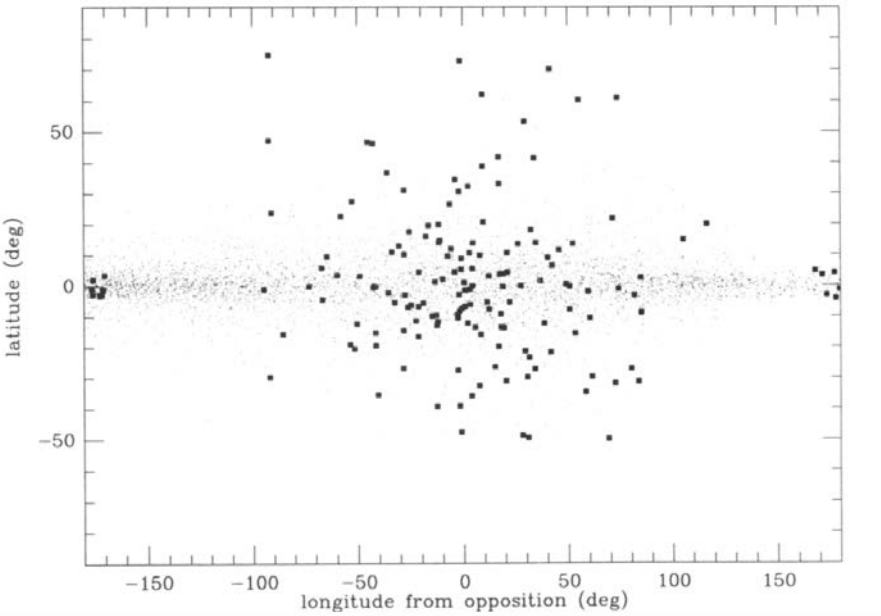


Figure 5. The distribution of NEA (boxes) and main-belt (dots) asteroids according to the theoretical size distribution given in the text.

belt objects. Retrograde NEAs moving faster than $>0.^{\circ}5$ day^{-1} at opposition are easily distinguished by their greater rates of motion. All prograde asteroids are identified as NEAs. Figure 6a shows the rates less than $0.^{\circ}5$ day^{-1} near opposition for a simulated population of asteroids such as is shown in Fig. 5, and illustrates that the slowly moving NEAs can be distinguished by their direction of motion. Figure 6b shows angular rates at opposition vs apparent magnitude for a simulated population of NEAs only. It is apparent that a system which can identify slowly moving ($<0.^{\circ}5$ day^{-1}) NEAs fainter than 17th magnitude by their direction of motion also samples the densest area of the distribution, and will identify the greatest number. These conclusions are being verified by the Spacewatch Telescope, which is finding slow, large NEAs farther than 1 AU from the Earth.

Rather than search in the opposition direction, or in the direction of the apex and antapex of the Earth's way, a third way is to conduct searches along specific orbits. The four asteroid/fireball streams of Drummond (1991) and Halliday et al. (1990) lend themselves to a search of a fixed volume in space throughout the year. (In an updated analysis, 1991FA has been added to the third stream, bringing to five the number of asteroids in each of the first three streams.) The bottlenecks that occur near perihelion in the orbits of the individual members are about 0.1 AU in diameter for the asteroids of the first three streams, and occur when the asteroids are traveling at 35 km s^{-1}. (The

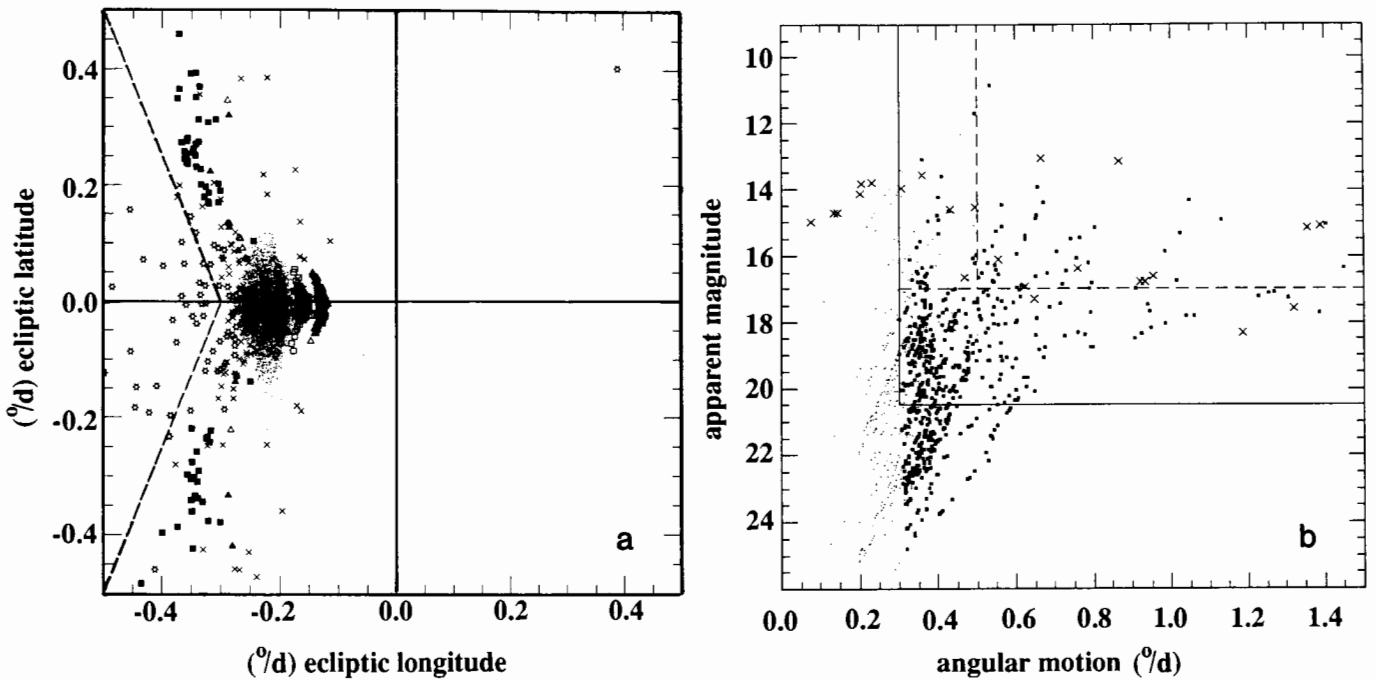


Figure 6. (a) Rates of motion for the theoretical NEA population shown in Fig. 5 and discussed in the text. Also shown are the expected rates for other asteroid types. Many NEAs move faster than 0°.5 per day, but the figure only covers slower rates. NEAs are depicted with open stars and are the only objects to the left of the dashed line, indicating that they can be uniquely identified by their direction of motion. Other kinds of asteroids depicted are Main Belt (dots), Trojans (open triangles), Phocaeas (solid triangles), Hildas (open squares), Hungarias (solid squares) and Mars crossers (X). (b) Apparent magnitude vs rates of motion for a theoretical distribution of NEAs only. Solid squares (prograde) and Xs (retrograde) show NEAs that can be identified strictly by their motions (from Rabinowitz 1991).

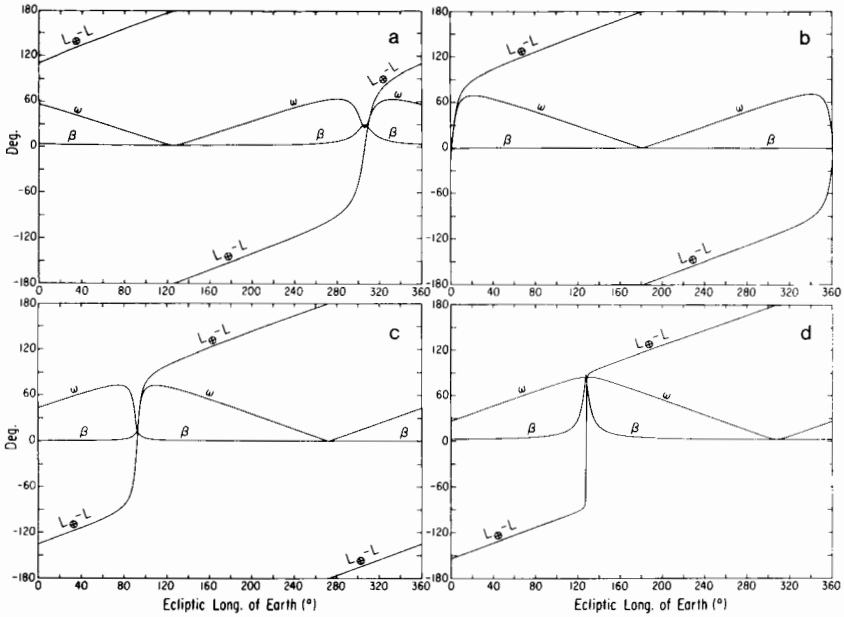


Figure 7. Observability of four (a–d) asteroid stream bottlenecks. Each panel shows three continuous curves, ecliptic longitude with respect to the opposition point ($L_\oplus - L$), ecliptic latitude (β), and solar phase angle (ω) throughout the year (ecliptic longitude of Earth). For example, the scrunch point for Association I (plot a) reaches a maximum in latitude ($\beta = 29°$) on July 29 (Ecl. Long. of Earth = 306°), when it is at opposition ($L_\oplus - L = 0°$), and the phase angle reaches a secondary minumum of 29°. Most of the year the scrunch point is at very low latitudes. The phase angle's deepest minimun occurs when the scrunch point is behind the Sun. The three associations are the asteroid streams of Drummond (1991), while the fourth is for a stream defined by five fireballs (Halliday et al. 1990) and 1989DA; the observability parameters are calculated assuming the characteristics of the three other asteroid streams and applying them to the orbit of 1989DA.

fourth stream, which is comprised of five fireballs and one asteroid, 1989DA, is given the same parameters centered on the orbit of the asteroid.) The density of the stream is maximum here, where the individual orbits pinch together and the flux through the volume is greatest, presenting the best opportunity for the discovery of new members. Figure 7 shows the ecliptic coordinates with respect to the opposition point for the fixed scrunch points or regions during the course of the year. Note that the scrunch point is generally very near the ecliptic, only changing in latitude as it approaches opposition. For objects passing through this volume, the solar phase angle shown on the same plot reaches two minima, with the lowest usually occurring when the scrunch point is at superior conjunction. In Fig. 8, the apparent diameter of the fixed volume is generally around 4° for most of the year, but shoots up to 40° at opposition for Stream I, and up to 88° for Stream III. The apparent

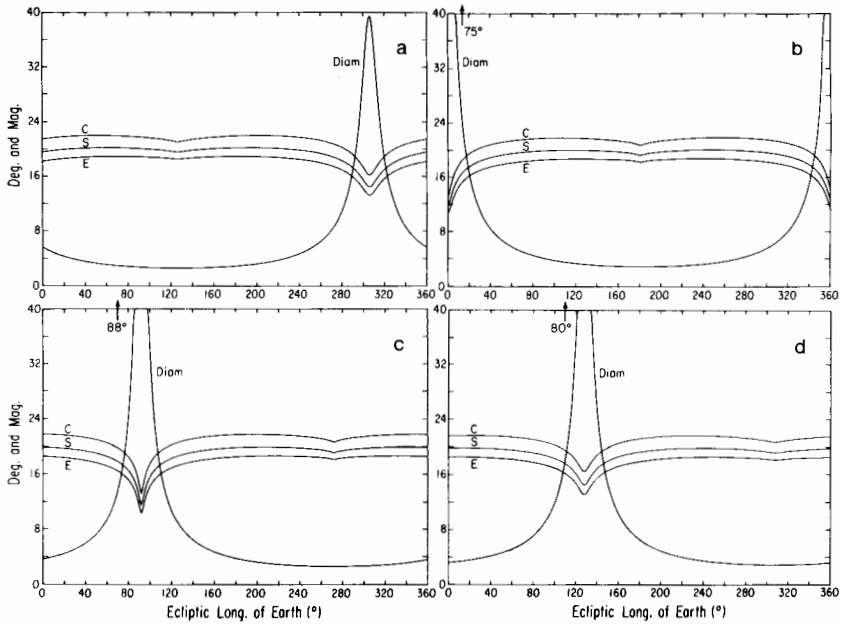


Figure 8. Apparent size of each (a–d) scrunch region and magnitude of a 1-km object passing through it. Throughout the year the 0.1 AU diameter of the bottlenecks have an apparent diameter of 4°, but can shoot up to more than 80° when the Earth is close to the scrunch point. The apparent magnitudes of a 1-km C, S and M asteroid are generally around magnitude 20 throughout the year.

magnitudes of a 1-km C, S and E object are shown to be rather steady until a month or two around opposition. This suggests that systems with bright limiting magnitudes but large fields of view should search for new members passing through the bottleneck at opposition, while systems with small fields of view but faint limiting magnitudes should follow the scrunch regions at other times of the year. The rates of motion for objects passing through all three regions is always greater than 1° per day, regardless of when the area is searched, allowing them to be distinguished from main-belt asteroids.

II. NEA BIASES

Both their accessiblity (represented by the distribution of orbital elements) and their composition (determined by the distribution of their sources—comets or asteroids) are important if NEAs are to be understood as carriers of resources. The result of the previous section, where the distribution of asteroids follows from the biased distribution of their detection location, allows for many possible true distributions of orbital elements. Even simple observational preferences and biases can skew any of several theoretical asteroid distributions to produce the observed distribution. Furthermore, the small

number of searches at small elongations from the Sun leave a considerable volume of space and orbital phase largely unexplored. On the other hand, a consideration of the efficiencies of the sources and sinks of NEAs, and of physical parameters and processes involved, does provide some constraints on the search for and detection of NEAs.

Today's distribution of NEAs is a composite distribution with varied origins. Aside from possible, but as yet undetected, fossil asteroids left over from the creation of the inner solar system (vulcanoids—planetesimals interior to Mercury, and Trojans of Mercury, Venus and Earth), the fraction of NEAs of cometary or asteroid-belt origin remains uncertain. Both of these groups should be subdivided in considering orbital element distributions. Extinct comets can be de-coupled from Jupiter family orbits by multiple dynamical interactions with the terrestrial planets (Wetherill 1991) and by nongravitational forces. Even if comets cannot survive devolitization in the inner solar system without being turned to dust, they can survive as asteroid-like objects in the outer, less irradiated parts of the solar system. Such objects, as represented by Chiron, may dynamically evolve inward. Most NEAs arrive from the main belt by inward cascade through orbital resonances, passing through a Mars-crossing stage. Perhaps a few may begin the inward journey in a collision.

These different (cometary/main-belt) original groups differ primarily in their diameter frequencies (Hughes 1990; Gradie et al. 1989) and in their orbital eccentricities and inclinations. Their inward evolution does not change the slope of the diameter frequency of the objects but acts as a neutral filter for the number of a given diameter that penetrate certain dynamical regimes. Therefore, the diameter of the largest member, for instance, of the classical subgroups of inner solar system asteroids gradually decreases with decreasing mean heliocentric distance for the group. For Mars crossers, Amors, Apollos and Atens, the largest asteroids are, respectively, 313, 1036, 2212 and 2100, with diameters (Tedesco 1989) of 101, 41, 9 and 3 km.

Another characterization of different dynamical regimes is given by a modification of the classical Tisserand invariant, which is usually expressed in terms of the restricted three-body system Sun-Jupiter-object

$$T = a_J/a_A + 2\sqrt{a_A(1 - e_A^2)/a_J} \tag{5}$$

where a_J and a_A are the semimajor axes of Jupiter and an asteroid, and e_A is the eccentricity of the asteroid. Replacing Jupiter by any of the terrestrial planets, the Earth in this instance, a significant number of NEAs cluster close to $T = 3.0$, indicating its dominating influence, while objects with $T < 3.0$ have a more scattered distribution.

Because different groups of NEAs have different rates of motion and dwell times in the observability windows, the observational selection effects for the groups are quite complicated. Cases like 1685 Toro exist, whose 5:8 ratio of orbital period to the Earth's excludes close approaches over large

areas of the sky. Resonant coupling of some orbital elements is responsible for other kinds of avoidance of close approaches to the Earth by some NEAs.

Some asteroid/comet dichotomy is reflected in the dynamical classification of Milani et al. (1989), although the existence of several transition paths between the classes introduces ambiguities with respect to evolutionary interpretations. Fine structure in the original distribution of new or long-period comets, such as concentrations of perihelion directions (Fernandez and Ip 1991), become entirely averaged during the dynamical evolution into stabilized near-Earth objects. As so little is known about the possible ranges of mineralogies of residual cometary cores, or of asteroids, surface spectral characteristics cannot be safely used as a test of cometary vs asteroidal origin. Rotation frequencies of asteroid groups close to resonances in the main belt, where the path to the near-Earth environment begins, have not been studied sufficiently to compare with NEAs. The same holds true for asteroid shapes and pole orientations. However, such comparisons do provide hope that the transformation processes from main-belt to Earth-approaching objects may be understood in the future in such a way that links can be identified between well-defined subgroups in the main belt and the NEAs. Even in the biased distributions some significant fine structure can be recognized and should be taken into account when attempting to model the real distribution. NEA perihelia and aphelia close to the terrestrial planets are underabundant (Fig. 9). For Mars the zone of greatest depletion (Fig. 9b) is at its aphelion where a gap in asteroid aphelia occurs. The semimajor axes of mean motion resonances have a tendency to be overabundant for objects with higher eccentricties (Apollos), while for Amors the Kirkwood gaps at the 3:1 and 5:2 resonances are depleted.

Our knowledge of the distribution of orbital elements of NEAs is based on chance discoveries and systematic searches centered on the opposition point. Only one object has been found by groundbased observations at solar elongations less than 90° (Hahn 1988) and the mean elongation at the time of their discovery of all NEAs through February 1988 is 149°.4. Figure 10 shows the location and motion at the times of discoveries for all fast-moving objects discovered within 60° at opposition reported on the IAU Circulars from 1981 through mid 1991. Dominated by the Palomar Schmidt surveys of E. M. Shoemaker, C. S. Shoemaker and E. F. Helin, the distribution corresponds to Fig. 2 with a system limiting magnitude of 14.5 for 1-km-sized asteroids, or alternatively, using the estimated average 16.5 limiting magnitude for their system, Fig. 10 suggests that their program is discovering the 400 m objects and larger. The limiting magnitude of 20.5 for the Spacewatch Telescope software detection implies that it would discover primarily 60 m objects and larger.

However, two man-made factors contribute to the nonisotropy of detections. First, because all objects are brightest near the opposition point, most surveys follow the route to easiest detection and concentrate in this region. Second, the window of observability in dark time at low air mass is much

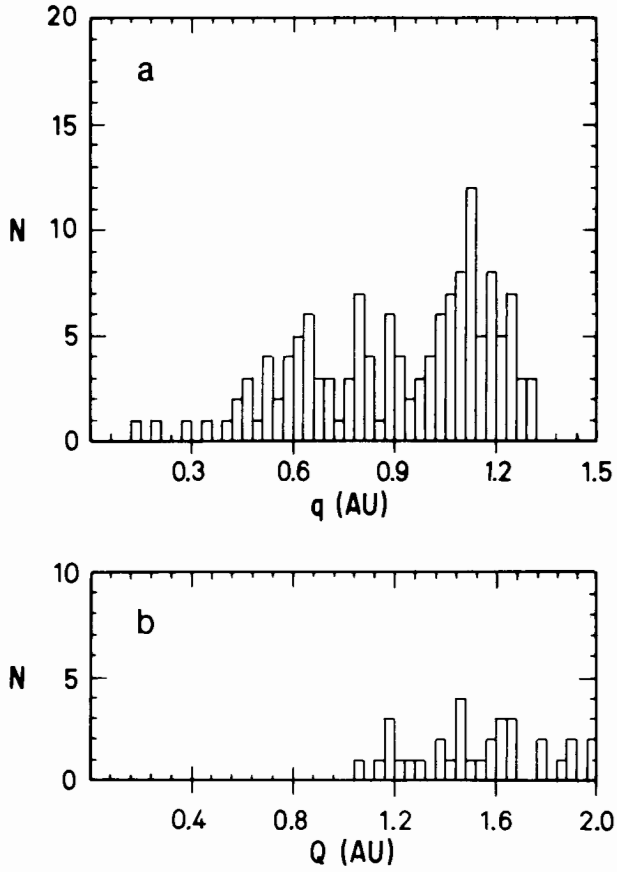


Figure 9. Histogram of NEA perihelion (a) and aphelion (b) distances. Note the gaps at the distances of the terrestrial planets: Mercury a = 0.39 AU, Venus 0.72, Earth 1.0, and near the aphelion distance of Mars at 1.67 (b).

smaller at low solar elongations. Furthermore, searches tend to concentrate towards the plane of the ecliptic and are influenced by the Milky Way. Because of such strong selection effects, the real distribution of NEAs may differ somewhat from the observed one. One way to model the selection effects is to assume an observing preference proportional to the $\sin l/2$, where l is the solar elongation. This accounts in a rough way for the observing time available at small solar elongations for observatories at intermediate latitudes. Another simple approach is to assume that all detections are made between fixed distances from the Sun, which is the case for a given observing station and a limiting size/magnitude, and then calculate the dwell time in this increment for various a's and e's. Table I gives such dwell times in fractions of orbital period for objects between 1 and 1.2 AU from the Sun. But even

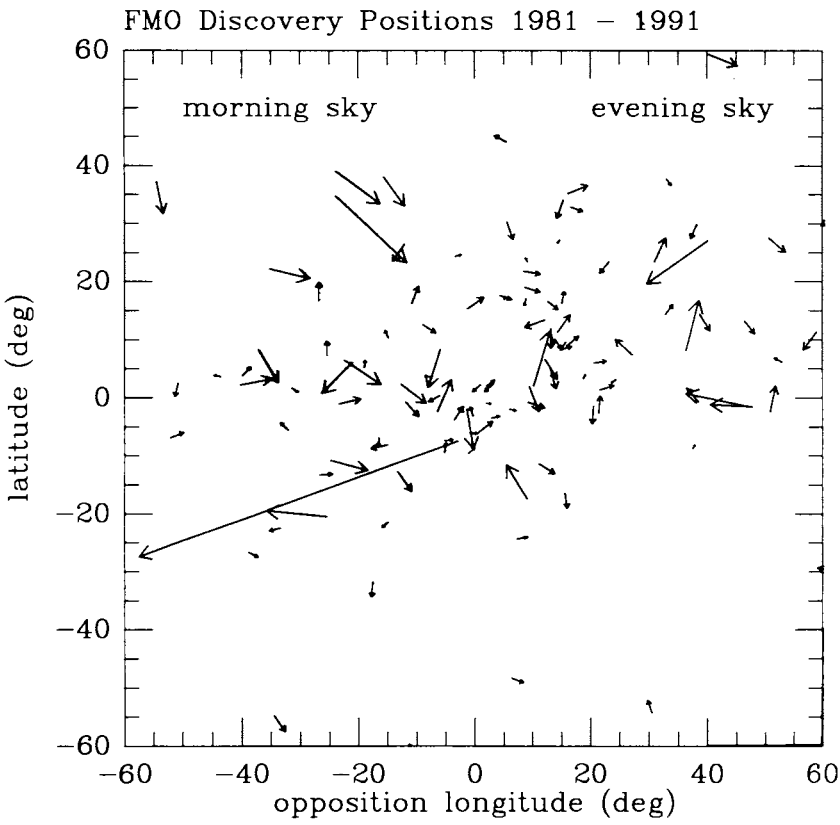


Figure 10. The discovery positions and angular velocities of NEAs discovered between 1981 and 1991. The longest velocity vector belongs to tiny 1991 BA, which was moving at about 1 arcsec per second at its discovery.

the dwell time as a measure of detectibility has to be modified by the angular velocity of the object with respect to any pre-set tracking rate. It is possible to use this angular velocity, or even simpler, direct or retrograde motion to help discriminate classes of NEAs. For example, if a uniform distribution of asteroids is assumed with $i = 0°$ and $0.6<a<4.0$ then of those asteroids observed to move in a prograde direction at opposition, 37% are Apollos, 55% are Amors, and 8% are Mars crossers.

Because many different distributions of orbital elements, sizes and reflectance properties can account for the observed distribution, it is not possible to derive a unique distribution from the observed set, because there are more variables than observed parameters. Furthermore, many of the relevant variables are not readily available, such as the total time a volume of space is searched to a given limiting magnitude, the results of negative searches, etc. However, it is possible to test models of orbital elements and size or reflective

TABLE I

Fraction of Orbital Period within 1.0 and 1.2 AU

		Eccentricity									
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.7	-	-	-	-	-	.254	.380	.466	.272	.211
	0.8	-	-	-	.239	.384	.471	.245	.189	.157	.135
	0.9	-	-	.365	.468	.276	.197	.157	.131	.113	.100
	1.0	1	.532	.564	.257	.184	.144	.119	.102	.089	.078
	1.1	1	.726	.300	.196	.146	.116	.097	.083	.072	.064
a	1.2	1	.463	.285	.171	.125	.099	.082	.070	.061	.054
(AU)	1.3	0	.195	.310	.166	.113	.087	.072	.061	.053	.047
	1.4	0	0	.197	.189	.107	.079	.064	.054	.046	.041
	1.5	0	0	0	.191	.107	.074	.058	.048	.041	.036
	1.6	0	0	0	.126	.117	.071	.054	.044	.037	.032
	1.7	0	0	0	.039	.145	.069	.050	.040	.034	.029
	1.8	0	0	0	0	.110	.071	.048	.037	.031	.027

property distributions against observations by holding at least one parameter constant. Such tests indicate that the assumption of a uniform distribution in a, e and i within their possible ranges, and an exponential distribution in absolute brightness (e^H) or a power-law distribution of diameter (d^k) leads to a good representation of the observed distribution of discovery parameters. However, because some volumes of parameter space have not been fully explored, (e.g., nearer the Sun, extremely small sizes, etc.), extrapolation to these regions is not necessarily warranted. In fact, recent discoveries by the Spacewatch Telescope suggest that there is an excess of small (<100 m) objects over an extrapolation from the power-law distribution observed with that telescope for larger NEAs (Rabinowitz 1992,1993).

No object is known to move permanently inside the orbit of the Earth. Making even a slight extrapolation from observed eccentricities and semimajor axes, however, indicates that such objects are to be expected. Furthermore, on dynamical grounds it is predicted that such orbits have been or will be achieved by a secular decrease of the eccentricities of Aten asteroids, or by a decrease of the semimajor axes of Apollos and Amors through close encounters with Mercury and Venus. An example of an infraterrestrial object is a fireball (no. 331) found by Halliday et al. (1989) to be moving in an orbit that just barely crosses the orbit of the Earth at its aphelion. Another hypothetical group of such objects are the vulcanoids, those interior to Mercury, left over from the formation of the planets. A search for these vulcanoids at thermal infrared wavelengths, and a compilation of previous searches in the vicinity of the Sun, is detailed by Leake et al. (1987). Other searches for faint objects close to the Sun during solar eclipses are described by von Klüber (1932) and Dossin (1966), while Kresak (1983) addresses their orbital stability. True vulcanoids, with orbits such that they can avoid being swept up or perturbed by Mercury, can best be detected at thermal wavelengths when the objects are at quadrature. Short of thermal wavelengths, infraterrestrial asteroids are at their brightest at superior conjuction, when the solar elongation is prohibitively small, but fade quickly as the elongation increases. Table II shows the brightness decrease or increase (negative values) of an asteroid with a Bowell-Lumme phase function of 0.25 as a function of four heliocentric distances and the Earth-Sun-asteroid angle (Bowell et al. 1989).

Recently a new survey at solar elongations between 28 and 40° was completed at the Llano del Hato Observatory of the Centro de Investigaciones de Astronomia near Merida, Venezuela by Hoffmann and Geyer. A 1 m Schmidt telescope was used each morning from January to March 1989 to obtain one plate per day (at best) for a given elongation, tracking at the sidereal rate. Although the results from this survey were negative, it can be used to set some upper limits to vulcanoid populations. Assuming a uniform distribution of orbital inclinations up to 60°, and a uniform distribution of the other orbital elements, 1.4% of such objects should have passed through the search fields during the survey. Taking the motion of the sky into account, the nondetection is equivalent to an upper limit of 240 objects with diameters of 6 km or greater,

TABLE II

Change in Apparent Magnitude as a Function of Heliocentric Distance (AU) and Earth-Sun-Asteroid Angle *C*

E-S-A Angle *C*	AU				E-S-A Angle *C*	AU			
	0.10	0.35	0.60	0.85		0.10	0.35	0.60	0.85
6	38.6	34.0	23.8	7.8	102	-6.0	-0.2	2.5	4.4
18	13.6	13.8	9.0	2.5	114	-6.8	-0.5	2.4	4.4
30	6.4	7.6	5.1	2.7	126	-7.4	-0.9	2.2	4.3
42	2.1	4.4	3.7	3.2	138	-8.0	-1.3	1.9	4.1
54	−0.8	2.6	3.1	3.7	150	-8.6	-1.7	1.6	3.9
66	−2.7	1.5	2.9	4.0	162	-9.3	-2.2	1.1	3.5
78	−4.1	0.8	2.7	4.2	174	-10.1	-3.0	0.5	2.9
90	−5.2	0.3	2.6	4.4					

or 75 objects with diameters of 10 km or more. No conclusion can be made for bodies with diameters of <5 km, or for populations always closer to the Sun than 28°, where searches are best conducted in the thermal infrared. Other previous and future surveys, systematic, casual, and during eclipses, that yield negative results should be analyzed and used to help sharpen the upper limits set here. Similar surveys taken from latitudes of about 50° on the Earth around the summer solistice, or from polar latitudes around the winter solistice, would take advantage of the prolonged twilight conditions that prevail as the Sun moves parallel to, but below the horizon. This condition allows protracted observational accessiblilty of objects at small elongations from the Sun.

III. SEARCH INSTRUMENTS AND TECHNIQUES

Summaries of current asteroid searches and their instrumentation and techniques are given by Helin and Dunbar (1990), Shoemaker et al. (1990), Gehrels (1991) and Rabinowitz (1991). The lion's share of discoveries of NEAs are credited to E. and C. Shoemaker and E. Helin. The method that has contributed the most discoveries in the past has relied on the use of the 0.46 m Schmidt on Palomar Mountain and the use of hyper-sensitized photographic plates. Two plates of the same field of view are taken about 45 minutes apart and visually compared with a stereo microscope. A moving object appears at a different depth than that of the fixed field stars if it has moved between exposures. NEAs are identified by their streaked shape or peculiar rate of motion relative to more distant moving objects. Recently, a new method of finding NEAs has come into play. At the Spacewatch Telescope of the University of Arizona at Kitt Peak, a TK2048 CCD is read out in drift-scan mode by a real-time computer system. The computer has been programmed to recognize trailed images and to detect moving objects by their change in position between three consecutive scans. This system is finding NEAs at a greater rate than any other method (15 objects in the first 10 months of regular operation) and is also extending the limits of detection. For the first time, asteroids smaller than 100 m are being found near the Earth, and larger NEAs are being found as far away as 2 AU.

Although the combined efforts of observers using Schmidt telescopes throughout the world has increased the rate of NEA discoveries in recent years, it is clear that the photographic method employed on these telescopes is reaching the limit of its capabilities. It is doubtful that emulsions with a quantum efficiency greater than 15% will ever be developed (K. Russell, personal communication). The potential of CCDs, on the other hand, is only just beginning to be explored. Quantum efficiencies as high as 80% are promised in the near future, and the CCDs will have greater resolution. A telescope designed to move rapidly across the sky, with the CCD read out at a high rate to match the motion of the sky across the field of view, could achieve the same area coverage as a Schmidt telescope without compromising its limiting magnitude for fast-moving objects. In principle, a marriage between CCD and

Schmidt telescopes could achieve the highest detection rate. Unfortunately, the difficulty of covering the large, curved field of view of a Schmidt with the small, flat areas of CCD chips would have to be solved. Current CCDs require an extensive housing for electrical connections and cooling with liquid nitrogen. It is doubtful that an array of CCDs could be built to take advantage of the wide field of the Schmidt unless the cooling requirement were eliminated, and the electrical connections reduced.

One other instrument that was successful at finding NEAs in the past was the Infrared Astronomical Satellite (IRAS). In the brief time it was operational (~8 months), it discovered two NEAs and one Earth-crossing comet (Green et al. 1985). These discoveries were made despite the fact that the satellite had not been designed to find NEAs. The area of the sky roughly 90° from opposition was scanned at intervals of roughly 100 min, but the overlap between consecutive scans was only 50%. As a result, there was a strong bias against multiple observations of retrograde NEAs. Also, some of the detector elements on the satellite were not functional. It is thus possible that the best instrument for detecting NEAs would not require high-resolution imaging at visual wavelengths. Instead, an infrared detector with low spatial resolution might accomplish the job more efficiently. For each NEA discovered by such an instrument, follow-up observations could be made with a CCD for confirmation and astrometry. The lesson to be learned here, however, is not that CCDs, or photographic plates, or infrared detectors offer the best solution. In the long run, the best method will be determined by whatever technology is currently available that can image a large area of the sky to faint limiting magnitudes without being swamped by detections of stars, galaxies, and other stationary objects. Perhaps a "smart" chip could be developed that would only respond to objects in motion, essentially an artificial retina.

IV. CONCLUSIONS

We have summarized in this chapter current strategies for finding near-Earth asteroids and have also suggested new ones. These strategies are appropriate given the distribution of orbits for the known NEAs. It is clear, however, that there are strong observational biases in this distribution. The NEA population is a highly structured complex. It has diverse origins, while Jupiter and the terrestrial planets have affected the distribution dynamically. The delivery process from the main belt to the Earth is particularly complex. The task of debiasing the observed distribution is therefore a formidable one. A physically plausible distribution must first be constructed and then tested by modeling the observational selection effects. If the predicted distribution matches the observed one, then the assumed distribution may be the true one, but not necessarily. While this task remains a challenge for the future, at least significant progress has been made in the development of search instruments and techniques. With the advent of CCD scanning, the current discovery rate of NEAs by observers world-wide has doubled. Improvements in the use of

CCDs within the next few years will double the discovery rate again. If the objective of these efforts is to detect all the NEAs in the solar system larger than a given size, however, the search must be directed in such a way as to minimize the observational biases. At the same time, our understanding of the physical processes that have created the NEA population must be used to guide the observations. Otherwise, it can never be known with certainty whether the near-Earth asteroids that are observed represent the complete population.

Acknowledgments. J. S. Lewis and T. Triffet are acknowledged for their continuing support of Spacewatch, and through SERC, for the support of J. D. while composing this chapter. Reviewer A. Milani and another anonomyous (but thorough) referee helped to improve the chapter, and E. Helin provided critical input.

REFERENCES

Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., and Harris, A. 1989. Application of photometric models to asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 524–556.

Dossin, F. 1966. Comètes faibles au voisinage du soleil. *Mem. (8°) Soc. Roy. Sci. Liège* 5:12.

Drummond, J. D. 1991. Earth-approaching asteroid streams. *Icarus* 89:14–25.

Fernandez, J. A., and Ip, W.-H. 1991. Statistical and evolutionary aspects of cometary orbits. In *Comets in the Post-Halley Era,* vol. I, eds. R. L. Newburn, Jr., M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 487–535.

Gehrels, T. 1991. Scanning with charge-coupled devices. *Space Sci. Rev.* 58:347–375.

Gradie, J. C., Chapman, C. R., and Tedesco, E. F. 1989. Distribution of taxonomic classes and the compositional structure of the asteroid belt. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 316–335.

Green, S. F., Davies, J. K., Eaton, N., Stewart, B. C., and Meadows, A. J. 1985. The detection of fast-moving asteroids and comets by IRAS. *Icarus* 64:517–527.

Hahn, G. 1988. A Data Base of Observing Conditions for Aten-Apollo-Amor Objects During the 20th Century. Uppsala Astronomical Obs. Rept. No. 45.

Halliday, I., Griffin, A. A., and Blackwell, A. T. 1989. Detailed records of many unrecovered meteorites in western Canada for which further searches are recommended. *Roy. Astron. Soc. Canada* 83:49–80.

Halliday, I., Griffin, A. A., and Blackwell, A. T. 1990. Evidence for the existence of groups of meteorite-producing asteroidal fragments. *Meteoritics* 25:93–99.

Helin, E. F., and Dunbar, R. S. 1990. Search techniques for near-Earth asteroids. *Vistas in Astron.* 33:21–37.

Hughes, D. W. 1990. Dust coma onset at large heliocentric distances. In *Asteroids, Comets, and Meteors II*, eds. C.-I. Lagerkvist, H. Rickman, B. A. Lindblad and M. Lindgren (Uppsala: Uppsala Univ. Reprocentalen), pp. 327–342.

Kresak, L. 1983. Dynamical interrelations between the smaller bodies of the Solar System. *Highlights in Astron.* 6:377–390.

Lagerkvist, C.-I., and Magnusson, P. 1990. Analysis of lightcurves. II. Phase curves in a generalized HG-system. *Astron. Astrophys. Suppl.* 86:119–165.

Leake, M. A., Chapman, C. R., Weidenschilling, S. J., Davis, D. R., and Greenberg, R. 1987. The chronology of Mercury's geological and geophysical evolution: The vulcanoid hypothesis. *Icarus* 71:350–375.

Milani, A., Hahn, G., Carpino, M., and Nobili, A. M. 1989. Dynamics of planet-crossing asteroids: Classes of orbital behavior. Project SPACEGUARD. *Icarus* 78:212–269.

Rabinowitz, D. L. 1991. Detection of earth-approaching asteroids in near real time. *Astron. J.* 101:1518–1529.

Rabinowitz, D. L. 1992. The flux of small asteroids near the Earth. In *Asteroids, Comets, Meteors 1991*, eds. A. W. Harris and E. Bowell (Houston: Lunar and Planetary Inst.), pp. 481–485.

Rabinowitz, D. L. 1993. The size distribution of the Earth-approaching asteroids. *Astrophys. J.*, in press.

Shoemaker, E. M., Wolfe, R. F., and Shoemaker, C. S. 1990. Asteroid and comet flux in the neighborhood of Earth. In *Global Catastrophes in Earth History*, eds. V. L. Sharpton and P. D. Ward, Geological Soc. of America SP-247 (Boulder: Geological Soc. of America), pp. 155–170.

Tedesco, E. F. 1989. Asteroid magnitudes, UBV colors, and IRAS albedos and diameters. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 1090–1138.

Tedesco, E. F., Williams, J. G., Matson, D. L., Veeder, G. J., Gradie, G. J., and Lebofsky, L. A. 1989. A three-parameter asteroid taxonomy. *Astron. J.* 97:580–606.

von Klüber, H. 1932. Über Nachforschungen nach intramerkuriellen Körpern. *Astron. Nach.* 244:307–314.

Wetherill, G. W. 1991. End products of cometary evolution: Cometary origin of earth-crossing bodies of asteroidal parents. In *Comets in the Post-Halley Era*, vol. II, eds. R. L. Newburn, Jr., M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 537–556.

DYNAMICAL RELATIONSHIPS OF NEAR-EARTH ASTEROIDS TO MAIN-BELT ASTEROIDS

RICHARD GREENBERG and MICHAEL C. NOLAN
University of Arizona

Near-Earth asteroids have relatively short (10–100 Myr) lifetimes, and the population must be replenished rapidly compared to the age of the solar system. We discuss the sources and mechanisms of replenishment, as evidenced by observations of asteroids, meteorites and meteors, and by dynamical theories. We also discuss the uncertainties in observations and theories, and the extent to which they influence our understanding of the composition and distribution of near-Earth asteroids.

I. INTRODUCTION

Near-Earth asteroids are a potential source of useful resources, as established by a number of other chapters in this book. A central issue in the study of near-Earth asteroids (NEAs) is identification of their place of origin and how they got "here."

To readers concerned with the practical issues of obtaining and using materials from these bodies, they may still seem very far away. However, from the point of view of orbital dynamics, they are in fact already "here." Most NEAs are "Earth-crossing" (their orbits cross the Earth's) so the Earth moves through that population, continuously sampling it in a statistical way.

While in conventional usage, the term NEA refers to a body large enough to be astronomically observable, we prefer here a broader definition that does not impose such an arbitrary cut-off. The population of NEAs has a size distribution ranging from (currently) largest bodies 1036 Ganymed (about 40 km diameter) and 433 Eros (about 20 km diameter) to great numbers at much smaller sizes. The actual size distribution is poorly known, due to the small number of objects that have been observed (a total of 201 as of March, 1992), complicated by observational biases.

Larger NEAs are (fortunately for civilization and life in general) so sparsely distributed that, even with the Earth's continuous sampling of the region, they hit relatively infrequently. A given single body probably remains an Earth-crosser for about 30 Myr before either hitting the Earth, hitting Venus, or being ejected from the solar system (all with comparable probability). As the Earth samples the population, the specimens that people collect are called meteorites. Meteorites represent that portion of the NEA population (in the

size range from about 0.1 kg to 10^4 kg) that hits the Earth frequently enough for human collection and can remain at least partially intact for some time after hitting the ground. Larger NEAs are too rare to impact often enough (and are vaporized upon impact in the extreme case). Smaller NEAs generally break up in the atmosphere, though this is somewhat dependent on what they are made of.

Observational evidence regarding the provenance of NEAs comes from several sources. Meteorites, the NEAs already delivered to our collections and laboratories, provide information by their physical and chemical properties, the statistics of those properties, and in some cases the documented circumstances of their falls to the ground. Astronomical observations of NEAs provide orbital properties and, in some cases, remote sensing of the character of their surfaces. Only 35 of 201 known (as of March, 1992) NEAs have assigned classes, meaning that they have been observed at several (3–8) wavelengths in the visible and near-infrared spectral regions. A few have been observed in greater detail.

Finally, we have similar observations (orbits and surfaces) of asteroids in the main belt, which, along with comets, are the most likely sources of NEAs.

A genetic relationship between NEAs and main-belt asteroids can be considered by theoretical studies of the dynamical processes (collisional and gravitational) that can change orbits, and by statistical studies that relate the physical properties of main-belt asteroids to those of NEAs (including meteorites). In summary, the objective of this line of research is to relate the statistics of meteorite types to statistics of asteroid classes, by consideration of dynamical processes.

One motivation for such work is to use meteorites to learn more about asteroids, which in turn can be interpreted (with care) as representatives of the original building blocks of planets. This work also sheds light on the properties of NEAs for those who would advocate practical utilization of NEA resources. To the extent that we understand the evolution and interrelations between the main belt and meteorites, our knowledge of the NEAs benefits from the larger statistical data base that the main belt and laboratory studies of meteorites provide.

Greenberg and Nolan (1989) reviewed this subject to considerable depth in an earlier volume in this series. In this volume we briefly summarize that work, but concentrate in greater depth on two topics:

1. A tutorial description of the currently known orbital-evolution routes that deliver asteroids from the main belt to various parts of Earth-crossing orbital space.
2. The results of numerical work that illustrates the important role of stochastic events in skewing systematic relationships between main-belt and NEA physical properties.

II. PROBLEMS AND CONSTRAINTS

An overriding problem that severely constrains any effort to relate meteorites to a main-belt provenance is the lack of evidence in the main belt for material similar to ordinary chondrites, the most common kind of meteorite seen to fall on the Earth. For a short interval in the mid-1980s, this difficulty seemed to disappear. Spectral studies by Feierberg et al. (1982) suggested that the common S-type asteroids might be made of ordinary-chondritic material.

That identification encouraged the development of two complementary models of meteorite-asteroid connection. One (Greenberg and Chapman 1983) stressed the role of the collisional processes that liberate material from parent bodies; the other (Wetherill 1985) considered the effects of orbital-evolutionary paths on observable properties. The principal observation driving the collisional model was the shortage of olivine among meteorites, relative to olivine's expected volumetric abundance in differentiated parent bodies. This observation was addressed by invoking cratering impacts as the dominant launching process, rather than catastrophic fragmentation, which would give volumetric sampling. The principal constraint on the orbital-evolution model was the time-of-day statistics for meteorite falls on the Earth, which indicate substantial orbital maturity for bodies that impact this planet. This constraint was addressed by modeling meteorites as collisional products of larger NEAs that were able to survive and evolve to orbital maturity.

Although these two models are complementary in the sense of having concentrated on different processes and constraints, there was an unfortunate symmetrical shortcoming: each violated the constraints that drove the other. The collision model, in down-playing catastrophic fragmentation as a launching process from the main belt, produced bodies too small to remain intact until reaching orbital maturity. The orbital evolution model adopts Dohnanyi's (1969) production rate for collisional products, for which catastrophic fragmentation is dominant, and would yield too much olivine in meteorites.

A more serious difficulty effectively overshadowed these shortcomings: further analysis by Gaffey (e.g., 1986) showed that the S-asteroids appear to be differentiated material with a significant metallic component, contradicting the identification with ordinary chondrites, which both models of meteoritic provenance had invoked. A possible resolution to the problem could be that the spectral data do not accurately represent the material below the asteroids' surfaces, perhaps due to regolith precesses (Housen et al. 1979) or "weathering" in the environment of space (Britt 1991) or the way that an ensemble of materials on a surface appears when averaged together and viewed through a telescope (Wetherill 1985). It is also possible that ordinary chondrite parent bodies are not among the S-asteroids studied by Bell et al. (1988), although they went to some effort to sample a large variety of S-asteroids. However, there is little evidence to support those possibilities other than that they explain meteorite provenance.

Therefore, the value of the models discussed above lies not in their

consistency with all observational constraints, but rather in their identification and treatment of some important processes that certainly play a role in bringing asteroids from the main belt to where the Earth can sample them as meteorites.

III. ORBITAL EVOLUTION

A. Current Distribution

The current orbital distribution of NEAs and the evolutionary paths that they may have taken in coming from the main belt are most easily visualized and described in terms of a, e space, where a is semimajor axis and e is orbital eccentricity. Bear in mind that the distribution into a third dimension of orbital inclination i can also be of interest and importance. Figure 1 shows such a plot with the regions occupied by main-belt asteroids shaded. Most main-belt asteroids (>97%) have eccentricities smaller than 0.3. Their a-distribution is gapped and clustered due to the effects of orbital resonances, primarily with Jupiter. For example the Hilda asteroids, near 4 AU, have a 3:2 ratio of orbital periods with Jupiter; the Trojans, near 5.2 AU, are in 1:1 commensurability with Jupiter. The gap at 2.5 AU corresponds to orbits that would be in 3:1 commensurability, and low-inclination orbits in the gap at 2.05 AU would be subject to the so-called ν_6 ($\dot{\nu}_6$) secular resonance. See Froeschle and Greenberg (1989) and Scholl et al. (1989) for reviews on mean-motion (commensurability) and secular resonances, respectively.

Any object with perihelion distance $q = a(1 - e) < 1$ AU and aphelion distance $Q = a(1 + e) > 1$ AU is an Earth-crosser. (NEAs just outside this range may also be Earth-crossers, due to the Earth's small orbital eccentricity.) These boundaries are shown on Fig. 1, so that the area between them is the Earth-crossing zone. NEAs are generally classified according to their position in this space, as shown in Fig. 1: the Apollos and Atens are Earth-crossing (with $a > 1$ AU and $a < 1$ AU, respectively), and the Amors are near Earth-crossing with $q > 1$ AU. Many Atens and Apollos have a as large as main-belt asteroids, but their eccentricities are much greater. In Fig. 2, we plot the orbits of the astronomically observed Earth-crossing and near-Earth-crossing asteroids (D. J. Tholen, personal communication).

Information on NEA orbital distribution also comes from fireballs, which track small NEAs as they enter (and are heated to incandescence by) the Earth's atmosphere. Extrapolation back from the direction and velocity of a fireball has been used to infer pre-encounter heliocentric orbits. Those fireballs that Wetherill and ReVelle (1981) identified as due to stony bodies gave a set of orbits that are also shown in Fig. 2. Interestingly, the distribution of semimajor axes is similar to that for the astronomically determined NEA orbits, and there seems to be some clustering near the positions of the 3:1 commensurability with Jupiter at 2.5 AU and near the 4:1 commensurability and the ν_6 secular resonance near 2.05 AU, though there are too few points to be sure the clustering is statistically significant. For comparison, in Fig. 3, we show all fireball orbits that Wetherill and ReVelle considered in generating

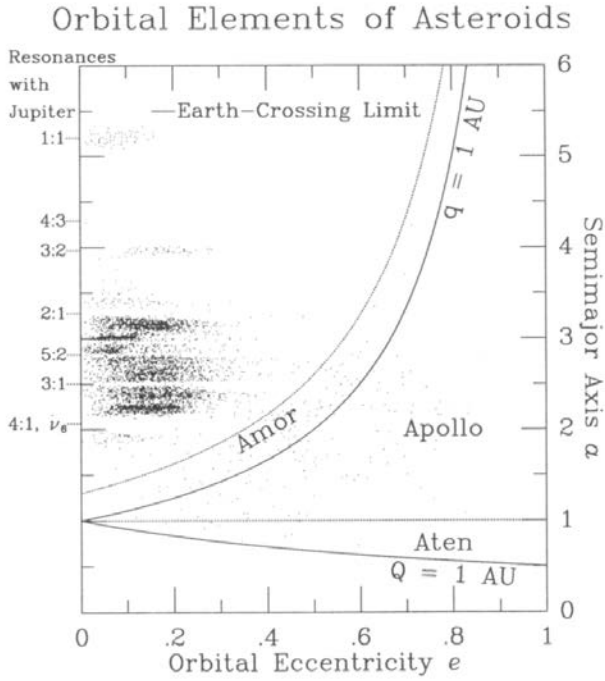


Figure 1. A plot of semimajor axis a vs eccentricity e for asteroids. The numbered asteroids (and all NEAs) are shown as dots. Main-belt and Trojan regions are visible as clumps. The Jovian mean-motion resonances are shown at left. Limits of Earth crossing are where perihelion q or aphelion Q distance equal 1 AU (more precisely, 1.017 and 0.983 AU when the Earth's e is taken into account). By definition, Amor class objects lie between the $q = 1.017$ AU and $q = 1.3$ AU lines; Apollos lie between the $a = 1$ AU and $q = 1.017$ AU lines; Atens lie between $Q = 0.983$ AU and $a = 1$ AU lines. Asteroid data from D. J. Tholen (personal communication).

the set of stony bodies included in Fig. 2; in Fig. 4 we show >45,000 orbits generated from the Lund meteor database (Lindblad 1987).

The fireball data are in several ways complementary to the telescopic data. Telescopic observations are biased towards light-colored material, whereas any material can make a meteor. Meteor cameras can detect much smaller objects than can telescopic detectors, but tend to miss the largest objects because they arrive too infrequently. The size ranges overlap, though only slightly. The smallest observed asteroid has a diameter between 6 and 12 m (Scotti et al. 1992), and the largest photographed meteor had an estimated diameter of 6 m before it hit the Earth's atmosphere (Ceplecha and McCrosky 1976).

The main difficulty in using meteors to relate main-belt asteroids and meteorites is in identifying the composition of the objects which form meteors. The methods currently used to determine the composition of meteors observed

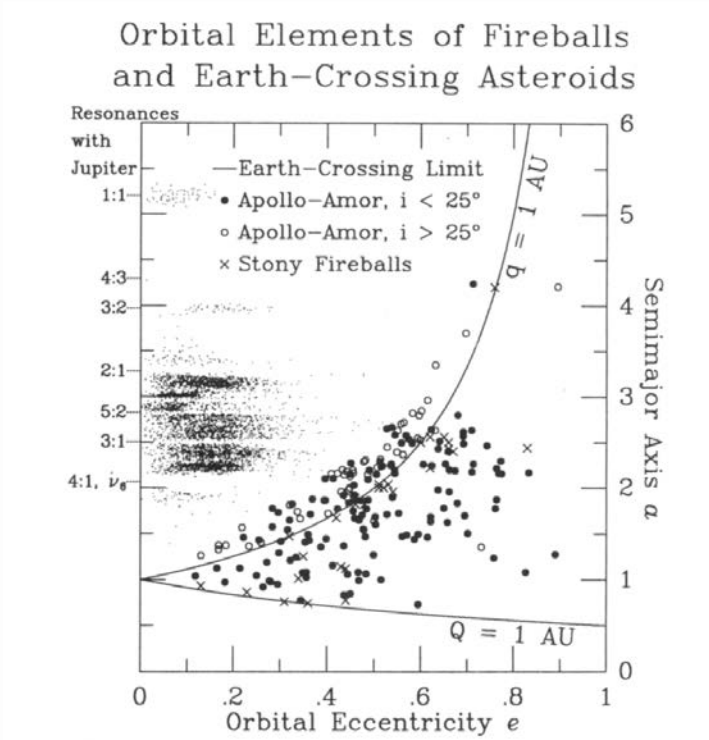


Figure 2. A plot as in Fig. 1, with NEAs highlighted, and with stony fireballs from Wetherill and ReVelle (1981).

photographically (McCrosky et al. 1978; Wetherill and ReVelle 1981) are very labor intensive, as the meteor velocity and brightness must be measured manually at many points along the trajectory. For this reason, only a small fraction of the available photographic data has been reduced. If a camera network based upon electronic (rather than photographic) detectors were built, an automated (or at least more automated) orbit and composition measurement system could allow the collection of enough data to be statistically meaningful (G. W. Wetherill, personal communication).

Radar can be used to find and measure the orbits of meteors, and semi-automated observatories exist which measure the orbits of thousands of meteors a day, but radar data provide no compositional information, and have strong observational biases; for example, radar can only observe very fast-moving meteors (Baggaley et al. 1992). This bias specifically excludes those regions of orbital space in which the NEAs that can easily be visited from the Earth (low delta-V) lie.

Accuracy may also affect the usefulness of available meteor measurements. The intrinsic accuracy of photographically measured meteors is quite good for orbit determination, as long as the time of occurrence of the meteor

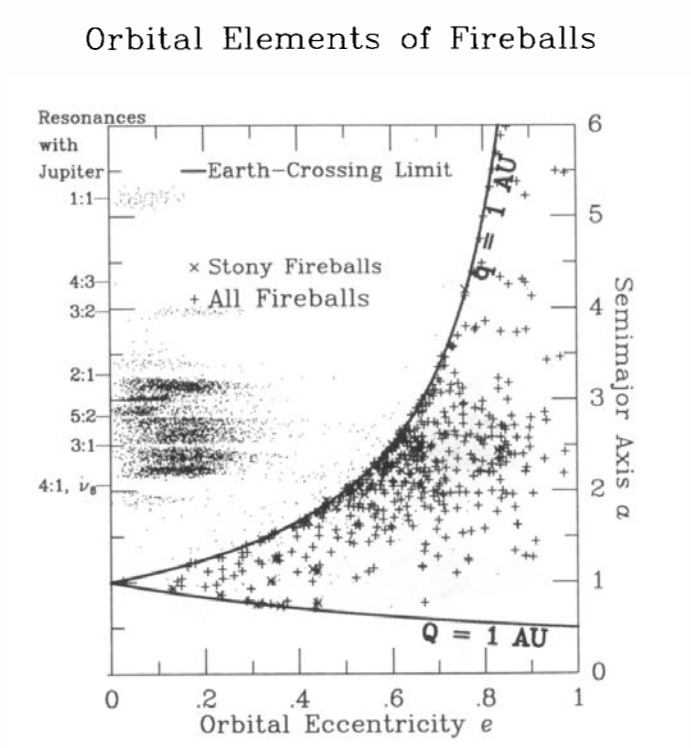


Figure 3. As in Fig. 1, with all of the fireballs considered by Wetherill and ReVelle (1981) in identifying the stony bodies shown in Fig. 2.

is well known (which is not true in all cases). The accuracy then depends on the geometry of the meteor trail with respect to the Earth and the observing stations in a complicated way. Muinonen and Bowell (1991) considered the related problem of accuracy of orbital elements based upon astrometric observations of asteroids, and have not found a general solution.

Meteor orbits are computed by extrapolating the trail of the meteor back into space. If at least two stations photograph the same meteor, the velocity and position in three dimensions can be computed by triangulation. Air resistance is assumed to be negligible before the appearance of the meteor fireball. The meteoroid is then assumed to have been on a hyperbolic geocentric orbit past the Earth, and that before it hit the Earth, its geocentric velocity was what it would have had on the asymptote of the hyperbola. Adding the Earth's rotational and heliocentric velocities then gives a heliocentric velocity for the meteoroid. From the heliocentric velocity thus computed, and assuming the position to be that of the center of the Earth, a heliocentric orbit is computed.

Nolan and Greenberg (1991) tested the implicit assumptions that three-body effects are negligible, and that the distance to the center of the Earth is unimportant, by direct numerical integration of the trajectories. They found that, in each of the 229 cases tested, a, e, and i were different by less than

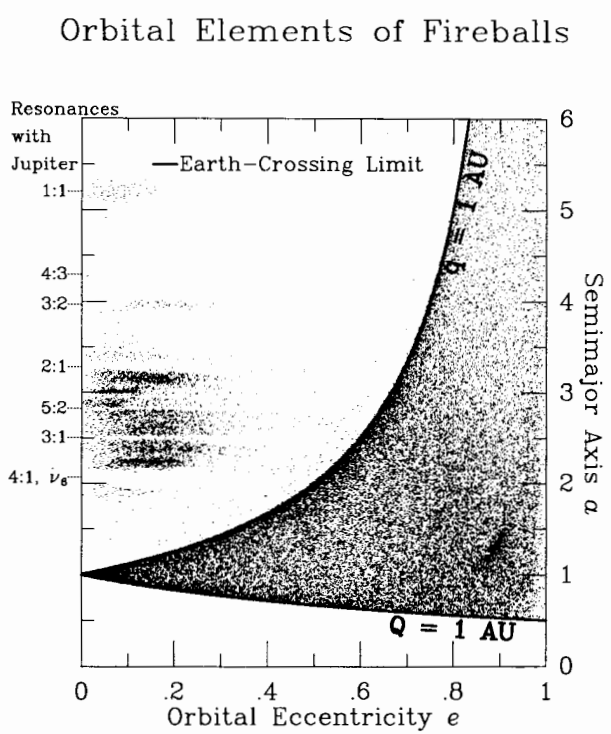


Figure 4. As in Fig. 1, showing >45,000 orbits from the Lund database. The complete database was not used, as some of the data were rounded in a way that generates visual artifacts when plotted.

1% from the value computed using the two-body approach described above.

Radar orbits have a lower intrinsic accuracy than do optically determined orbits, because the meteors can be observed for only a very short time. In addition, the directional accuracy is much lower (approximately 1° compared to approximately 1 arcsec for photographic film). The accuracy of properly reduced radar data is however comparable to the currently available optical measurements, due to the timing uncertainty and simplified data reduction used in the optical surveys to date (W. J. Baggaley, personal communication).

B. Evolutionary Paths

Although we do not yet understand all of the paths through orbital parameter space that an asteroid may take en route to Earth, there are a number of constraints and numerically modeled processes that give us a general idea of how most material gets here.

One important constraint comes from the Jacobi integral of the restricted three-body problem. Suppose the NEAs' heliocentric orbits were modified only by the gravitational effect of the Earth, and not by any other planets. Also, assume that the Earth's orbital eccentricity is negligible. In a rotating

coordinate system, in which the positions of the Sun and Earth are fixed, an NEA's pseudo-energy must be conserved; this pseudo-energy, or Jacobi constant, is the sum of its gravitational potential, its kinetic energy in the rotating frame, and a pseudo-potential that accounts for centrifugal force.

Tisserand (1882) showed that the Jacobi constant can be expressed in terms of the orbital elements a, e and i:

$$C = 1/a + 2\left[a\left(1 - e^2\right)\right]^{1/2} \cos i. \tag{1}$$

Thus the NEA would be constrained to a single surface in a, e, i space. Figure 5 shows the loci of those surfaces for $i = 0$. For larger i, each contour would shift slightly to lower values of e, but the shapes remain similar (Kresák 1967, Fig. 3).

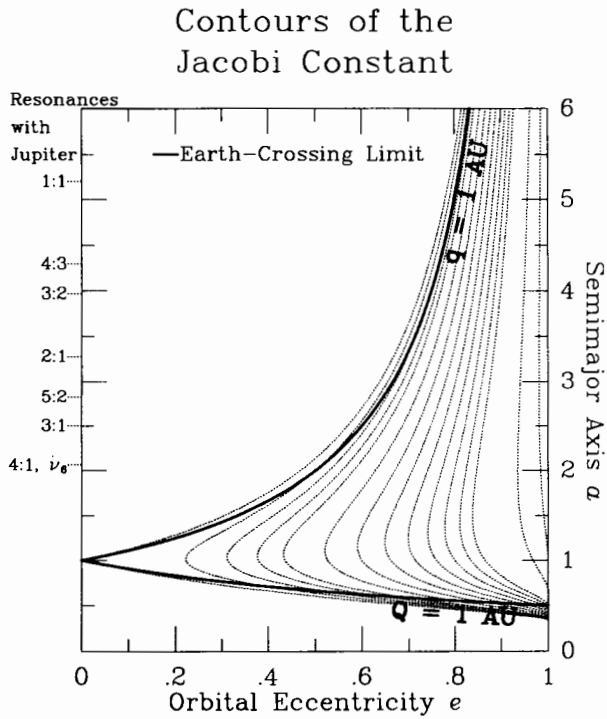


Figure 5. A plot of semimajor axis a vs eccentricity e for contours of the Jacobi constant C at zero inclination i. The contours are at $C =$ (from left to right) 3, 2.95, 2.9, 2.85, 2.8, 2.7, 2.6, 2.5, 2.4, 2.3, 2.2, 2.1, 2, 1.5 and 1.

Most of the time a NEA's orbit would change very little; upon close encounter with the Earth it might change significantly, but always remaining on a surface of constant C. (Tisserand derived Eq. [1] as a criterion for identifying comets before and after encounters with a planet.) If we ignore effects

of changes in i, encounters with other planets, and the orbital eccentricity of the Earth, the NEA would random walk along one of the contours of constant C shown in Fig. 5, with the maximum step size corresponding to the change in velocity for a surface-grazing encounter. In fact this is a reasonably good qualitative description of the evolution, at least between close encounters with other planets.

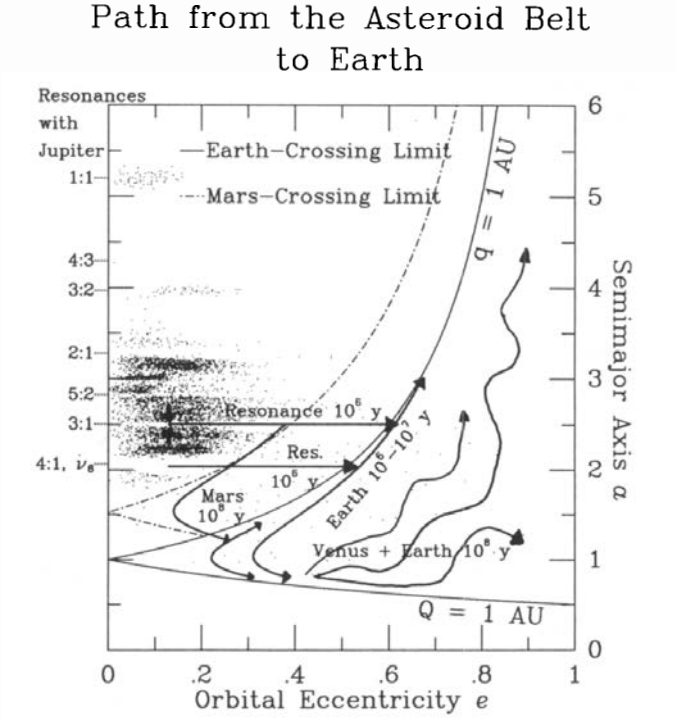


Figure 6. A figure showing the paths in a, e space that an asteroid may take en route to Earth-crossing, as discussed in Sec. III.B.

Based on a long series of numerical experiments, Wetherill (1988, and references therein) has discussed a set of evolutionary paths from the main belt to Earth-crossing and on to ultimate fates. These paths are illustrated in Fig. 6. In this scenario, most Earth-crossing asteroidal material originates in the main belt near the 3:1 resonance with Jupiter. Impacts among the asteroids yield debris on slightly different orbits from the parent bodies. Some of this material (including bodies tens of kilometers in size in Wetherill's model) is given an a value near that of the resonance. Whether there are enough asteroids near enough to the resonance that collisions can give large pieces a big enough kick to get into the resonance zone is unknown. A delta-V of $75\,\mathrm{m\,s^{-1}}$ is required to change an object's semimajor axis a by 0.02 AU at

the 3:1 resonance. This first orbital change (shown as a small change in a in Fig. 6) occurs instantaneously in this model.

Once in the 3:1 resonance, a body enters a regime of chaotic dynamics, in which orbital eccentricity can change drastically and unpredictably. From Fig. 6 we see that a body in the 3:1 resonance must have its eccentricity raised to 0.6 to become Earth-crossing. The time variation of e for a typical numerical integration (Wisdom 1983) is shown in Fig. 7. In this case, Earth-crossing is reached after about 1 Myr, and e remains that large for about 0.01 Myr. Wetherill (1985) assumes that these time scales are characteristic for asteroidal material injected from the main belt into the resonance. In fact there is some uncertainty about whether the cited numerical examples are truly typical, or whether any time scale can be adopted as characteristic or typical in the context of chaotic behavior (Wetherill 1987; Greenberg and Nolan 1989). Nevertheless, a widely accepted model is that for many bodies e can increase to Earth-crossing as shown by the horizontal arrow in Fig. 6.

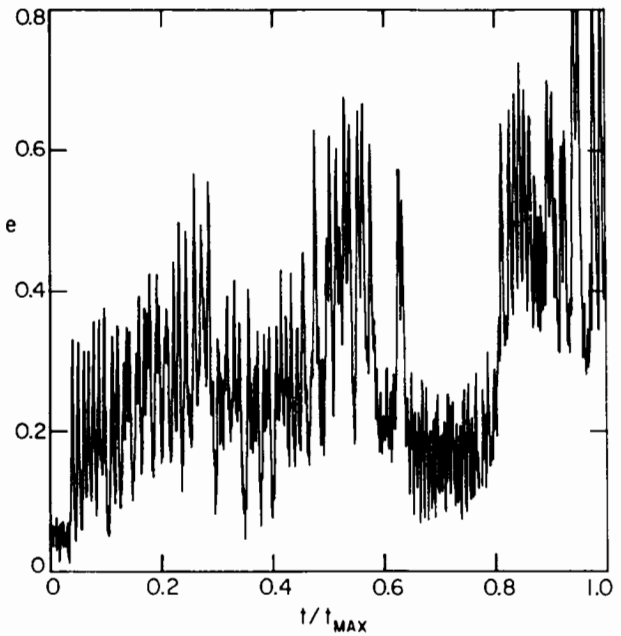


Figure 7. Evolution of eccentricity e for a chaotic trajectory in the 3:1 resonance (Wisdom 1983), with initial inclination of a few degrees. Earth crossing ($e = 0.6$) is reached in about 1 Myr ($t_{max} = 2.4$ Myr). Wetherill (1985) assumed that this value is typical and that, contrary to the behavior shown here, the asteroid is Mars crossing ($e > 0.6$) for most of that time.

Once a body reaches Earth-crossing, it begins to random-walk along one of the contours of Tisserand invariance, moving about 0.03 AU during the 10^4 yr before the chaotic behavior reduces e below 0.6. This change in a is enough

to remove the body from resonance. Subsequently, the now NEA evolves largely under the influence of Earth encounters, random-walking along a constant-C contour near the $q = 1$ AU line (i.e., remaining a shallow Earth crosser). The population of such bodies will spread out along the contour. An individual will random walk down to $a = 1$ AU in about 10 Myr (consistent with the usual time-squared dependence of random-walk distance).

As the evolution approaches $a = 1$ AU, the constant-C contour curves away from the $q = 1$ AU line down towards the $Q = 1$ AU line (cf. the contours in Fig. 5). As the value of a continues to decrease, e stops decreasing. The NEA becomes a more deeply Earth crossing body. At this stage in its evolution, the NEA begins to cross the orbit of Venus as well as that of the Earth. At this point, the premise that Earth encounters dominate the evolution becomes invalid. Encounters with Earth will still give changes with constant-C, but similarly frequent encounters with Venus will give changes along the nearly orthogonal set of C contours for that planet. Then an NEA is free to wander over the entire Earth- or Venus-crossing part of a, e space. After about 100 Myr, it might have evolved to almost anywhere in this region, and is likely to have impacted a terrestrial planet or become a Jupiter crosser, with comparable probability for each. Once a Jupiter crosser, the NEA is effectively lost, as it is ejected from the solar system by Jovian encounters in about 1 Myr.

The above scenario is a "fast track": it converts main-belt material to NEAs in about 1 Myr. There is also a "slow track" of comparable potential importance. Long before the 3:1 resonance pumps e up to Earth crossing, Mars crossing is reached at e approximately 0.3. Wetherill (1985) found from Monte Carlo studies that encounters with Mars can move a body out of resonance (change a by 0.03 AU) in 1 Myr, 100 times longer than the analogous effect of the Earth. According to Fig. 7, for a substantial portion of the 1 Myr before Earth crossing is achieved, the body is a Mars crosser. Thus Wetherill assumed that about half of the objects in the 3:1 resonance are removed from resonance by Mars before becoming Earth-crossers.

Actually, depending on the changing value of Mars' own eccentricity, values of e greater than 0.35 (or even more) may be required to reach Mars crossing. According to Fig. 7, values that great are achieved during less than 20% of the 1 Myr before Earth crossing is achieved, which suggests that the estimate that half the bodies are removed from resonance by Mars may be exceedingly high.

These objects that are removed then evolve primarily under the influence of repeated encounters with Mars, random-walking along constant-C lines analogous to those of the Earth. They reach the locus $q = 1$ AU and become Earth-crossers after about 100 Myr. They then evolve along the Earth's constant-C lines until they soon reach Venus crossing as well.

There is another "slow track" to the Earth via the ν_6 secular resonance at 2.05 AU, and perhaps other resonances as well. Collisional debris in the inner asteroid belt can have its eccentricity raised to Mars-crossing values

by action of the ν_6 secular resonance, but according to Wetherill (1987) such resonances cannot deliver significant amounts of material to Earth crossing, primarily because there are so few asteroids near enough to the actual resonance (Fig. 2). Material in the inner asteroid belt (near 2.2 AU) can have its eccentricity periodically raised to Mars crossing by weak interaction with the the ν_6 resonance. Repeated Mars encounters could then random walk bodies closer to the ν_6 resonance, allowing eccentricities to be pumped up further. Eventually some bodies can be brought close enough to the ν_6 resonance that they become deep Mars crossers, and then random walk to Earth crossing along the 100 Myr slow track.

More recent numerical experiments (Scholl and Froeschlé 1991*a*; Scholl and Froeschlé 1991*b*) indicate that significant amounts of material between 2 and 2.5 AU (affected by the $\dot{\nu}_6$, $\dot{\nu}_5$ and $\dot{\nu}_{16}$ secular resonances, by the 4:1 Jovian resonance at 2.065 AU, and by Mars encounters), may be pumped directly into Earth crossing by the effects of multiple resonances. The close proximity of these resonances allows interactions among them to greatly enhance their abilities to deliver material to the Earth. Scholl and Froeschlé (1991*a*) have begun to explore these possibilities, by showing that objects as far out as 2.1 AU are delivered to Earth-crossing orbits in 1 Myr or less, indicating that interactions between resonances have important effects. Thus once an object becomes a Mars crosser, it may take any of several resonance "fast tracks" to the Earth, besides the 100 Myr slow track.

Milani et al. (1989) have performed numerical integrations of 410 asteroid orbits for 2×10^5 yr. Their results show the processes discussed here to be important, and also that higher-order Jovian resonances and resonances with the Earth's orbit play a role in the orbital behavior of NEAs.

C. Orbital Maturity

Collisional disruption significantly affects the orbital distribution of asteroids undergoing the evolutionary processes described above. Depending on its size, a body leaving the asteroid belt may have a collisional lifetime so short that it cannot complete any of these orbital evolutionary sequences. For example, a stony body 1 m in diameter (a typical size for a meteorite before it hits the Earth's surface) would probably be destroyed by impact with another small body in about 12 Myr. This value is based on the cosmic-ray exposure ages of most stony meteorites (Wetherill 1988).

Bodies with collisional lifetimes of 20 Myr would have a distinctive distribution in a, e space. Few could survive to Earth crossing by the slow track via Mars encounters. Those that take the fast track would spread along constant-C lines near $q = 1$ AU, but would not survive long enough to spread over Earth-crossing space. Such a sub-population of NEAs can be described as orbitally immature.

These orbitally immature bodies are characterized by perihelia near 1 AU. When they encounter Earth, they are moving tangentially to Earth's orbit, and at higher heliocentric velocity (Fig. 8). They overtake the Earth from its

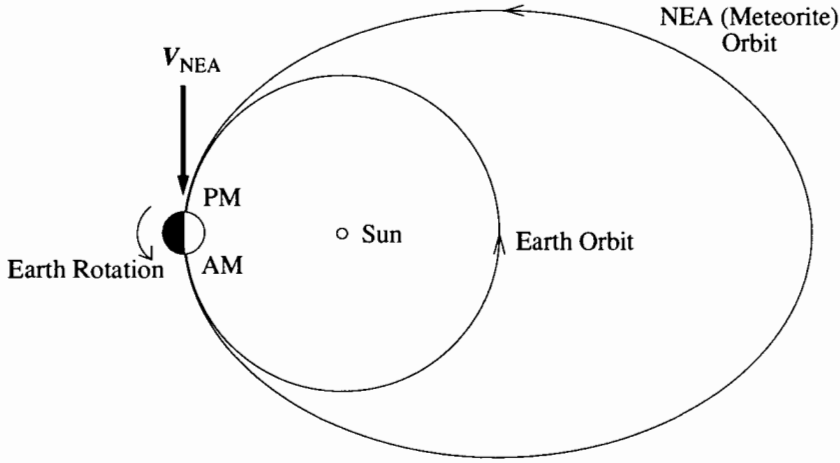


Figure 8. A figure showing the origin of AM/PM fall time asymmetry. An object with aphelion in the asteroid belt and perihelion near 1 AU will tend to encounter the Earth tangentially. As it has a higher heliocentric velocity than the Earth, it will tend to overtake it, and thus arrive on the trailing side, where the time is PM. If the asteroid's perihelion is significantly less than 1 AU, the radial component of the motion will be larger than the tangential component, and this bias towards PM arrivals will be eliminated. Asteroids with large inclinations can have large vertical (out of ecliptic) velocities, which will also dilute the PM preference.

trailing side, where local times are PM. According to this scenario, most stony meteorites should fall after noon and before midnight. In contrast, longer-lived bodies (such as larger NEAs or strong metallic ones) would reach orbital maturity; they would be scattered across Earth-crossing space, not only near $q = 1$ AU, and their fall times on Earth would be more evenly distributed over local time. In fact, this effect is complicated by the fact that asteroids falling on different parts of the Earth go through different atmospheric paths, resulting in different survival rates. For example, an asteroid hitting the Earth directly from behind can hit in the center, resulting in a vertical trajectory, or it can hit on the edge, and traverse the atmosphere tangentially, resulting in a nearly horizontal trajectory, and thus resulting in a very different atmospheric deceleration and heating profile.

In fact, small stony bodies that fall on Earth are not as strongly clustered on immature orbits as the above discussion would suggest. This result follows from the orbital distribution of stony fireballs shown in Fig. 2, where a substantial fraction of the orbits are closer to $Q = 1$ AU than to $q = 1$ AU. Of course, that line of evidence depends on the correct identification as stones for bodies that were never recovered. However, fall-time statistics for stony meteorites that were recovered also point toward considerable orbital maturity for these small bodies. Only $64\pm4\%$ of ordinary chondrites and $53\pm14\%$ of achondrites have fallen in the PM (Wood 1961; Wetherill 1968). These bodies

are evidently far more orbitally mature than one might expect based on the collision lifetimes of bodies of their size.

IV. EXPLANATION OF ORBITAL MATURITY

A. Ordinary Chondrites

This problem and its significance were identified by Wetherill, who also developed the following plausible explanation (Wetherill 1985). Most meter-scale (pre-atmosphere meteorite-size) bodies that are injected into resonance zones probably do not survive collisions long enough to reach orbital maturity. However, larger bodies leaving the main belt can evolve orbitally for much longer times. When such NEAs are broken up by collisions with very small objects, some of their pieces (or pieces of their pieces created by subsequent impacts) will be meteorite size and will reflect the orbital maturity of their parents.

Wetherill showed that if the bodies that started on this orbital evolution had a size distribution typical of a collisionally mature population, roughly $d(\text{number}) = \text{radius}^{-3.5}\, d(\text{radius})$ (Dohnanyi 1969), nearly all the meter-size bodies in a steady-state NEA population would be quite orbitally mature. This result seems intuitively reasonable, because the -3.5 power law has most of the mass of the population in larger bodies, so after considerable collisional evolution en route to Earth, most of the surviving meter-scale objects must be later-generation pieces that reflect orbital maturity.

In that case the PM fall times should be about 50% of all falls. Now the problem becomes the high value 64% (i.e., the substantial orbital immaturity) for ordinary chondrites, the most common stones. Wetherill (1985) explained this value by noting that, if the main belt is in collisional equilibrium (power law exponent -3.5), the collision debris that it produces has a power law exponent of -4: there is a significant portion of the total mass in very small bodies. With this power law for the material originally placed into resonance, the steady-state NEA population would be consistent with the PM fall times for ordinary chondrites according to Wetherill's model.

B. Achondrites

The much less common achondrites, stony meteorites from differentiated parents, have PM fall fractions of $53\pm14\%$. Contrasting this value with the ordinary chondrites ($64\pm4\%$), Wetherill (1987) inferred that these meteorites come primarily from near the ν_6 resonance.

The basis for this inference was the belief that bodies in this resonance cannot have their orbits pumped up to Earth crossing, so that they can only reach Earth via the Mars-crossing slow track. Such objects take 100 Myr to reach Earth crossing, by which time their power-law index will have flattened to -3.5, even if it started at -4 when the population left the main belt. Nearly all meteorite-sized Earth crossers from this population will be offspring of

much larger bodies and thus will reflect the large bodies' orbital maturities, with a PM-fall portion near 50%, according to that model.

Although key elements of this scenario were very innovative and probably play important roles in governing the character of NEA populations, the scenario as a whole is problematical for several reasons:

First, as discussed above, there is now evidence that material from the nu-six region reaches Earth crossing by a direct fast track. If this route is significant, then following the above logic it would presumably raise the PM-fall fraction for meteorites originating in the nu-six region.

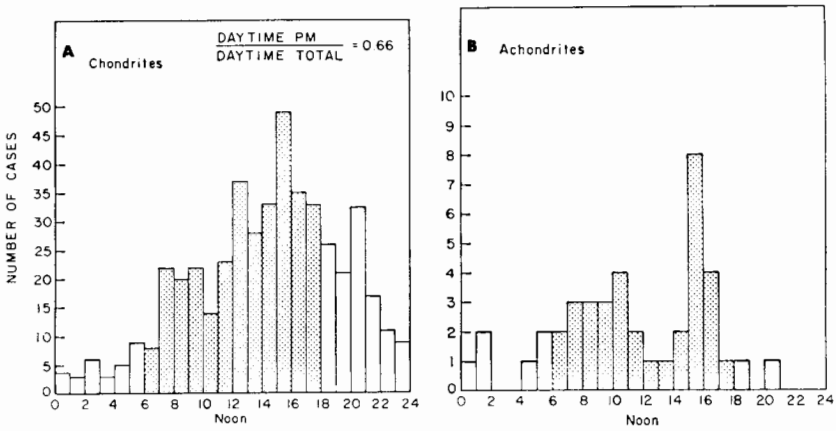


Figure 9. Observed local times of fall of stone meteorites (Wetherill 1968). The numbers of achondrites are too small to show any significant difference in AM/PM ratio compared with the chondrites, so that datum is not a convincing way to discriminate among source regions. More intriguing, albeit of marginal statistical significance, is the achondrite lunch break: without this prominent minimum, the achondrite distribution would look much like the chondrites.

Second, the difference between the achondritic and chondritic fall-time distributions is not well represented by the single parameter of PM-fall fraction. The fall-time distributions cited by Wetherill (1968) are shown in Fig. 9. The AM/PM fall fractions refer only to "daytime" falls (after 6 AM and before 6 PM), as observational biases are expected to dominate the "nighttime" statistics. The distributions are clearly quite different in shape. One interpretation might be that the shapes would be similar were it not for the "lunch-break" gap in the achondrites. The peak remains clearly between 3 PM and 4 PM in both cases. Until the gross structure is better understood, explanations of small difference in AM/PM ratio seem inappropriate. Third, this model does not address the lack of olivine among stony meteorites from differentiated parents. Nearly all achondrites are of crustal-type materials, yet volumetrically their parents must have been mostly olivine, the material of their mantles. Greenberg and Chapman (1983) addressed this problem by

assuming that most meteoritic material was liberated from main-belt parents by relatively surficial impacts, but that idea is at odds with the above scenario, which requires large objects to be ejected from the main belt and which (at least for the chondrites) implicitly requires catastrophic fragmentation. Note that, as discussed in Sec. II, the model by Greenberg and Chapman did not address many of the dynamical constraints, but it does highlight the issue of the missing olivine.

V. STEADY-STATE MODEL

Finally, we note concerns about the steady-state assumption for the modeling of populations migrating to Earth-crossing orbits, i.e., portions of the population en route to Earth that are identified by some set of orbital parameters (Earth crossers, for example) are essentially unchanging: sizes and types of bodies that are removed by collisions with small bodies or planets or by orbital change are continually replenished by similar bodies.

In fact, much of the replenishment is by collisional disruption of the larger bodies, which is clearly a discrete, infrequent process. The population must undergo occasional sudden changes in numbers of bodies in the middle and small size range, and corresponding changes in the character as large bodies of distinctive composition break up and temporarily dominate the population.

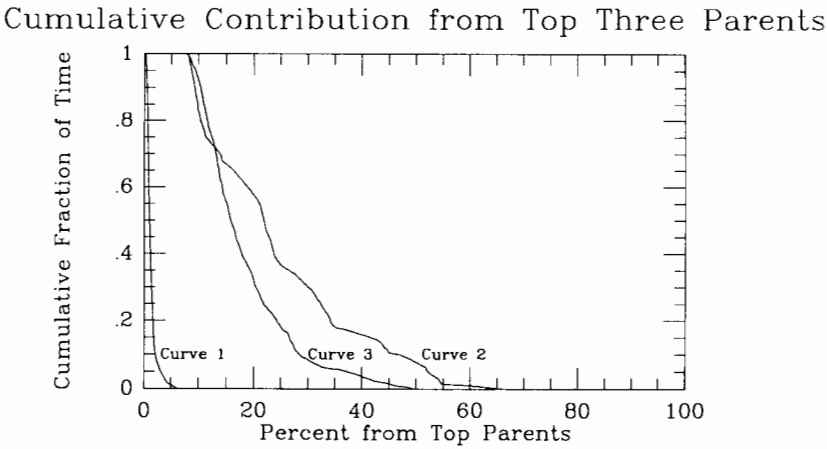


Figure 10. Plot of the fraction of the time over the last 2 Gyr (abscissa) that a given fraction of meteorite-sized Earth-crossing asteroids (ordinate) come from only three parent bodies. See text for discussion.

Nolan and Greenberg (1989) modeled this effect, and found that much of the time a large fraction of Earth crossers comes from only a few individual progenitors. Figure 10, for example, shows the fraction of time that a certain percentage of the Earth crossers is likely to come from only three progenitors.

For bodies that take the fast track to Earth (curve 1), most of the time only a few percent of the Earth crossers come from the three ancestors that contributed the most. These bodies experienced very little collisional evolution during their rapid orbital evolution. For the population that comes via the slow-track (curve 2), half the time 20% of the bodies come from only three ancestors, and 10% of the time about 50% of the bodies do. The slow-track bodies can be dominated by a few parents because there is plenty of time for collisional evolution. Curve 3 shows a population split evenly between fast- and slow-track bodies.

These results demonstrate that theories that rely heavily on a steady state should be evaluated with care. One must use caution in the systematic interpretation of NEA statistics. These results also help explain relatively short-term changes in the NEAs sampled as meteorites. For example, there are systematic differences between Antarctic meteorites, which fell over a period of about 3×10^5 yr, and modern falls elsewhere on Earth, which have been collected over about 200 yr. The variations are not great enough, however, to explain the missing olivine as a quirk of our collection interval.

Meteorite streams would further complicate this issue. Drummond (1991) suggests that the orbits of some NEAs are more correlated than is expected for a random distribution, and that therefore these "streams" of asteroids are likely the products of a recent collision, which would also generate smaller fragments which could become meteorites. This phenomenon would concentrate the fragments in a small region of space, so that in that region, products of the body that formed the stream could dominate the random background population. If the Earth happened to pass through this region, we would see an excess of meteorites from the stream, and therefore from a single parent body, rather than a uniform sample of the meteorite source region. This same phenomenon produces meteor showers. Due to the small sample of NEAs and the possible statistical biases in their observations, this subject is currently under debate, and the implications for delivery of material from the main belt are uncertain.

It is very unlikely that a stream could survive the dynamical processes that bring material from the main belt to the near-Earth region, as small differences in orbital parameters greatly affect the rate of evolution towards Earth crossing. Thus any streams which do exist must have formed since the parent bodies "arrived" in the Earth-crossing region. For that reason the meteorite stream phenomenon should not significantly affect the population of NEAs, rather the process modeled by Nolan and Greenberg (1991) pertains. However, meteorite streams may well affect our samples of NEAs, the meteorites, from which the compositions of the NEAs are inferred.

VI. CONCLUSIONS

We have shown the various plausible paths of orbital evolution for NEAs from the main belt. The statistical character of this population, in terms of both

orbital and physical properties, is largely controlled by the orbital evolution paths taken. However, we have also shown that many problems remain in interpreting NEA provenance from their properties. This is a challenging area for continued study, and the processes and logical connections revealed by past models will serve as tools for the continuing effort.

REFERENCES

Baggaley, W. J., Steele, D. I., and Taylor, A. D. 1992. A southern hemisphere radar meteor orbit survey. In *Asteroids, Comets, Meteors 1991*, eds. A. W. Harris and E. Bowell (Houston: Lunar and Planetary Inst.), pp. 37–40.

Bell, J. F., Owensby, P. D., Hawke, B. R., and Gaffey, M. J. 1988. The 52-color asteroid survey: Final results and interpretations. *Lunar Planet. Sci.* XIX:57–58 (abstract).

Britt, D. T. 1991. "Space weathering": Are regolith processes an important factor in the S-type controversy? *Meteoritics* 26:324 (abstract).

Ceplecha, Z., and McCrosky, R. E. 1976. Fireball end heights: A diagnostic for the structure of meteoritic material. *J. Geophys. Res.* 81:6257–6275.

Dohnanyi, J. W. 1969. Collisional model of asteroids and their debris. *J. Geophys. Res.* 74:2531–2554.

Drummond, J. D. 1991. Earth-approaching asteroid streams. *Icarus* 89:14–25.

Feierberg, M. A., Larson, H. P., and Chapman, C. R. 1982. Spectroscopic evidence for undifferentiated S-type asteroids. *Astrophys. J.* 257:361–372.

Froeschlé, Cl., and Greenberg, R. 1989. Mean motion resonances. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 827–844.

Gaffey, M. J. 1986. The spectral and physical properties of metal in meteorite assemblages: Implications for asteroid surface materials. *Icarus* 66:468–486.

Greenberg, R., and Chapman, C. R. 1983. Asteroids and meteorites: Parent bodies and delivered samples. *Icarus* 55:455–481.

Greenberg, R., and Nolan, M. C. 1989. Delivery of asteroids and meteorites to the inner solar system. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 778–804.

Housen, K. R., Wilkening, L. L., Chapman, C. R., and Greenberg, R. 1979. Asteroidal regoliths. *Icarus* 39:317–351.

Kresák, Ľ. 1967. Relation of meteor orbits to the orbits of comets and asteroids. In *Meteor Orbits and Dust*, ed. G. S. Hawkins, NASA SP-135, pp. 9–34.

Lindblad, B. A. 1987. The IAU meteor data center in Lund. In *Interplanetary Matter*, eds. Z. Ceplecha and P. Pecina (Prague: Czechoslovak Academy of Sciences), pp. 201–204.

McCrosky, R. E., Shao, C.-Y., and Posen, A. 1978. Prairie network fireballs. I. Summary and orbits. *Meteoritika* 38:44–59. In Russian; also in English as CFA preprint 655.

Milani, A., Carpino, M., Hahn, G., and Nobili, A. M. 1989. Dynamics of planet-crossing asteroids: Classes of orbital behavior. *Icarus* 78:212–269.

Muinonen, K., and Bowell, E. L. G. 1992. Asteroid orbital error analysis: Theory and application. In *Asteroids, Comets, Meteors 1991*, eds. A. W. Harris and E.

Bowell (Houston: Lunar and Planetary Inst.), pp. 429–432.
Nolan, M. C., and Greenberg, R. 1989. Stochastic evolution of asteroids to produce the ordinary chondrites. *Bull. Amer. Astron. Soc.* 21:964–965 (abstract).
Nolan, M. C., and Greenberg, R. 1991. Delivery of meteorites from the asteroid belt. *Meteoritics* 26:380–381 (abstract).
Scholl, H., and Froeschlé, Ch. 1991*a*. The ν_6 secular resonance region near 2 AU: A possible source of meteorites. *Astron. Astrophys.* 245:316–321.
Scholl, H., and Froeschlé, Ch. 1991*b*. The effect of secular resonances in the asteroid region between 2.1 and 2.4 AU. In *Asteroids, Comets, Meteors 1991*, eds. A. W. Harris and E. Bowell (Houston: Lunar and Planetary Inst.), pp. 205–209.
Scholl, H., Froeschlé, Ch., Kinoshita, H., Yoshikawa, M., and Williams, J. G. 1989. Secular resonances. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 845–861.
Scotti, J. V., Gehrels, T., and Rabinowitz, D. L. 1992. Automated detection of asteroids in real-time with the Spacewatch Telescope. In *Asteroids, Comets, Meteors 1991*, eds. A. W. Harris and E. Bowell (Houston: Lunar and Planetary Inst.), pp. 541–544.
Tisserand, M. F. 1882. *Ann. Obs. Paris* 16:E1.
Wetherill, G. W. 1968. Stone meteorites: Time of fall and origin. *Science* 159:79–82.
Wetherill, G. W. 1985. Asteroidal source of ordinary chondrites. *Meteoritics* 20:1–22.
Wetherill, G. W. 1987. Dynamic relationships between asteroids, meteorites, and Apollo-Amor objects. *Phil. Trans. Royal Soc. London* A323:323–337.
Wetherill, G. W. 1988. Where do the Apollo objects come from? *Icarus* 76:1–18.
Wetherill, G. W., and ReVelle, D. O. 1981. Which fireballs are meteorites? A study of the Prairie Network photographic meteor data. *Icarus* 48:308–328.
Wisdom, J. 1983. Chaotic behavior and the origin of the 3/1 Kirkwood gap. *Icarus* 56:51–74.
Wood, J. A. 1961. Stony meteorite orbits. *Mon. Not. Roy. Astron. Soc.* 122:79–88.

REVIEW OF ASTEROID COMPOSITIONS

MARCIA L. NELSON, DANIEL T. BRITT and
LARRY A. LEBOFSKY
University of Arizona

Asteroids are a potentially rich source of a variety of materials in the solar system. Some of them are also relatively easy to access because they have been perturbed into near-Earth orbits. Most of our information on asteroid surface compositions comes from telescopic reflectance spectroscopy, and most of our information on asteroid bulk compositions has been determined from the study of meteorites. In this chapter, we summarize the current state of knowledge of asteroid compositions, using the information from both these sources. We conclude with a discussion of some of the remaining controversies.

I. INTRODUCTION

Asteroids were the last class of objects to be discovered in the solar system. However, most astronomers did not consider asteroids as subjects worthy of study in their own right, therefore very little was learned about them until the latter half of this century. Now much has been learned about the brighter, main belt asteroids, and improvements in instrumentation have made detailed investigations of the faint asteroids feasible. Several dedicated asteroid search programs (see, e.g. Shoemaker and Shoemaker 1988; Helin and Dunbar 1991; Gehrels 1991) have significantly increased the number of known asteroids in recent years. The growing interest among astronomers in asteroids essentially coincided with a growing realization among meteoriticists that the vast majority of the meteorites in our collections have been derived from the asteroids. Astronomers and meteoriticists have been debating which classes of meteorites are derived from which classes of asteroids ever since.

It is sometimes difficult to determine which asteroids and meteorites are similar for several reasons. Telescopic reflectance spectroscopy gives compositional information only about the surfaces of the asteroids, and those surfaces are subject to alteration by a variety of irradiation and collisional processes. These processes can cause changes in the surface which obscure the diagnostic absorption features of component minerals in the reflectance spectra. There are a few meteorites in the collections which have been identified as surface material (the gas-rich chondrites), but the majority of the

meteorites are derived from asteroid interiors, which have not been subject to the same alteration processes. These differences can make it extremely difficult to determine compositional similarities for some of the asteroid and meteorite classes. It is even more complicated to attempt to identify the parent body of a specific meteorite, because there is no guarantee that the asteroid was not destroyed in the collision which created the meteoroid. It will probably not be possible for most meteorites, without detailed chemical comparisons from asteroid sample return missions.

There are several reasons for continuing to try to make these correlations. Interpretations of telescopic reflectance spectra, without meteorite analogs, are limited to discussions of the spectrally distinguishable minerals such as olivine, pyroxene, plagioclase and water. The presence of metal can sometimes be inferred from spectral slope, but it cannot be confirmed due to the lack of diagnostic absorption features. Meteorite analogs make it possible to apply the extensive mineralogic information derived from meteorites to the compositional interpretation of spectrally similar asteroids. These correlations are also essential for models of the origin of the solar system. They make it possible to place the information derived from the meteorites on conditions in the solar nebula in the correct spatial context.

Asteroids are also potentially rich sources of ores for space mining. The near-Earth asteroids are particularly attractive because they are in favorable orbits. In the chapters by Nichols and by Lewis and Hutson, the types of materials which could be mined from asteroids with compositions similar to the various meteorite types are discussed. These chapters can be combined with the discussion in this chapter to identify potential targets for future exploration.

Two recent review articles have discussed asteroids from a spectroscopic perspective (M. J. Gaffey et al. 1989,1992). When combined with the McFadden et al. (1989) review of the near-Earth asteroids, these papers give a complete review of the published asteroid spectral literature up to the time of their submittal. We refer the reader to these papers for detailed summaries of the compositional analyses of individual asteroids. In our section on asteroid spectroscopy, we have focused on the compositional similarities and differences represented within the different asteroid classes. We then discuss meteorite compositions, and identify analogs for the asteroid classes when applicable. We conclude with a discussion of the remaining problems and controversies in the field.

II. DISTRIBUTION AND CLASSIFICATION

The majority of the asteroids are believed to be material which never accreted to form planets (Wetherill 1989). Most of them are probably collisional fragments of the original planetesimals, but at least one asteroid, 4 Vesta, is believed to be intact because it still possesses a basaltic crust (Davis et al. 1989). Figure 1 shows the distribution of known asteroids as a function of

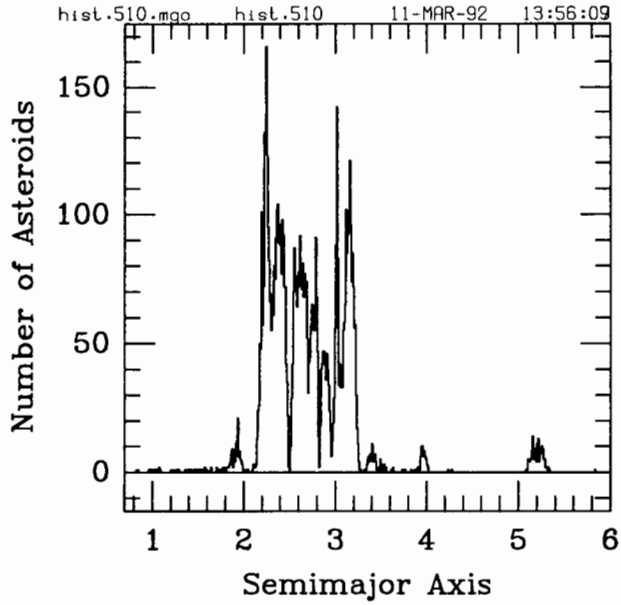
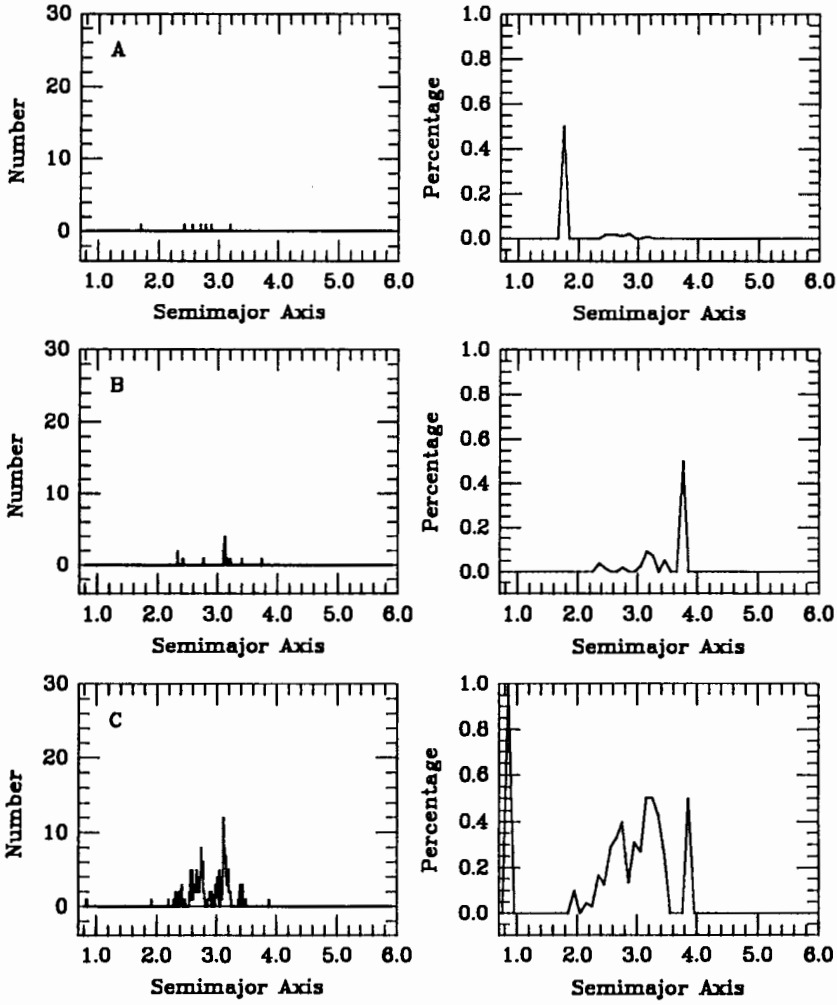


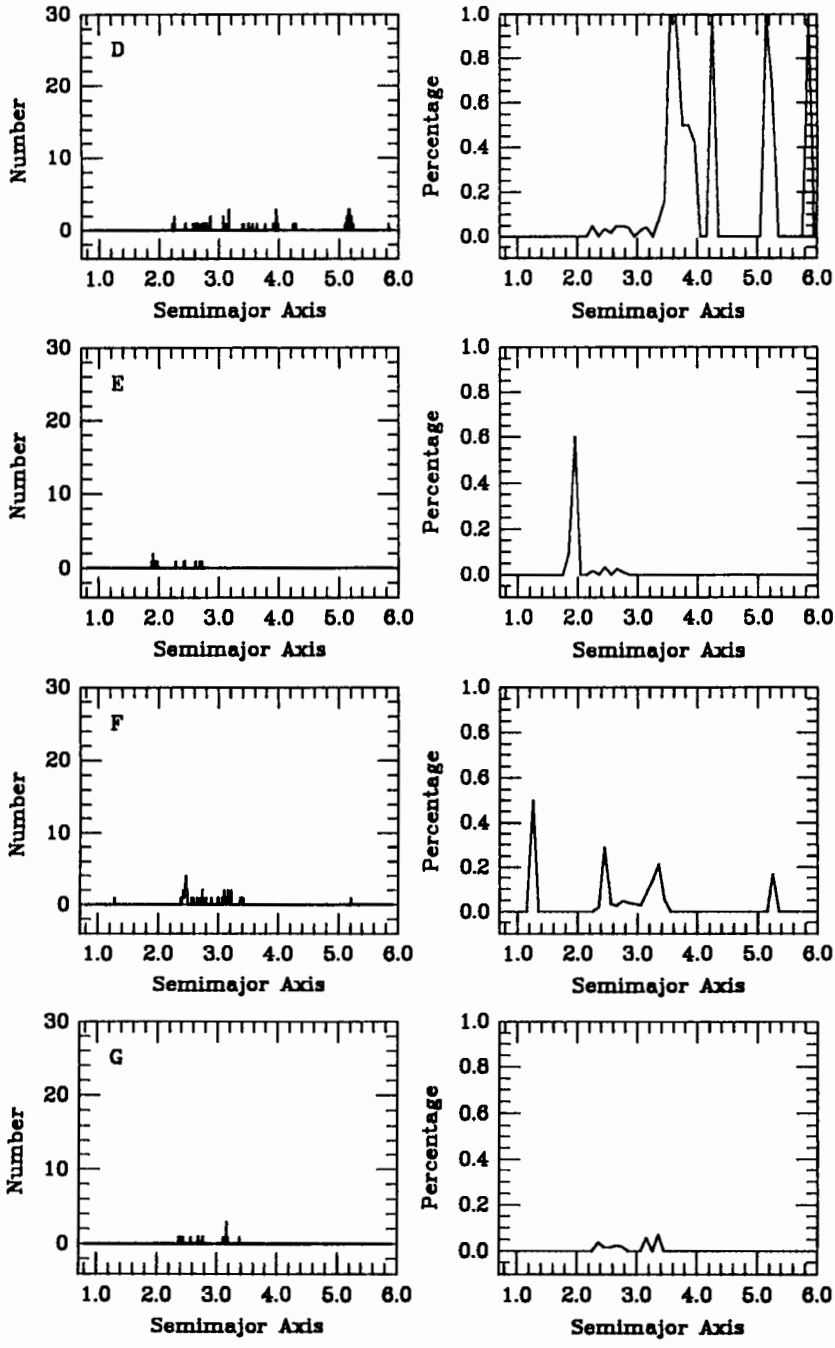
Figure 1. Distribution of asteroids as a function of orbital semi-major axes. This figure includes all the asteroids in the Minor Planets Center database which have been observed at more than one apparition, asteroid type, at a spatial resolution of 0.01 AU. The right panels show as of February 1991.

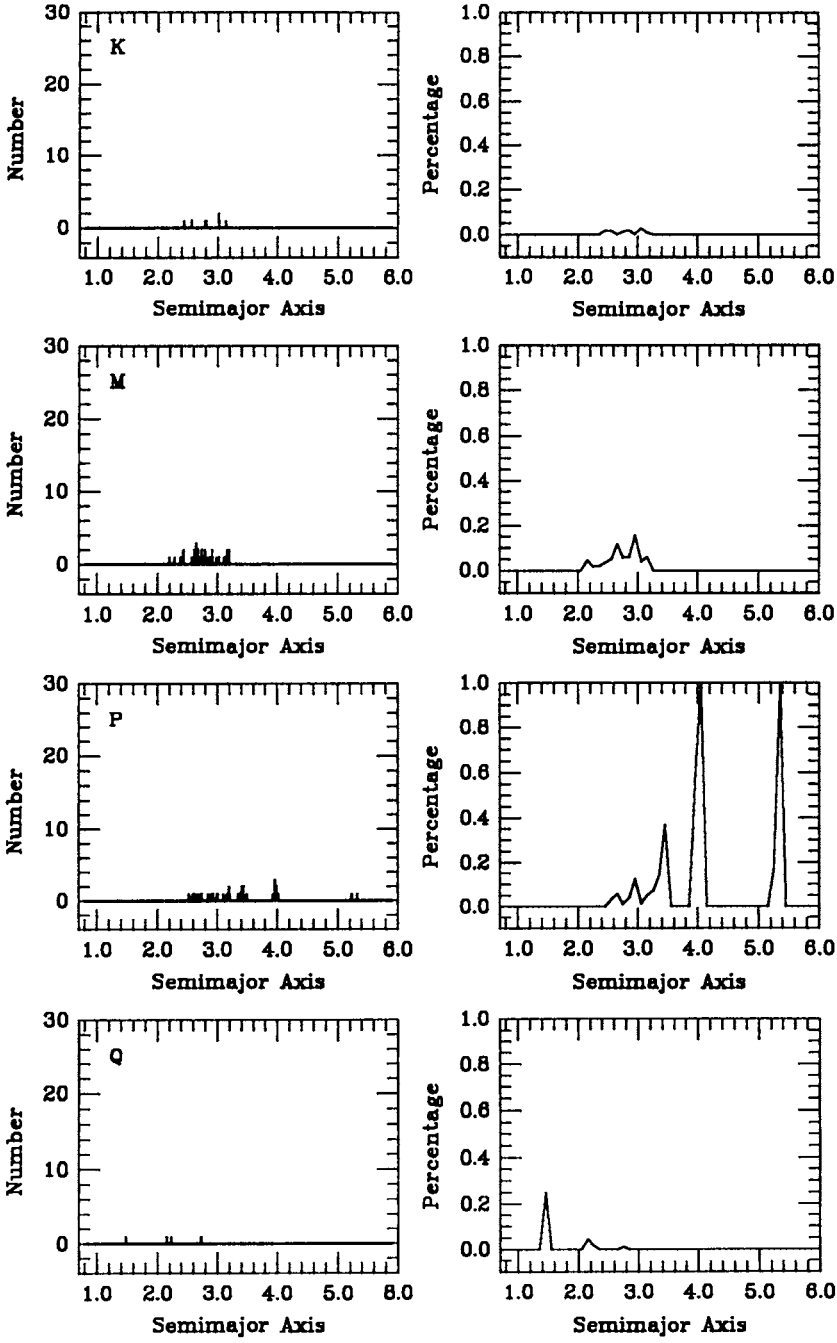
semimajor axis. Most of the asteroids are found in the main asteroid belt between the orbits of Mars and Jupiter (with semimajor axes between 2.1 and 3.3 AU). Orbital resonances with Jupiter create the gaps in the distribution (Froeschlé and Greenberg 1989). The dominant composition of main belt asteroids changes with heliocentric distance (Gradie et al. 1989), which suggests that they have remained near their original formation regions since the origin of the solar system (Ruzmaikina et al. 1989). The Trojan asteroids are located at the L4 and L5 Legrangian points of Jupiter's orbit (with semimajor axes of 5.2 AU). They are probably a small subset of the original outer solar system planetesimal population which were trapped in these resonances (Shoemaker et al. 1989). The near-Earth asteroids have semi-major axes ranging from >1 AU to 1.5 AU. Their lifetimes against collision with one of the terrestrial planets, or ejection from the solar system, are calculated to be approximately 10^7 to 10^8 yr (Chapter by Greenberg and Nolan). The majority of the present near-Earth asteroids are believed to have been perturbed from the main belt into planet-crossing orbits (Greenberg and Nolan 1989), although these dynamical mechanisms are still poorly understood. Some of the current population of near-Earth asteroids may also be extinct comet nuclei, (Weissman et al. 1989), however no definitive identification of an extinct nucleus has been made. Recent work by Binzel et al. (1992) has shown that

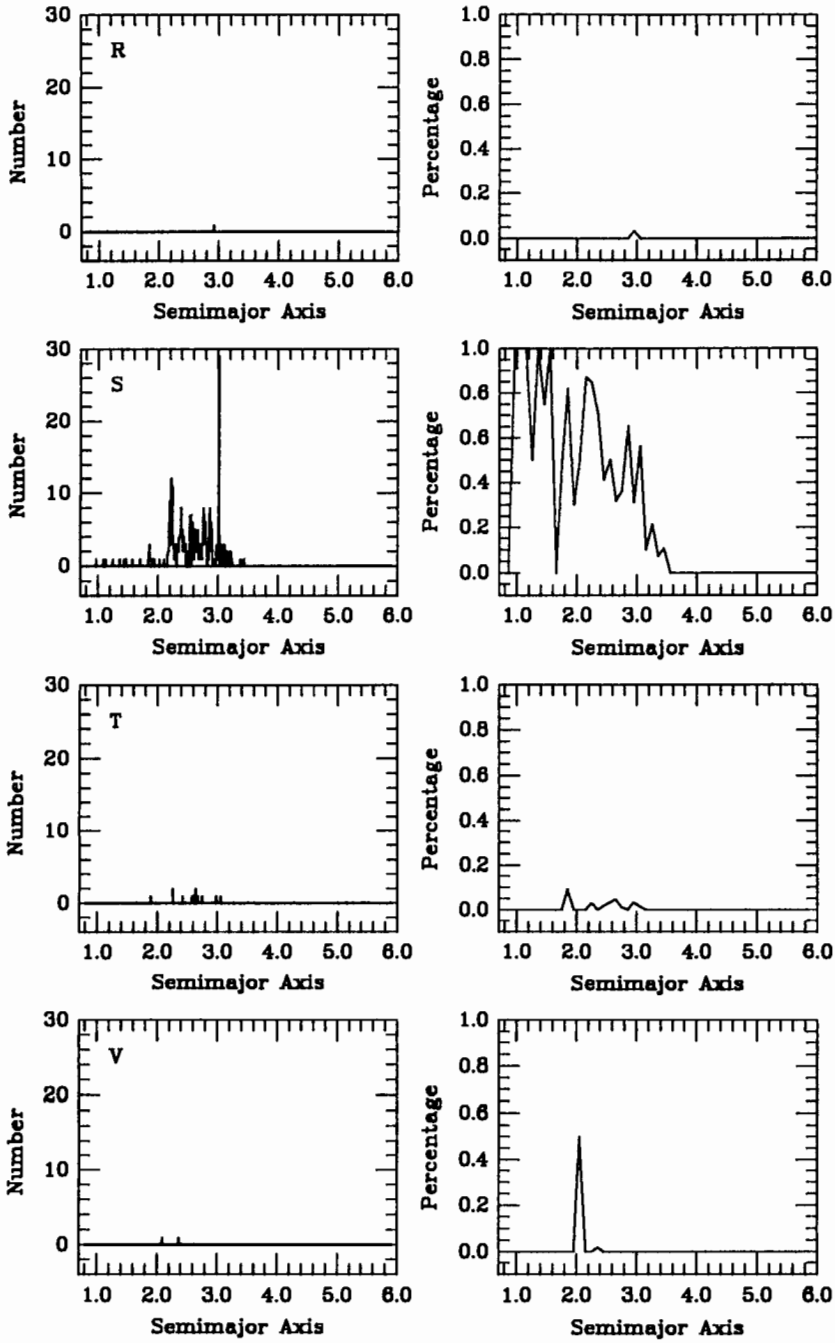
it is theoretically possible for up to 40% of the near-Earth asteroid population to be extinct comment nuclei.

Asteroids can be classified into groups based on orbital parameters (Valsecchi et al. 1989; Zappala et al. 1990) or on the basis of reflectance spectra (Tholen and Barucci 1989). The different orbital classifications all use different criteria to define groups of asteroids with similar orbits which are called families. Families are thought to be the pieces of disrupted bodies, but recent spectral analyses have called the validity of many of the smaller families into question (Chapman et al. 1989). There are also a number of spectral classification systems currently in use, and earlier papers use still other, older, less detailed systems. Tholen and Barruci (1989) present the




D
E
F
G
Number
Percentage
Semimajor Axis
0
10
20
30
0.0
0.2
0.4
0.6
0.8
1.0
1.0
2.0
3.0
4.0
5.0
6.0


K
M
P
Q
Number
Percentage
Semimajor Axis
0
10
20
30
0.0
0.2
0.4
0.6
0.8
1.0
1.0
2.0
3.0
4.0
5.0
6.0



Figures 2a–d. Distribution of classified asteroids, as a function of orbital semi-major axis. The left panels show histograms for each asteroid class, at spatial resolution of 0.01 AU. The right panel shows the fractional distribution of each type, at a spatial resolution of 0.1 AU. Only the typed asteroids were used in determining the fractions.

best summary of the different systems, illustrating some of the differences in the criteria used, and the results obtained. This chapter will use the Tholen classification, which is based on the analysis of Eight Color Asteroid Survey (ECAS) spectra (0.3–1.1 μm) of 405 asteroids (Zellner et al. 1985). The classification uses principal components analysis to define asteroids with similar spectral properties (Tholen and Barucci 1989).

Figure 2 shows the spatial distribution of each taxonomic class of the asteroids. The left panels show histograms for the individual types, and the right panels show the fractional abundance of each type as a function of semimajor axis. It is obvious from this figure that the abundance of the different asteroid classes varies with heliocentric distance. If the near-Earth asteroids are ignored (all asteroids with semimajor axes $\leq$1.5 AU) the variations are highly systematic. This is the best evidence that the main belt and Trojan asteroids have not been perturbed far from their original formation orbits (Gradie et al. 1989). The near-Earth asteroids are a heterogeneous group, including most of the taxonomic types found in the main asteroid belt, as well as one, (Q), which has not yet been found in the main belt. This may be a real difference in populations, although it is more likely to be an observational bias. It is possible to detect much smaller objects in the near-Earth population than in the main-belt population, and if the distribution of taxonomic classes varies as a function of size, some classes might not be detectable in the main belt with current technology (Bell et al. 1989).

Our Fig. 2 differs from many published figures of distribution by type because it shows only the known asteroids. The true population of asteroids is much larger than the observed population, due to limitations on the number of hours available on telescopes to observe, and to the instrument limitations which define the darkest, and therefore the smallest object which can be detected. Many of the published figures, such as those in Gradie et al. (1989) have made attempts to extrapolate to the actual number of asteroids by correcting for the observational bias against dark objects. All of these extrapolations depend on the taxonomy and the assumptions used, therefore no two are alike.

III. ASTEROID SPECTROPHOTOMETRY

Much of what we know about the asteroids has been learned from telescopic spectrophotometry. The principal is very simple; different minerals absorb light at different wavelengths, producing reflectance spectra with characteristic, wavelength-dependent absorption features (see, e.g. S. J. Gaffey et al. 1992). The reflectance spectra of asteroids can be analyzed by comparing the features observed telescopically with the extensive data base of laboratory reflectance spectra of minerals and meteorites (M. J. Gaffey et al. 1989). Figure 3 shows typical reflectance spectra of some of the common minerals found in meteorites. Pyroxene and olivine are present in most meteorite types. They are readily identified by their distinctive absorption features in the near

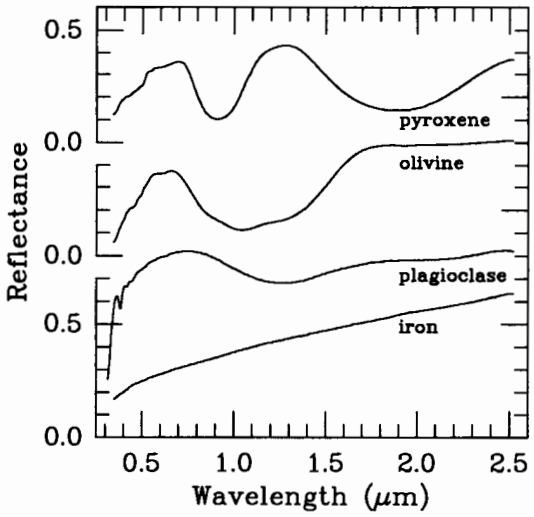


Figure 3. Reflectance spectra of pure minerals. The pyroxene is hypersthene from the Johnstown diogenite meteorite. The olivine is from the Chassigny meteorite. The plagioclase is a lunar anorthite. The metal is an iron meteorite. All of these spectra are directional hemispherical reflectance. The three meteorite spectra are from M. J. Gaffey (1976) and the lunar anorthite is an unpublished spectrum from J. Adams.

infrared. Olivine has an asymmetric absorption feature at 1 μm. Most pyroxenes have two absorption features at 1 μm and 2 μm. Laboratory studies have shown that the positions of these absorption features vary as a function of mineral chemistry (King and Ridley 1987; Cloutis and Gaffey 1991). Plagioclase is found mainly in the chondrite meteorites, the eucrite/diogenite/howardite meteorites, and in silicate inclusions in iron meteorites. It is more difficult to detect spectroscopically because of the weakness of its 1.25 μm absorption feature. Opaque components such as metal and troilite are found in most meteorite types. These compounds are difficult to detect spectroscopically because they do not produce distinct absorption features. Their presence in a mineral mixture can often be deduced from an overall lowering of the albedo and decrease in the spectral contrast. The lack of distinctive absorption features makes it impossible to determine which opaques are present from the spectra alone.

Interpretation of telescopic spectra is complicated by two factors, grain size and intimate mineral mixtures. Figure 4 illustrates the effect that grain size has on both the albedo and the depth of absorption features. These features are functions of the pathlength of the light through the mineral grains (M. J. Gaffey et al. 1992). Light is scattered out of small grains very rapidly, so the

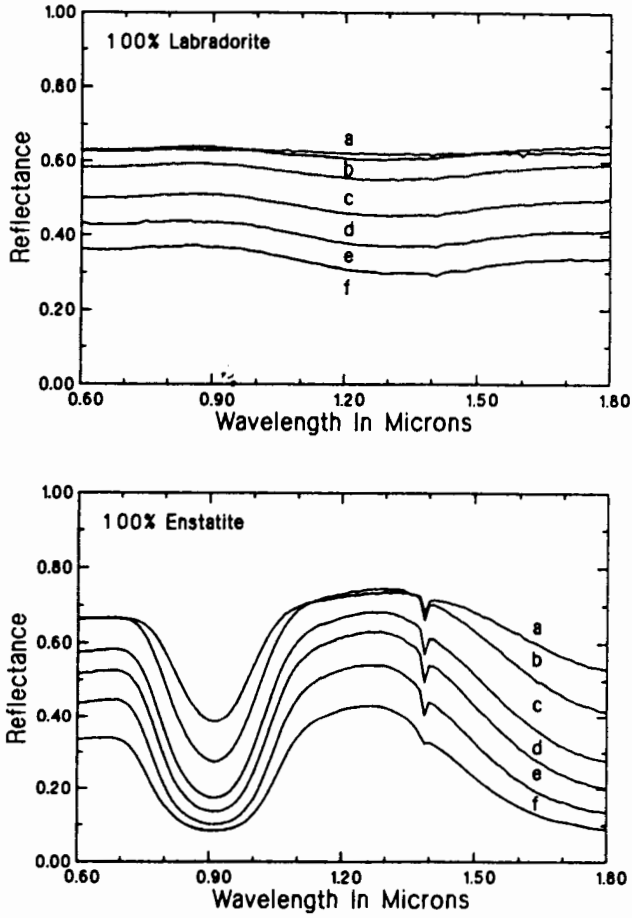


Figure 4. Bidirectional reflectance spectra (0.60 to 1.80 μm) of particle size separates of labradorite and enstatite. Particle sizes are (a) <25 μm; (b) 25–45 μm; (c) 45–75 μm, (d) 75–125 μm; (e) 125–250 μm; and (f) 250–500 μm (figure from Crown and Pieters 1987).

path lengths are short, which minimizes absorption. The path lengths in large grains are long, which maximizes absorption at all wavelengths. In addition, most natural surfaces are intimate mixtures of several minerals. The spectra of mixtures are dominated by the spectrum of the darkest component at a given wavelength, so the resultant mixture spectrum is a non-linear combination of the spectra of the individual components, as seen in Fig. 5 (S. J. Gaffey et al. 1992). Adams (1974) describes the basic approach for analyzing spectra of intimate mineral mixtures, and M. J. Gaffey et al. (1989) review most of the quantitative interpretive techniques which have been developed; they also discuss the logic involved in the application of reflectance spectroscopic

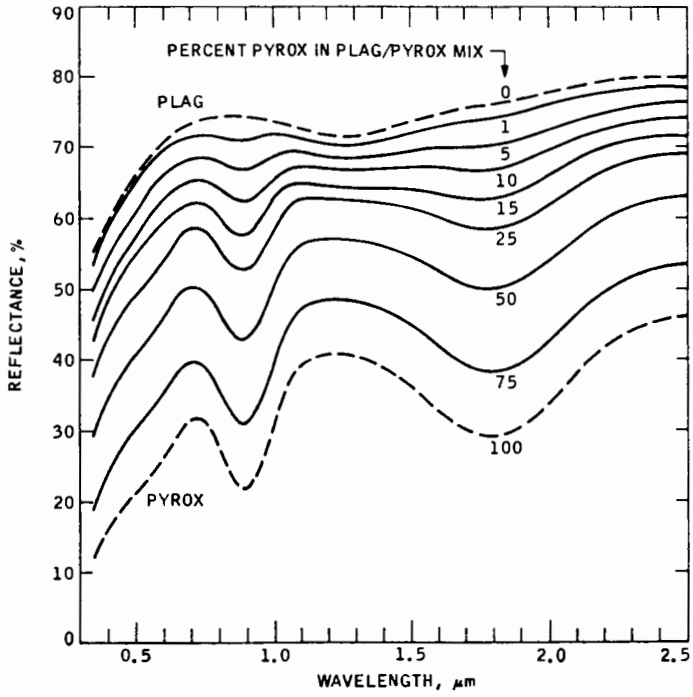


Figure 5. Reflectance spectra of a series of mixtures of labradorite (plagioclase) and hypersthene (pyroxene) (figure from Nash and Conel 1974).

techniques specifically to asteroids.

Mean (ECAS) reflectance spectra of the various Tholen asteroid classes are shown in Fig. 6. In general, these spectra lack the spectral resolution and wavelength coverage to really make compositional interpretations, but it is useful to show them because this is the dataset that was used to define the classes. The classes identify objects with similar spectra over this wavelength range, but they do not necessarily define actual compositional classes (M. J. Gaffey et al. 1989). Longer wavelength infrared data has shown that several of the classes must contain members with significantly different compositions (see, e.g., Lebofsky et al. 1990). Britt (1991) and Britt and Pieters (1991) have raised the additional possibilities that asteroids with different compositions may actually look very similar, and those with similar compositions may look very different. This will be discussed in more detail in the following section.

Comprehensive reviews of asteroid spectrophotometry have recently been published by M. J. Gaffey et al. (1989,1992). We have summarized the main points here for each Tholen asteroid class. The figures used in the rest of this section illustrate the spectral variability within each class. They include spectra from Bell et al. (1992, in preparation), and Lebofsky et al. (1990), which survey the asteriods at longer wavelengths than those used to

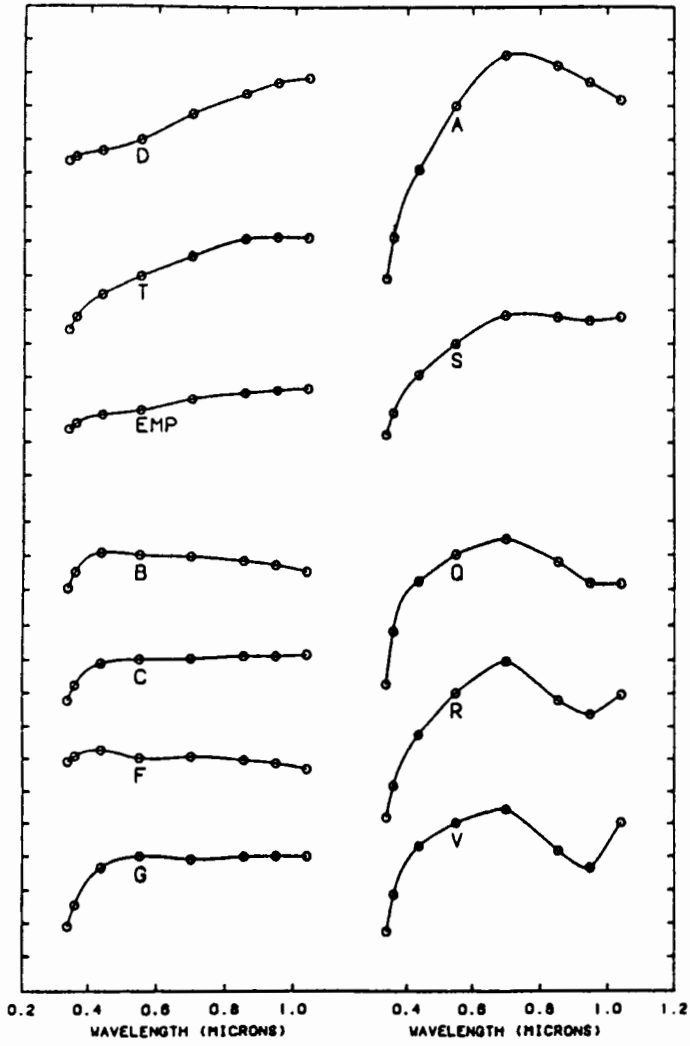


Figure 6. Mean reflectance spectra for the Tholen asteroid classes. Tick marks on the ordinate are spaced 0.2 magnitudes apart (figure from Tholen and Barucci 1989).

define the spectral classes.

A. A-Type Asteroids

Spectra of the A asteroids show an asymmetric absorption band with a minimum near 1 μm, which is characteristic of olivine (Fig. 7). The surface composition has been interpreted to be either pure olivine, or a mixture of olivine and metal, depending on the actual grain sizes and surface textures. A-type asteroids are relatively rare in the main asteroid belt, although there is some indication that they are more abundant in the population of small

(<10 km) asteroids (R. Binzel, personal communication). One A-type asteroid has been found in the near-Earth asteroid population.

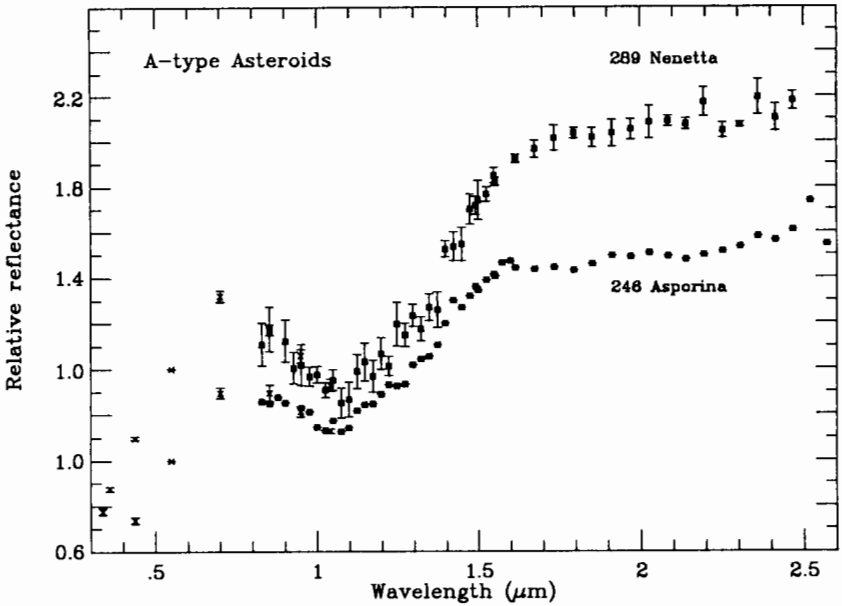


Figure 7. Reflectance spectra of two A-type asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

B. B-, C-, F-, and G-Type Asteroids

The spectra of B, C, F, and G asteroids are generally relatively flat and featureless, with an ultraviolet drop-off (Figs. 8 and 9). The types are distinguished by differences in the ultraviolet absorption. Comparison with meteorite spectra and various laboratory analogs has led to a compositional interpretation of silicate chondrules (composed of olivine and pyroxene), in a matrix of phyllosilicates mixed with carbon, organics, and opaques. Cruikshank and Brown (1987) may have spectrally detected organics on one G-type asteroid, but the observation is still unconfirmed. Clark et al. (1990) have also reported the presence of an ammonium-bearing component on Ceres, which is another G-type. Lebofsky and co-workers have surveyed these asteroid classes in the 3 *mu*m region looking for absorption features caused by water in the phyllosilicates. Figure 10 shows a few of their best spectra. The latest results from their survey work (Lebofsky et al. 1992) indicate that the phyllosilicates are hydrated on approximately half of the C-type asteroids and most of the B- and G-type asteorids. No absorption features attributed to water have yet been identified on the F-type asteorids. The differences in

the 3 micrometer spectra amoung the C-type asteroids imply important compositional differences which are discussed in more detail in the next section, and in Lebofsky and Britt (1992). A new spectral classification by Howell et al. (1993), including the 52-color data identifies two additional sub-groups within the large C-class, but they point out that the dataset is still too small for at these wavelengths to fully define new classes. Figure 8. C-type asteroids are very common, but the other types are significantly less common, perhaps due to observational biases. All of these asteroid types are found primarily in the main asteroid belt. The relative abundance of C-types increases with semimajor axis through the main belt. One C- and one F-type asteroid have also been found in the near-Earth asteroid population.

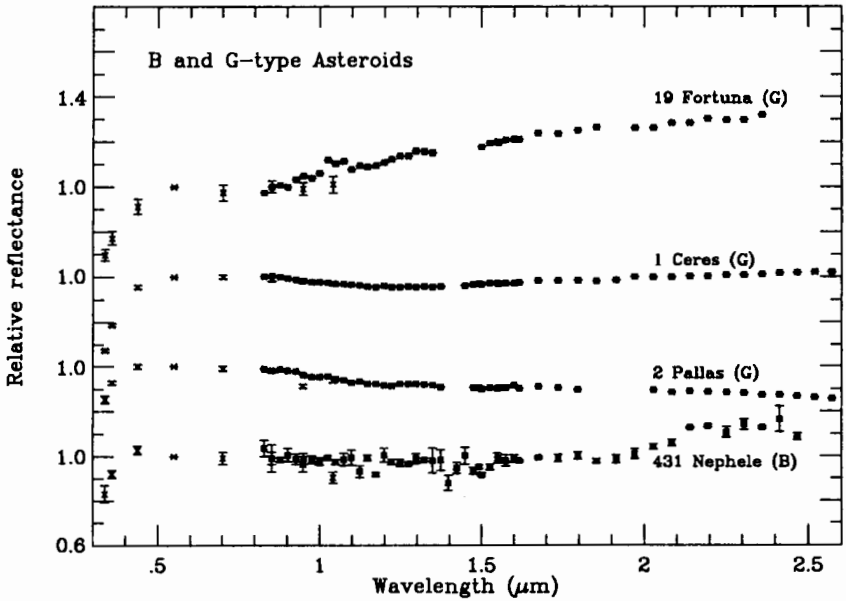


Figure 8. reflectance spectra of one B-class and three G-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

C. D-, P- and T-Type Asteroids

The spectra of D-, P-, and T-type asteroids are red sloped and featureless (Figs. 11 and 12). Lebofsky et al. (1990) observed some as yet unidentified features in the 3 μm region on asteroids of these types, but the water absorption feature has only been detected in one main-belt D-type asteroid (Lebofsky et al. 1992). These spectra are difficult to interpret, and there are no known meteorite analogs. Cosmochemical arguments have been advanced to suggest a composition of organic rich silicates, carbon, and anhydrous silicates, with

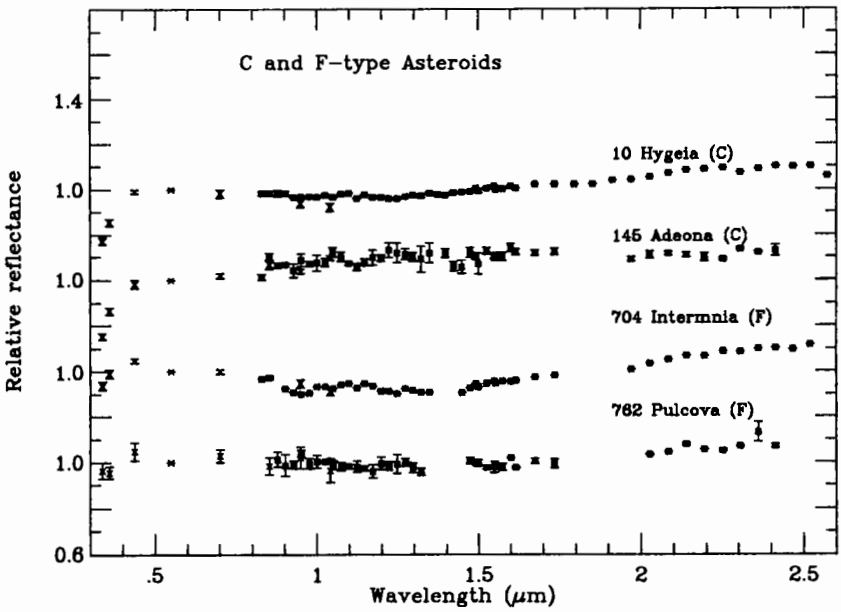


Figure 9. Reflectance spectra of two C-class and two F-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

water ice possibly preserved in the interiors (Bell et al. 1989). The D- and P-type asteroids are concentrated in the outer asteroid belt and beyond. The T-type asteroids are found in the inner main asteroid belt.

D. E-Type Asteroids

Spectra of the E asteroids are also flat and featureless (Fig. 12). They differ from those of the other spectrally similar asteroids because they have much higher albedos. Comparison with spectra of meteorites and terrestrial minerals has suggested an essentially pure enstatite composition. They are found primarily in the inner main asteroid belt.

E. M-Type Asteroids

Spectra of the M-type asteroids are also flat and featureless, with intermediate albedos (Fig. 13). Lebofsky et al. (1990) have also reported 3 μm absorption features, indicative of water on some M-type asteroids, as seen in Fig. 10. M-type asteroids have been interpreted as metal surfaces, with a possible silicate component, similar to the iron meteorites. Lupishko and Belskaya (1989) have argued based on phase functions and polarimetry that the silicates may comprise up to 50% of the surface. The water absorption feature observed in some M-types also suggests that objects with significantly different compositions may be included in this class (Lebofsky and Britt 1992). This question is still unresolved at this time. M-type asteroids are found in the main belt.

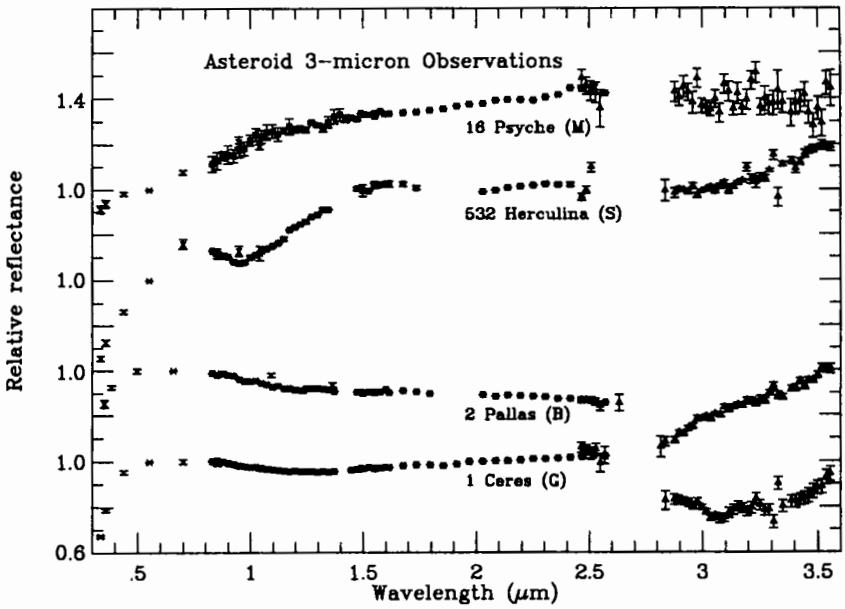


Figure 10. Reflectance spectra of two C-type and two F-type asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

F. Q-Type Asteroids

Only one Q asteroid is known to date, although three other asteroids are spectrally more similar to Q-type than any other type, so have been included in this class at this time. The significance of this class far exceeds its abundance because the Q-type asteroid is spectrally more similar to ordinary chondrite meteorites than any other asteroid type. The spectrum shows a clear 1-μm absorption feature (Fig. 6). The shape of the band indicates that both olivine and pyroxene are present, and the overall slope indicates the presence of metal. The principle Q-type asteroid is found in the near-Earth population, and the other similar asteroids are found in the inner belt.

G. R-Type Asteroids

The one R-type asteroid is spectrally very similar to the V-type (Fig. 14). The spectrum shows distinct olivine and pyroxene absorption features at 1 μm and 2 μm, with the possibility of plagioclase. The R-type asteroid is found in the main belt.

H. S-Type Asteroids

Controversy has raged for years over the composition of the S-type asteroids, and whether or not they are the parent bodies of the ordinary chondrite meteorites. This is discussed in more detail in the next section. The spectra show

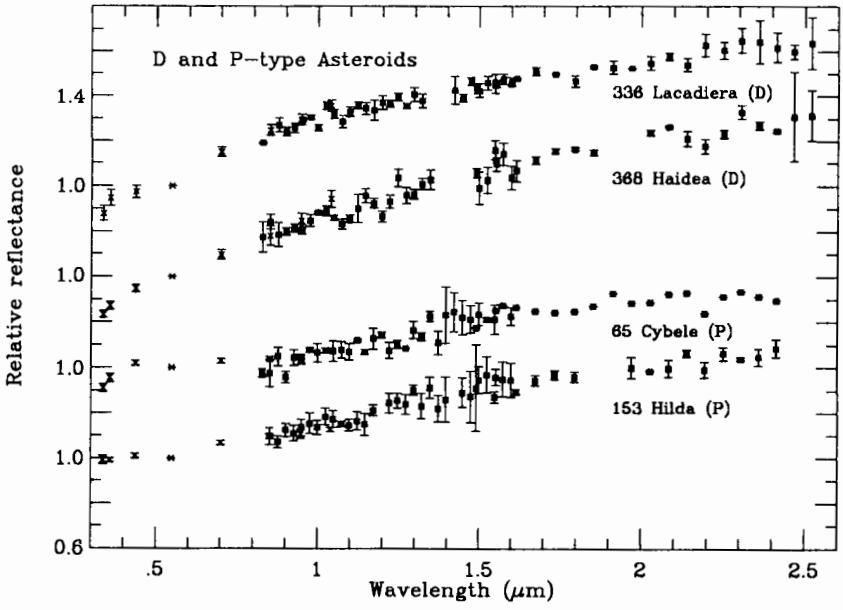


Figure 11. Reflectance spectra of two D-class and two P-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from 52-color asteroid survey (Bell et al. 1992, in preparation).

1 μm and 2 μm absorption features indicative of olivine and pyroxene, and slopes indicative of metal (Fig. 15). The spectral slopes, band shapes and band depths vary considerably within the class, which has led several authors to suggest additional subdivisions (Bell 1988; Tedesco et al. 1989; M. J. Gaffey et al. 1992; Howell et al. 1993). A consensus seems to have emerged that the S class includes asteroids with a range of compositions. S-type asteroids are found primarily in the inner main belt, and they constitute a significant fraction of the near-Earth asteroid population (Fig. 16.)

I. V-Type Asteroids

Vesta is the type example of the V-type asteroids. Until the recent discovery of several more in the near-Earth asteroid population (Cruikshank et al. 1991), it was a single member class. The spectra show clear 1 and 2 μm absorption features characteristic of olivine and pyroxene, and a weak absorption feature at approximately 1.25 μm which is indicative of plagioclase (Fig. 14). Comparison with spectra of meteorites and terrestrial analogs shows that the predominant rock type on the surfaces of the V asteroids is basalt.

IV. ASTEROID AND METEORITE RELATIONSHIPS

There are strong observational, dynamical, orbital, and spectroscopic arguments that most meteorites do indeed come from the asteroid belt and therefore

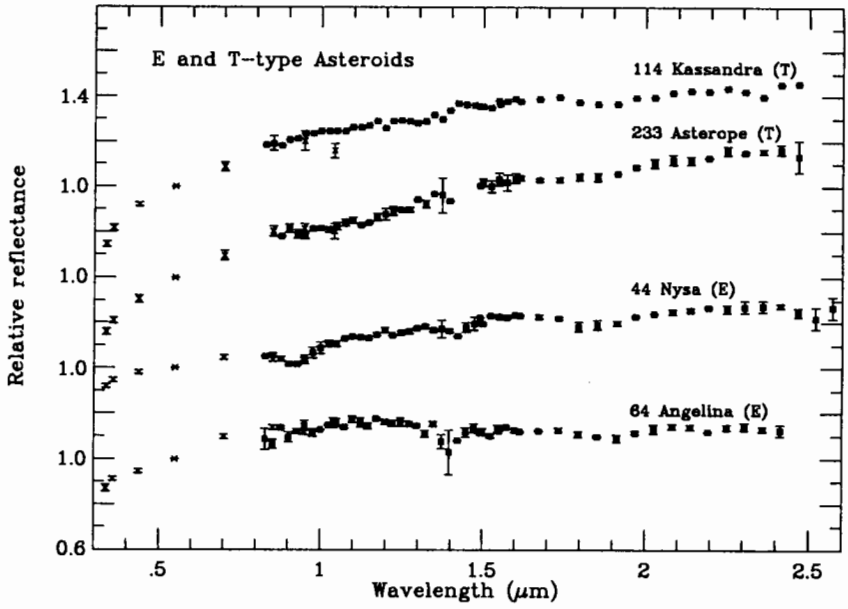


Figure 12. Reflectance spectra of two T-class and two E-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

represent samples of asteroidal material (Wetherill and Chapman 1988; Greenberg and Nolan 1989; Bell et al. 1989). These "free" samples greatly simplify the task of unraveling asteroidal mineralogy and evolution. However, even with this material, linking individual asteroid types with meteorite compositional classes is not trivial. There are several factors that combine to bias the population of meteorites arriving on Earth and therefore limit our sample of the asteroid belt. First, the dynamical processes that deliver meteorites from the asteroid belt to Earth are probably strongly biased toward sampling relatively narrow zones in the asteroid belt. Calculations suggest that the vast majority of meteorites and planet crossing asteroids originate from near the 3:1 Kirkwood gap and the ν 6 resonance (Wetherill 1985; Wisdom 1987). Both these zones are in the inner asteroid belt and probably sample primarily differentiated and relatively less primitive material (Bell et al. 1989). A second factor that introduces bias in the meteorite collection is the relative strength of the meteorites (Wasson 1985). Many meteorites begin the process of evolving into an Earth-crossing orbit by being ejected at high velocity from the parent body by a major impact (Greenberg and Nolan 1989). To survive the stress of impact and acceleration without being crushed into dust the meteorite must have substantial cohesive strength (Stoffler et al. 1988). This would strongly select against the relatively weak (and rare) carbonaceous chondrites. At the opposite extreme, the almost completely metallic

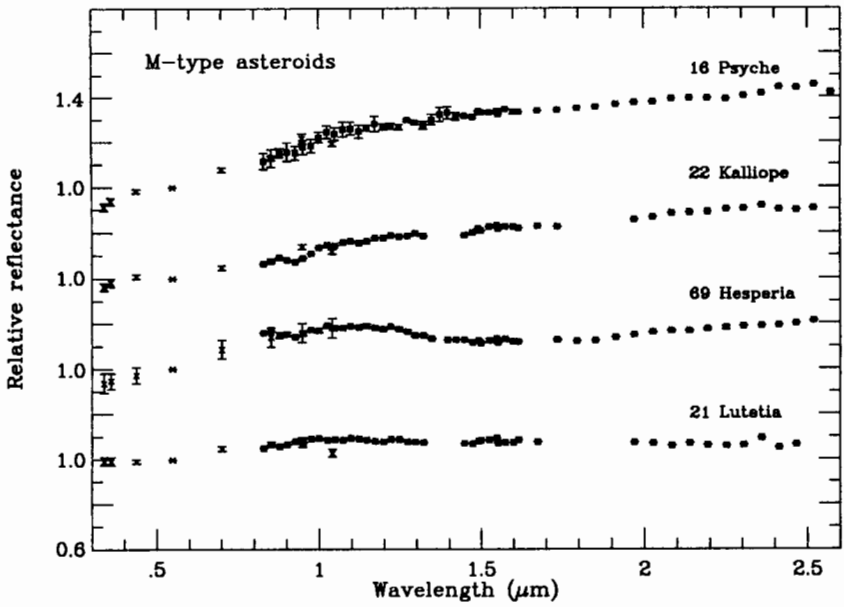


Figure 13. Reflectance spectra of four M-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

iron meteorites and the mostly metallic stony-irons have such great cohesive strength that it would be difficult to break off pieces for ejection. Cohesive strength is also an important selection factor for the meteorites that survive collisions while they are in near-Earth space and for material that survives the deceleration and heating of atmospheric entry (Wasson 1985). Data from cosmic-ray exposure ages of meteorites indicate that collisions in near-Earth space are common and that many meteorites are fragments of larger fragments that were broken up while in Earth-crossing orbits (Wetherill and Chapman 1988; Wasson 1985). Atmospheric entry usually involves a variety of thermal and dynamical stresses that typically break up most stony meteorites from one large individual into showers of much smaller stones (Wasson 1985).

Another major problem is the effect the space environment can have on the surface material of asteroids. Most of our knowledge of asteroid composition comes from the interpretation of remotely sensed reflectance spectra. Our studies of the lunar surface have shown that the spectral characteristics of the surface material, the regolith, can be strongly altered by its exposure to the impact, thermal, and radiation environment of space (Pieters et al. 1985). These regolith processes have been shown to also alter the spectra of some meteorite types, further complicating the identification of asteroid mineralogy (Britt and Pieters 1991). These selection and alteration effects introduce biases of unknown magnitude into the link between asteroids and meteorites that can

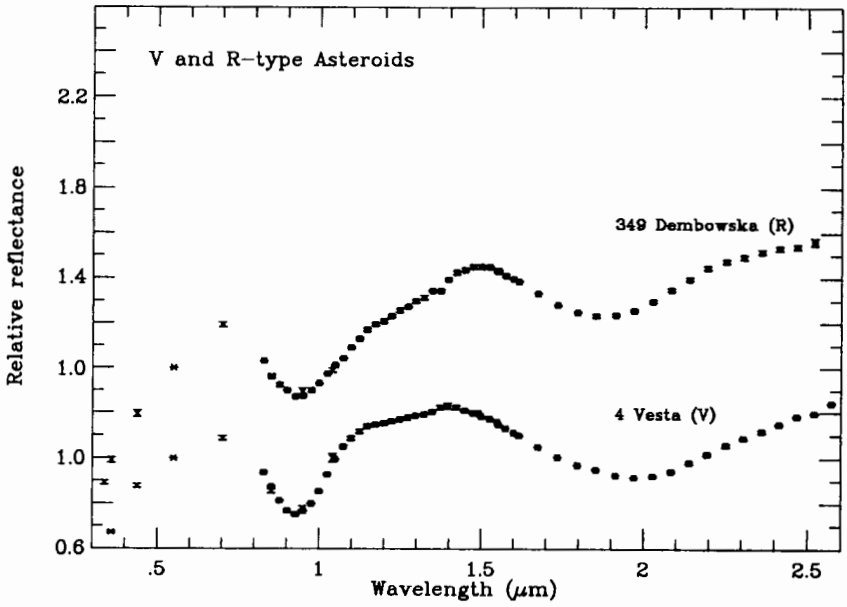


Figure 14. Reflectance spectra of one V-class and one R-class asteroid, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

limit the usefulness of the meteorite collection as a representative sample of the asteroid belt.

Any link between asteroids and meteorites must be an extrapolation of a limited and biased sample of a much larger and more complex population. With the caveats of the previous paragraphs in mind, shown in Table I are the mineralogical interpretation and possible meteoritic analogs for the Tholen (1984) asteroid classes. The organization and composition of the meteorite classes are reviewed in the chapter by Lewis and Hutson. Table I is organized in two groups, the first being asteroid types common in the inner asteroid belt (and thus more likely to have meteoritic counterparts) and the second being asteroid types common in the outer belt. The outer belt is dominated, as shown in Fig. 2, by the P and D types. The spectra of these types are dark, red to very red, generally anhydrous, and relatively featureless. No direct analogs for these asteroid types exist in the meteorite collections. The analogs most commonly cited are cosmic dust or CI carbonaceous chondrites that are enriched in organics (French et al. 1989; Lebofsky et al. 1990). However, the spectral characteristics of these asteroids are difficult to duplicate with material that is delivered to the inner solar system. Probably P and D asteroids are composed of primitive materials that have had a different geochemical evolution than cosmic dust or CI chondrites. Inner asteroid belt objects are much more spectrally varied than those of the outer belt and

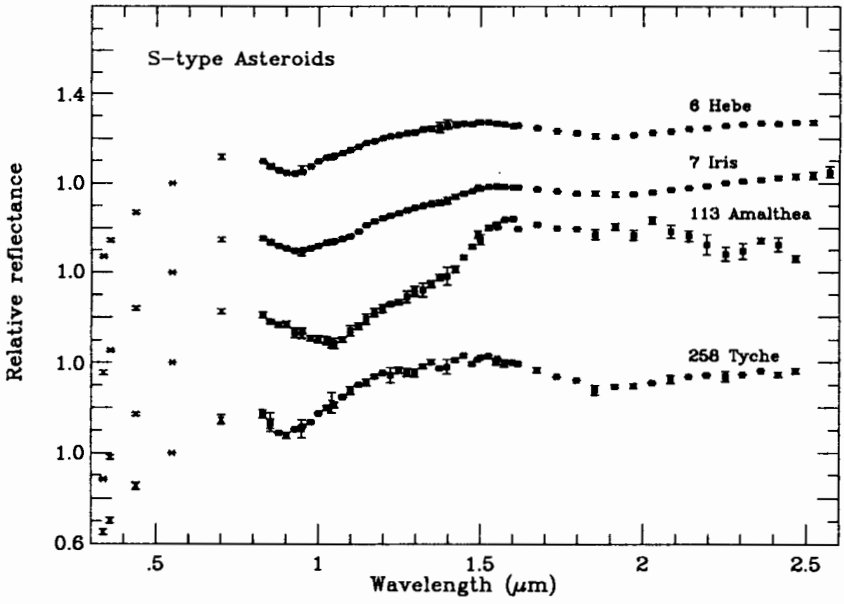


Figure 15. Reflectance spectra of four S-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

include materials which have probably been heated to some degree. This variable heating probably produced varying degrees of alteration, ranging from aqueous alteration by migrating fluids, through thermal metamorphism, to melting and differentiation. The degree of alteration is thought to be a function of composition, radius, and heliocentric distance (Bell et al. 1989).

Perhaps the best asteroid/meteorite spectral match are the V-type asteroids (Fig. 10) with the basaltic achondrite meteorites (McCord et al. 1970; Cruikshank et al. 1991). Spectrally V-types are interpreted to be a differentiated assemblage of primarily orthopyroxene with varying amounts of plagioclase (M. J. Gaffey et al. 1989), which makes them very close analogs to the basaltic howardite/eucrite/diogenite (HED) association of meteorites. The petrology of these meteorites indicate that they are basaltic partial melts originating on asteroids that underwent extensive heating and differentiation. Thus, these meteorites probably represent the surface melts and upper-crustal rocks of a differentiated asteroid (Dodd 1981). In addition, similarities in petrology, chemical trends, and oxygen isotopes suggest that all the HED meteorites are closely related and probably come from a single parent body. Fragments of what once may have been the V-type parent body may have been identified in near-Earth space. Cruikshank et al. (1991) have recently identified three V-type asteroids in Earth-crossing orbits. These objects have almost identical V-type spectra, have spectral features that suggest they are rocky fragments, and are small enough (<4 km in diameter) to be fragments of the crust of a

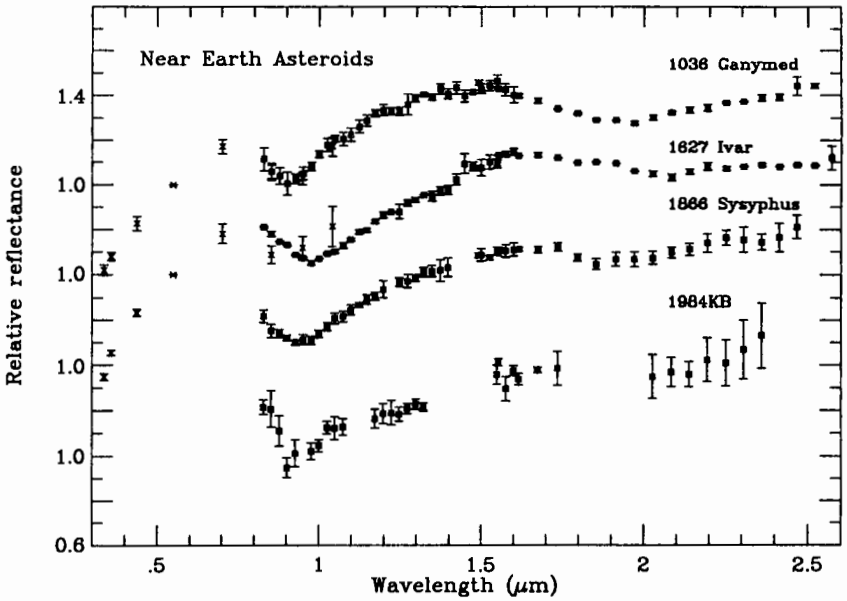


Figure 16. Reflectance spectra of four near-Earth S-class asteroids, scaled to 1.0 at 0.55 μm. The x's are data from the ECAS survey (Zellner et al. 1985) and the squares are data from the 52-color asteroid survey (Bell et al. 1992, in preparation).

disrupted differentiated asteroid (Cruikshank et al. 1991). Interestingly, their orbits are very similar, suggesting that they were all part of a larger fragment that was collisionally disrupted while in an Earth-crossing orbit. In addition, the orbits of these near-Earth asteroids may be related to the peak fall times for HED meteorites. This may be a case where the study of meteorite chemistry, spectra, and fall statistics has come together with remote sensing and orbital dynamics to describe the origin, evolution, and current location of a major meteorite parent body.

The A-types are thought to represent the next lower zone of a differentiated asteroid (Fig. 7). These asteroids are interpreted to be nearly pure olivine and may be derived from the mantle of extensively differentiated parent bodies (Bell et al. 1989; Cruikshank and Hartmann 1984). The best meteorite analog for this interesting asteroid type are the extremely rare brachinites. The 200-gram meteorite Brachina is the only non-Antarctic member of this type. This points up what some workers (see, e.g., Bell et al. 1989) have termed "the great dunite shortage." If some asteroids are really differentiated then there should be, along with the identified fragments of crust (V-types) and core (M-types) material, a substantial amount of mantle material, that is olivine-rich, dunite-like rocks, in the asteroidal and meteorite populations. However, A-type asteroids are relatively rare and dunite-like meteorites are very rare. Another suggested meteorite analog for the A-type asteroids are the

TABLE I

Inner Belt Asteroids

Type	Interpreted Surface Mineralogy	Meteorite Analogs
V	Pyroxene, feldspar	HED association
A	Olivine or olivine-metal	Brachinites
S	Metal, olivine, pyroxene	Pallasites, mesosiderites CV/CO chondritex (K-type), ordinary chondrites
M	Metal, trace silicates (enstatite?)	Irons (enstatite chondrites?)
R	Pyroxene, olivine	none
B,C, F,G	Hydrated silicates, carbon, organics, opaques, shock-darkened silicates?	CI, CM chondrites, black/gas-rich ordinary chondrites?
Q	Olivine, pyroxene, metal	Ordinary chondrites
E	Enstatite	Enstatite achondrites
T	Organic-rich silicates, carbon	none

Outer Belt Asteroids

Type	Interpreted Surface Mineralogy	Meteorite Analogs
D	Organic-rich silicates, carbon	none
P	Organic-rich silicates, carbon	none

somewhat more abundant pallasite meteorites (M. J. Gaffey et al. 1989), but they are also a suggested analog for some S-type asteroids. Another possible mantle-derived asteroid is the R-type which is a single-member class made up of the asteroid 349 Dembowska. Analysis of its reflectance spectra suggests a mineralogy that contains both olivine and pyroxene and may be a partial melt residue of incomplete differentiation (M. J. Gaffey et al. 1989). Unfortunately the meteorite collection contains no potential analogs for this mineralogy.

A more common asteroid type are the M-types (Fig. 13) which have the spectral characteristics of almost pure metal assemblages and are direct analogs to the metallic meteorites (M. J. Gaffey et al. 1989; Ostro 1989). This material may represent the cores of differentiated asteroids. However, there is a great deal of geochemical variety in the iron meteorite population. The 13 different classes of iron meteorites suggest origins from a number of different parent bodies and/or a variety of geochemical conditions. Several M-type asteroids have been identified in Earth-crossing orbits and their metallic composition has been strongly confirmed by radar (Ostro 1989). However, water bands have been observed on two main-belt M-type asteroids (Jones et al. 1990). This strongly indicates that at least some M asteroids are not metal rich and that there may be significant mineralogical variation within the M class.

Perhaps the most complex class of asteroids is the S-type (Fig. 15). Their

spectra, on average, indicate subequal amounts of olivine and pyroxene with a substantial metallic component (M. J. Gaffey et al. 1989). The standard analog for S-types are the pallasites which show an interesting assemblage of large olivine grains set in a matrix of iron-nickel metal. This indicates that some S-types may represent the core-mantle boundary of a differentiated asteroid where the silicates (principally olivine) of the mantle are in direct contact with the metallic core (M. J. Gaffey 1984,1986). However, the S-type group is very large and includes a number of objects that may not conform to this standard interpretation (M. J. Gaffey et al. 1992). A number of S-types are rich in pyroxene and may represent a larger cross section of the mantle and lower crust of an asteroid. Some S-types have lower metal contents and may be the parent bodies of ordinary chondrite meteorites. Other S-types (principally the Eos family) have already been split-out to form the K class (Bell et al. 1989). The K-type asteroids have lower albedos and flatter spectral features than most S-types and are interpreted as analogs for the CV and CO carbonaceous chondrites (Fig. 18). Because of their number, inner belt origins, and moderate albedo, the S-types are well represented in the planet-crossing asteroid population. About half of the identified near-Earth asteroids are S-types and examples of the spectra of four planet crossing S-types are shown in Fig. 16 .

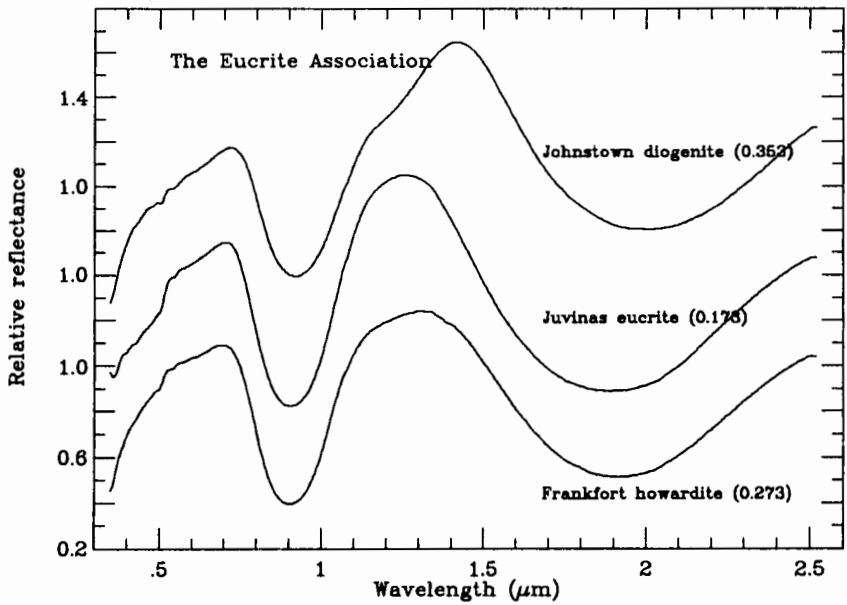


Figure 17. Reflectance spectra three eucrite meteorites from M. J .Gaffey (1976).

A long running controversy in asteroid science has been the identification of the asteroidal source of the ordinary chondrite meteorites. Ordinary

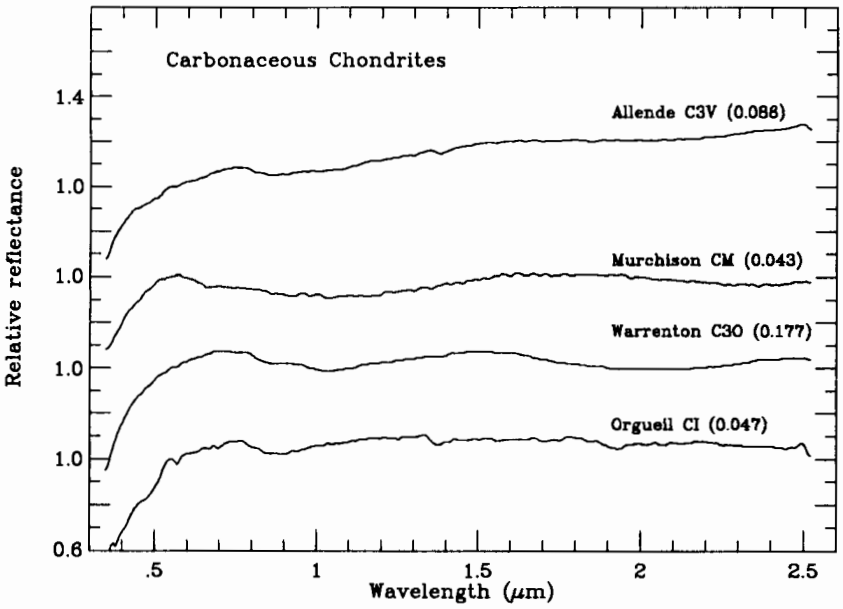


Figure 18. Reflectance spectra of four carbonaceous chondrite meteorites from M. J. Gaffey (1976).

chondrites are by far the largest meteorite type, accounting for approximately 80% of observed meteorite falls, but so far only one asteroid, the small Earth-crosser 1864 Apollo has been identified as an ordinary chondrite analog. A number of S-type asteroids have spectral absorption bands roughly similar to those of ordinary chondrites, but S-types typically have a strong red continuum slope that is not seen in ordinary chondrites (M. J. Gaffey 1986). A number of explanations for the lack of ordinary chondrite parent bodies have been put forward:

1. At least some S-types are ordinary chondrites, but regolith processes enrich the metal content of the regolith and increase their apparent spectral red slope (Bell et al. 1989).
2. Ordinary chondrite parent bodies are in the asteroid belt, but 4.6 Gyr of collisions have ground them down to sizes that are too small to see with current telescopes (Bell et al. 1989).
3. Regolith processes can darken ordinary chondrites, so their parent asteroids actually have the spectra of the dark C-type asteroids (Britt and Pieters 1991).
4. Ordinary chondrites do not even come from asteroids, but are actually cometary material (M. J. Gaffey et al. 1989).

Whatever the final answer, this subject promises to be a source of lively debate for some time in the future.

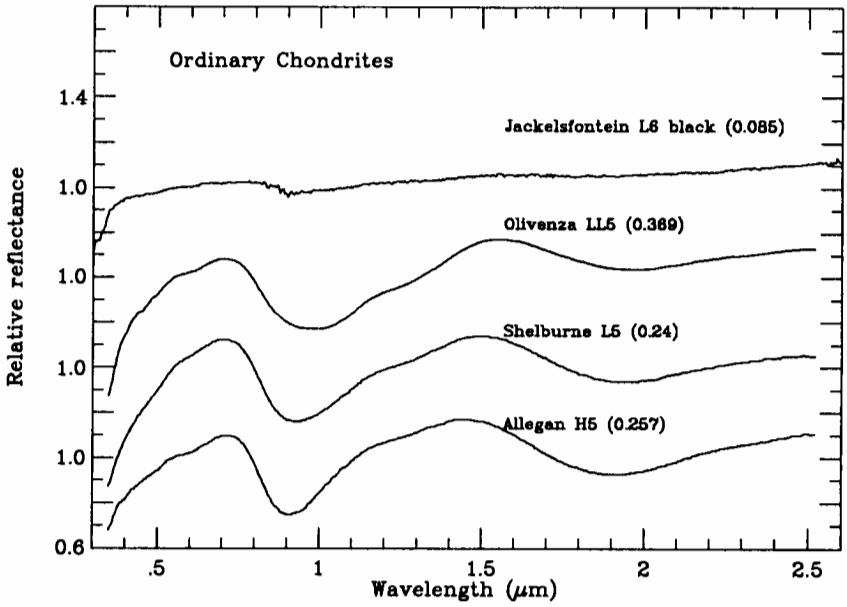


Figure 19. Reflectance spectra of four ordinary chondrite meteorites from M. J. Gaffey (1976) and Britt (1991).

In general, the differentiated asteroids of the V, A, R, S, and M types probably

represent a transect from the crust to the core of differentiated asteroids and as such they can tell us a great deal about the geochemical evolution of a differentiating body. The V-type asteroids would be the surface and crustal material. The A-types would be from a completely differentiated mantle while the R-type would represent a mantle that experienced only partial differentiation. S-types would be the core-mantle boundary with large silicate crystals directly in contact with the metallic core or more mantle material. And finally, M-types represent samples of the metallic cores of these asteroids. Metallic meteorites are shown in Fig. 20.

The dark inner belt asteroids of the B, C, F, and G types (Figs. 8 and 9) are characterized by relatively featureless flat spectra. The proposed analogs for these asteroids are the dark CI and CM carbonaceous chondrite meteorites (Fig. 18) and the spectral differences between the asteroid types are thought to represent varying histories of aqueous alteration or thermal metamorphism (Bell et al. 1989). Many of these asteroids show hydration features indicating they may be valuable sources of mineralogically bound water (Lebofsky et al. 1990). However, the compositions of these asteroid types are probably more complex than this simple model suggests. The B, F, and G types all tend to show hydration features, but at least 40% of the C-types are anhydrous

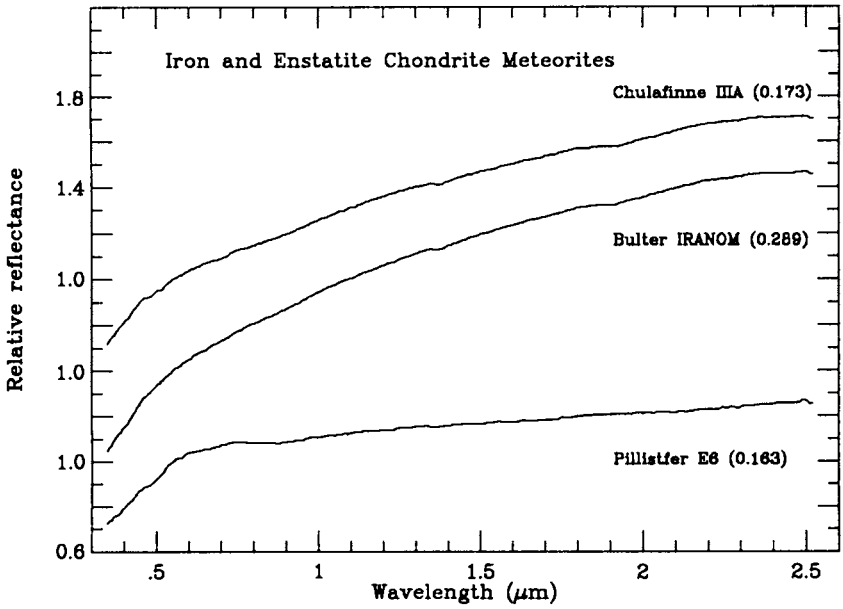


Figure 20. Reflectance spectra of two iron and one enstatite chondrite meteorites from M. J. Gaffey (1976).

(Lebofsky et al. 1990). The C-type may be, like the S-type asteroids, a collection of objects with roughly similar reflectance spectra but varying mineralogy. One suggestion is that regolith processes can darken and flatten the spectra of ordinary chondrite material, making an ordinary chondrite parent body with a mature regolith appear to be spectrally a C-type asteroid (Britt and Pieters 1991*a*, *b*). The E-type asteroids are excellent analogs for bright, red-sloped, but spectrally featureless enstatite achondrites. These asteroids are probably composed of the same differentiated enstatite assemblages as the enstatite achondrites (M. J. Gaffey et al. 1989). The T-type asteroids are something of an enigma. They are rare, probably anhydrous, inner-belt asteroids with dark, featureless, and moderately red spectra. These asteroids have no direct meteorite analogs and may be related to the P- and D-type asteroids of the outer belt (Lebofsky et al. 1990).

V. CONCLUSIONS

As we have seen in the previous discussions, there are many questions remaining about the detailed composition of the asteroids. Specifically, there are major types of meteorites for which we still do not know the asteroid parent bodies and there are major classes of asteroids for which we do not seem to have meteorite analogs. Also, because the taxonomic classes are determined solely from visual spectra, concerns have been raised as to the relationship of

classes to mineralogy and the possibility that an individual class may contain asteroids of greatly differing mineralogy.

We must therefore caution any mineralogical interpretation of asteroids based solely on asteroid taxonomic classes. Until we can observe a range of classes close-up, there will remain doubts in the minds of asteroid scientists about the true compositional relationships between the various taxonomic classes and their meteorite analogs.

REFERENCES

Adams, J. B. 1974. Visible and near-infrared diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system. *J. Geophys. Res.* 79:4825–4836.

Bell, J. F. 1988. A probable asteroidal parent body for the CV or CO chondrites. *Meteoritics* 23:256–257.

Bell, J. F., Davis, D. R., Hartmann, W. K., and Gaffey, M. J. 1989. Asteroids: The big picture. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 921–945.

Binzel, R. P., Xu, S., Bus, S. J., and Bowell, E. 1992. Origins for the near-Earth asteroids. *Science* 257:779–782.

Britt, D. T. 1991. The Meteorite Record As Clues To Asteroidal Regolith Processes. Ph. D. Thesis, Brown Univ.

Britt, D. T., and Pieters, C. M. 1991. Darkening in gas-rich ordinary chondrites: Spectral modelling and implications for regoliths of ordinary chondrite parent bodies. *Lunar Planet. Sci.* XXII:141–142 (abstract).

Chapman, C. R., Paolicchi, P., Zappala, V., Binzel, R. P., and Bell, J. F. 1989. Asteroid families: Physical properties and evolution. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels, and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 386–415.

Clark, R. N., King, T. V. V., Klejwa, M., and Swayze, G. A. 1990. High spectral resolution reflectance spectroscopy of minerals. *J. Geophys. Res.* 95:12653–12680.

Cloutis, E. A., and Gaffey, M. J. 1991. Pyroxene spectroscopy revisited: Spectral-compositional correlations and relationship to geothermometry. *J. Geophys. Res.* 96:22809–22826.

Crown, D. A., and Pieters, C. M. 1987. Spectral properties of plagioclase and pyroxene mixtures and the interpretation of lunar soil samples. *Icarus* 72:492–506.

Cruikshank, D. P., and Brown, R. H. 1987. Organic matter on asteroid 130 Elektra. *Science* 238:183–185.

Cruishank, D. P., and Hartmann, W. K. 1984. The meteorite-asteroid connection: Two olivine-rich asteroids. *Science* 223:281–283.

Cruikshank, D. P., Tholen, D. J., Hartmann, W. K., Bell, J. F., and Brown, R. H. 1991. Three basaltic Earth-approaching asteroids and the source of the basaltic meteorites. *Icarus* 89:1–13.

Davis, D. R., Weidenschilling, S. J., Farinella, P., Paolicchi, P., and Binzel, R. P. 1989. Asteroid collisional history: Effects on sizes and spins. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 805–826.

Dodd, R. T. 1981. *Meteorites: A Petrologic-Chemical Synthesis* (New York: Cambridge Univ. Press).

French, L. M., Vilas, F., Hartmann, W. K., and Tholen, D. J. 1989. Distant asteroids and Chiron. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 468–486.

Froeschlé, Cl., and Greenberg, R. 1989. Mean motion resonances. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 827–844.

Gaffey, M. J. 1976. Spectal reflectance characteristics of the meteorite classes. *J. Geophys. Res.* 81:905–920. Gaffey, M. J. 1984. Rotational spectral variations of asteroid (8) Flora: Implications for the nature of the S-type asteroids and for the parent bodies of the ordinary chondrites. *Icarus* 60:83–114.

Gaffey, M. J. 1986. The spectral and physical properties of metal in meteorite assemblages: Implications for asteroid surface materials. *Icarus* 66:468–486.

Gaffey, M. J., Bell, J. F., and Cruikshank, D. P. 1989. Reflectance spectroscopy and asteroid surface mineralogy. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 98–127.

Gaffey, M. J., Lebofsky, L. A., Nelson, M. L., and Jones, T. D. 1992. Asteroid surface compositions from earthbased reflectance spectroscopy. In *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, eds. C. M. Pieters and A. J. Englert (New York: Cambridge Univ. Press), in press.

Gaffey, S. J., McFadden, L. A., and Nash, D. 1992. Ultraviolet, visible and near-infrared reflectance spectroscopy. In *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, eds. C. M. Pieters and A. J. Englert (New York: Cambridge Univ. Press), in press.

Gehrels, T. 1991. Scanning with charge-coupled devices. *Space Science Reviews* 58:347–375.

Gradie, J. C., Chapman, C. R., and Tedesco, E. F. 1989. Distribution of taxonomic classes and the compositional structure of the asteroid belt. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 316–335.

Greenberg, R., and Nolan, M. C. 1989. Delivery of asteroids and meteorites to the inner solar system. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 778–804.

Helin, E. F., and Dunbar, R. S. 1991. Search techniques for near-Earth asteroids. *Vistas in Astronomy* 33:21–37.

Howell, E. S., Merenyi, E., and Lebofsky, L. A. 1993. Classification of asteroid spectra using a neural network. *J. Geophys. Res.*, submitted.

Jones, T. D., Lebofsky, L. A., Lewis, J. S., and Marley, M. S. 1990. The composition and origin of the C, P, and D asteroids: Water as a tracer of thermal evolution in the outer belt. *Icarus* 88:172–192. King, T. V. V., and Ridley, W. I. 1987. Relation of the spectroscopic reflectance of olivine to mineral chemistry and some remote sensing implications. *J. Geophys. Res.* 92:11457–11469.

Lebofsky, L. A., and Britt, D. T. 1993. Spectral variation within the asteroid classes. *J. Geophys. Res.*, submitted.

Lebofsky, L. A., Jones, T. D., Owensby, P. D., Feierberg, M. A., and Consolmagno, G. J. 1990. The nature of low-albedo asteroids from 3-μm multi-color photometry. *Icarus* 83:16–26.

Lebofsky, L. A., Howell, E. S., and Britt, D. T. 1992. Characterization of low albedo asteroids. *Bull. Amer. Astron. Soc.* 23:1140.

Lupishko, D. F., and Belskaya, I. N. 1989. On the surface composition of the M-type asteroids. *Icarus* 78:395–401.

McCord, T. B., Adams, J. B., and Johnson, T. V. 1970. Asteroid Vesta: Spectral

reflectivity and compositional implication. *Science* 168:1445–1447.
McFadden, L. A., Tholen, D. J., and Veeder, G. J. 1989. Physical properties of Aten, Apollo, and Amor asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 442–467.
Nash, D. B., and Conel, J. E. 1974. Spectral reflectance systematics for mixtures of powdered hypersthene, labradorite, and ilmentite. *J. Geophys. Res.* 79:1615–1621.
Ostro, S. J. 1989. Radar observations of asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 192–212.
Pieters, C. M., Adams, J. B., Mouginis-Mark, P., Zisk, S. H., Head, J. W., McCord, T. B., and Smith, M. 1985. The nature of crater rays: the Copernicus example. *J. Geophys. Res.* 90:12393–12413.
Ruzmaikina, T. V., Safronov, V. S., and Weidenschilling, S. J. 1989. Radial mixing of material in the asteroidal zone. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 681–700.
Shoemaker, C. S., and Shoemaker, E. M. 1988. The Palomar Asteroid and Comet Survey (PACS), 1982–1987. *Lunar Planet. Sci.* XIX:1077–1078 (abstract).
Shoemaker, E. M., Shoemaker, C. S., and Wolfe, R. F. 1989. Trojan asteroids: Populations, dynamical structure and origin of the L4 and L5 swarms. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 487–523.
Stoffler, D., Bischoff, A., Buchwald, V., and Rubin, A.E. 1988. Shock effects in meteorites. In *Meteorites and the Early Solar System*, eds. J. F. Kerridge and M. S. Mathews (Tucson: Univ. of Arizona Press), pp. 165–203.
Tedesco, E. F., Williams, J. G., Matson, D. L., Veeder, G. J., Gradie, J. C., and Lebofsky, L. A. 1989. A three-parameter asteroid taxonomy. *Astron. J.*97:580–606.
Tholen, D. J. 1984. Asteroid Taxonomy From Cluster Analysis of Photometry. Ph. D. Thesis, Univ. of Arizona.
Tholen, D. J., and Barucci, M. A. 1989. Asteroid taxonomy. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 298–315.
Valsecchi, G. B., Carusi, A., Knezevic, Z., and Williams, J. G. 1989. Identification of asteroid dynamical families. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 368–385.
Wasson, J. T. 1985. *Meteorites: Their Record of Early Solar-System History* (New York: W. H. Freeman).
Weissman, P. R., A'Hearn, M. F., McFadden, L. A., and Rickman, H. 1989. Evolution of comets into asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 880–920.
Wetherill, G. W. 1985. Asteroidal source of ordinary chondrites. *Meteoritics* 20:1–22.
Wetherill, G. W. 1989. Origin of the asteroid belt. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 661–680.
Wetherill, G. W., and Chapman, C. R. 1988. Asteroids and meteorites. In *Meteorites and the Early Solar System*, eds. J. F. Kerridge and M. S. Matthews (Tucson: Univ. of Arizona Press) pp. 35–69.
Wisdom, J. 1987. Chaotic dynamics in the solar system. *Icarus* 72:241–275. Zappala, V., Cellino, A., Farinella, P., and Knezevic, Z. 1990. Asteroid families. I. Identification by hierarchical clustering and reliability assessment. *Astron. J.* 100:2030–2046.
Zellner, B., Tholen, D. J., and Tedesco, E. F. 1985. The eight-color asteroid survey: Results for 589 minor planets. *Icarus* 61:355–416.

ASTEROIDAL RESOURCE OPPORTUNITIES SUGGESTED BY METEORITE DATA

JOHN S. LEWIS and MELINDA L. HUTSON
University of Arizona

The meteorites that fall on Earth are necessarily samples of the near-Earth population of small solar system bodies, and hence are a very useful guide to the nature of the materials available in near-Earth space. We review the compositions of the most abundant classes of noncarbonaceous meteorites that fall on Earth for the purpose of identifying economically attractive uses of near-Earth asteroid materials.

I. INTRODUCTION

The near-Earth asteroids (NEAs) are best known as sources of meteorites. They also represent threats to Earth as giant impactors that can disrupt the biosphere catastrophically, as is believed to have happened at the end of the Cretaceous era. But in the long term, there is little doubt that the vast resource potential of the near-Earth asteroids will be of far greater interest. Interestingly (and fortunately), the near-Earth asteroids are compositionally diverse fragments exiled from the main asteroid belt by gravitational perturbations. They sample the belt widely, with km-sized chunks of natural stainless steel, the cores (possibly ice-rich) of extinct comets, primitive unmelted planetary materials, and differentiated rocks similar to lunar basalts all present in glorious profusion. The panoply of materials is vastly broader and richer than those known to be present on the Moon. At least one fifth of the near-Earth asteroids are volatile-rich, with abundances of hydrogen, carbon, nitrogen, etc. over 100 times as high as in the most volatile-rich lunar materials. Almost all the others are metal-rich, again with over 100 times the free metal content of lunar material. Also, about one fifth are energetically more accessible than the surface of the Moon. The main drawback is that the trip times to visit these nearby asteroids and return are about 1 to 3 yr, similar to Mars missions, compared to a one-week round trip to the Moon. Fortunately, schemes are already known by which spacecraft dispatched on round-trip missions to the best of these asteroids could return over 100 times their own mass of asteroidal resources to near-Earth space.

The first products of nonterrestrial raw materials to be transported for use will be materials that are relatively simple to fabricate, and which are needed in very large masses. Thus the first products will probably be rocket

propellants, life-support fluids, structural metals, refractory heat shields (and fire bricks), and radiation shielding. Assuming that civil uses predominate, the raw material of choice for propellants, life-support materials, and shielding (against solar flare protons, trapped particles in Earth's radiation belts, and galactic cosmic ray primaries) is water. Because, in meteorites, water and other useful volatiles generally are found together, volatile-rich meteoritic material is likely to be of prime importance. Ferrous metals seem to be the likely second product, and refractory oxide compounds would either become available as coproducts from processing of volatiles and ferrous metals, or follow as the third-generation product. Thus, as we examine the composition of meteorites, it is these three types of materials that will most strongly attract our near-term interest.

Because of NEAs' role as the supplier of almost all of the meteorites that fall on Earth, we know a vast amount about their chemical and physical properties. These properties in turn have permitted us to reconstruct a great deal of information about conditions in the early solar system and evolutionary processes in populations of small bodies. But because our knowledge of the compositions of individual asteroids is limited to the identification and approximate relative abundances of a few of the most distinctive minerals in only 44 of the 155 known nearby asteroids, our ability to link specific meteorite classes with their asteroidal parents is severely limited. Further, some of the NEAs for which we have some spectral data have no known exact counterparts among our meteorite samples. Thus a broad overview of the nature of meteorites and their resource significance is both desirable and possible, whereas the linkage of the meteorite classes to specific asteroids remains speculative.

A number of introductory books on meteorites are available. We especially recommend Dodd (1981) for its chemical and petrological perspective and Wasson (1974) for its coverage of iron meteorites. The most recent general overview is found in the Space Science Series book *Meteorites and the Early Solar System* edited by Kerridge and Matthews (1988).

II. GENERAL TAXONOMY OF METEORITES

The broad outlines of meteorite taxonomy can best be conveyed by the outline given in Table I. In general, most meteorites date from the time of the origin of the solar system. Their radiometric ages, whether based on retention of radiogenic gases (K-Ar and He-(U,Th) dating) or on nonvolatile isotope systems (Pb-Pb, Pb-(U,Th), Rb-Sr, etc.) usually are concordant at an age of about 4.55 Gyr. A number of meteorites, as a consequence of violent collisional events in the asteroid belt, have received shock damage and suffered sufficient shock heating to at least partially reset the gas-retention clock.

Most meteorites (about 87% of those observed to fall) have about the same elemental composition as the nonvolatile component of solar (galactic) material. Some contain delicate assemblages of materials that are not compat-

TABLE I
Meteorite Taxonomy

Stones: silicate-dominated (96% of all meteorite falls)
Chondrites (88%)
Primitive, unmelted, undifferentiated materials, 4.6 Gyr old
Abundances of rock-forming elements close to solar proportions
Usually contain glassy "droplets" called **chondrules.**
Achondrites (8%)
Very silicate-rich igneous objects (about 99% silicates and oxides)
Formed by density-dependent differentiation
Almost all solidified about 4.6 Gyr ago
Stony-Irons (1% of all meteorite falls)
About 50% ferrous metal alloys, 50% silicates
Irons (3% of all meteorite falls)
About 99% metallic Fe-Ni-Co alloys
Inclusions of FeS, phosphides, carbides, graphite, silicates

ible at chemical equilibrium, including mm-sized "droplets," often containing glass, which are called *chondrules*. Others often show signs of storage at (or heating to) elevated temperature, leading to partial or complete crystallization of glass and equilibration of the other minerals. These primitive types of meteorites, which have not undergone melting and density-dependent differentiation into core, mantle, and crustal-type layers, are collectively referred to as *chondrites*, even though some of them have few or no chondrules. The chondrites span a wide range of oxidation states of iron, from over 30% metal to some that contain no free metal at all, having all their metal in the form of FeO-bearing minerals. The most highly oxidized chondrites are also the richest in volatiles, containing up to 20% water and up to 40% readily extractable volatiles. The chondrites have experienced varying degrees of heating and internal equilibration.

Almost all of the other meteorites are clearly products of melting and density-dependant separation. These include *irons* and *stony irons*, which are chemical analogs of core and core-mantle boundary samples, and *achondrites*, which are analogs of familiar mantle and crustal silicate rocks. The irons make up about 3% of all the meteorites that are observed to fall, the stony-irons contribute a little over 1% of meteorite falls, and the achondrites account for a little over 8%. Thus about 96% are stones. Fully 88% of the observed meteorite falls are primitive, unmelted materials, and only 12% are igneous.

The near-Earth asteroid population contains bodies whose spectra suggest the presence of all of these groups, from a 2-km piece of metal (the asteroid 3554 Amun) to a 4-km piece of basalt (1985 DO2) to a 12-km apparent extinct comet core (1983 SA). The nature of the near-Earth asteroid population is described in the Chapter by Nelson et al.

III. ORDINARY CHONDRITES

The large majority (87%) of all the well-characterized chondrites (Table II) have very similar mineral assemblages, in that they feature abundant olivine $(Mg,Fe)_2SiO_4$, pyroxene $(Mg,Ca,Fe)SiO_3$, plagioclase $NaAlSi_3O_8$-$CaAl_2Si_2O_8$ solid solutions, diopside $CaMgSi_2O_6$, metallic iron-nickel alloy, and troilite FeS (Dodd 1981). They also contain minor amounts (0.2 to 0.6%) of chlorapatite $Ca_5(PO_4)_3(Cl)$, whitlockite $Ca_3(PO_4)_2$, ilmenite $FeTiO_3$, and chromite $(FeCr_2O_4)$. It has been known since the work of Urey and Craig (1953) that these meteorites, called *ordinary chondrites*, are subdivided into three classes with different oxidation states, and hence different proportions of Fe metal and FeO, and also with slightly different ratios of Fe to Si. The most metal-rich of the ordinary chondrites have both a high Fe°:FeO and a high total Fe abunance (Fe:Si = 0.8), and are called the *high-iron* (*H* group) chondrites. Next, with a lower Fe°:FeO and a lower total Fe abundance (Fe:Si = 0.6), are the *low-iron (L group)* chondrites. Finally, less abundant and more recently discovered, are those with the lowest total iron content (Fe:Si = 0.5) and the highest oxidation state, are the *LL chondrites*. These meteorites have nearly identical mineralogy, and seem to reflect formation under very similar conditions on parent bodies with slightly different metal:silicate ratios. Such fractionation may be a natural part of the process of planetary formation. The differences in oxidation state probably reflect differences in their temperature of origin, as metallic iron in the early solar nebula would be most abundant at high temperatures, and would be progressively oxidized at lower temperatures. Therefore the H, L, and LL parent bodies may represent a sequence in heliocentric distance. The iron content and oxidation state of the chondrites groups are compared in Fig. 1.

The metal in the ordinary chondrites varies systematically in composition from class to class (Rambaldi 1977*a*, *b*, *c*). The LL chondrites have the highest Ni content, the lowest Ir:Ni ratio, the lowest Co:Ni ratio, and the highest Au:Ni ratio. As Ir is one of the most refractory elements and Au is one of the most volatile siderophilies, the metal composition implies that the LL chondrites formed at the lowest temperature. This is completely in harmony with the degree of oxidation of the iron, because the FeO content is highest (and the metallic iron abundance the lowest) in the LLs. The details of the metal composition in chondrites will be discussed in the context of our description of iron meteorites.

In addition to bulk chemical composition, it is useful to describe chondrites in terms of their degree of approach to equilibrium or, equivalently, their degree of thermal metamorphism (Van Schmus and Wood 1967). The least-equilibrated chondrites, which are also the most affected by aqueous alteration, consisting entirely of very dark, fine-grained matrix material, metal-free, and containing abundant (3–6%) organic matter and water (about 20%), are referred to as petrologic type 1. Next in the sequence are the chondrites with about half as much volatiles, with very sharply defined chondrules and

TABLE II
Chondrite Taxonomy

Chondrites	
Primitive, unmelted, undifferentiated materials, 4.6 Gyr old	
Abundances of rock-forming elements close to solar proportions	
Usually contain glassy "droplets" called **chondrules.**	
Enstatite (E) chondrites: Less than 1% FeO in silicates	(4% of all chondrite falls)
High-iron enstatite (EH) chondrites: Fe:Si >0.7	(2%)
Low-iron enstatite (EL) chondrites: Fe:Si <0.7	(2%)
Ordinary chondrites: FeO:Si = 0.1 to 0.5	(87% of all chondrite falls)
High-iron (H) chondrites: Fe:Si = 0.8	(31%)
Low-iron (L) chondrites: Fe:Si = 0.6	(46%)
Low-low-iron (LL) chondrites: Fe:Si = 0.5	(10%)
Carbonaceous (C) chondrites: FeO:Si >0.5	(9% of all chondrite falls)
Ivuna/Orgueil-type (C1I) chondrites: No free metal; 10–20% water; 6% organics	(1%)
Murchison-type (C2M) chondrites: No free metal; 5–10% water; 2% organics	(4%)
Ornans-type (CO) chondrites: Coarse-grained; 1% C; 3% water	(2%)
Vigarano-type (CV) chondrites: Fine-grained; 0.3% C; 1% water	(2%)

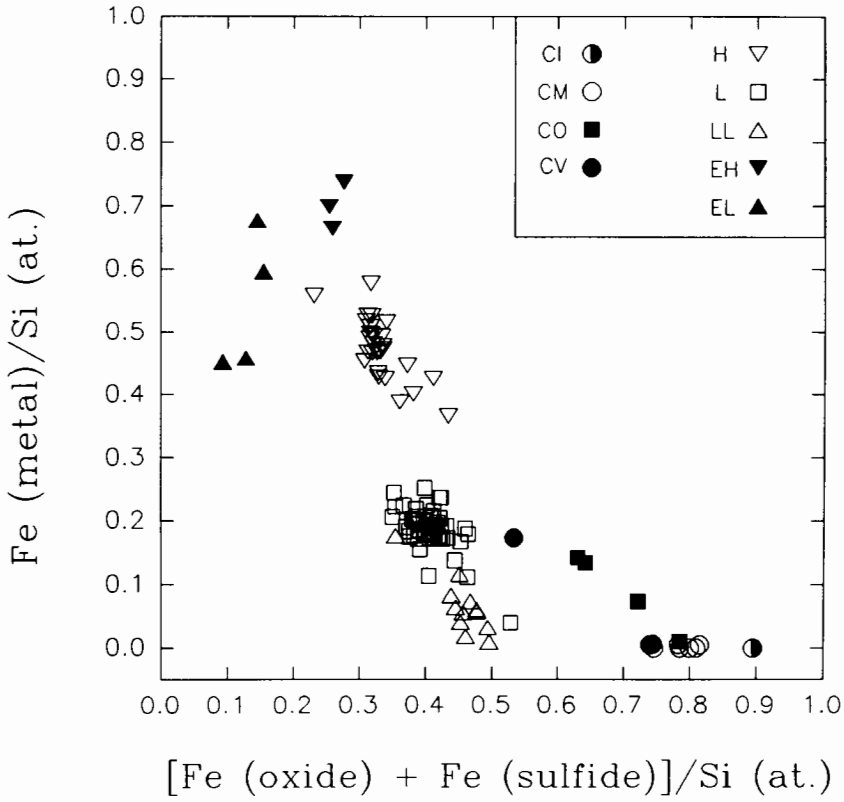


Figure 1. A modern version of the Urey-Craig diagram. The abundances of reduced iron (Fe metal) and oxidized iron (FeO plus FeS) are plotted for the E, H, L, LL and C chondrites to illustrate the reality of the compositional gaps between them. Both abundances are normalized to the abundance of elemental silicon. Note that diagonal lines with slope −1 correspond to contours of constant total iron content.

a somewhat coarser but still opaque matrix and coexisting unequilibrated mineral and glass grains the type 2 chondrites. All type 1 and 2 chondrites are carbonaceous, and all have suffered some degree of aqueous alteration. Next come type 3 chondrites with lighter matrix material, mostly made of crushed mineral grains, less than 1% carbon and less than 3% bound water, and at least small amounts of metal. The ferromagnesian minerals in them still exhibit at least 5% variation in their iron content. These are in general the most primitive meteorites, essentially unaltered by either aqueous or metamorphic processes. Type 4 chondrites show some recrystallization of feldspar and crystallization (devitrification) of glass. Less that 0.2% C and 1% water are present, and metallic grains made up of intergrowths of kamacite and taenite coexist with well-defined chondrules. The type 5 chondrites contain orthorhombic, not monoclinic, pyroxenes, coarse recrystallized feldspar, and

fully devitrified (glass-free) chondrules with slightly degraded boundaries. Type 6 chondrites are similar, but so highly recrystallized that the chondrules have become poorly defined.

A. Carbonaceous Chondrites

The carbonaceous chondrites (9% of all chondrites) are highly oxidized, containing little or no free metal, and contain volatile materials with abundances ranging from modestly larger than ordinary chondrites up to 20% bound water and 6% organic matter. In addition to these volatiles, autoreduction of magnetite and other iron oxides by carbon in these meteorites permits the extraction of up to 40% by weight of total HCNO volatiles. These fascinating meteorites are not only of great interest for the perspective they shed on the origin of life and on the origin of planetary materials far from the Sun, in the heart of the asteroid belt, but also for their outstanding resource value. They are the subject of a separate chapter by Nichols, and will not be further discussed here.

B. Enstatite Chondrites

The enstatite chondrites, representing 4 to 5% of all chondrites, are an unusual class of chondrites consisting of highly reduced phase assemblages. The silicates contain almost no FeO, while the Fe-Ni metal contains up to 3.6% dissolved Si (Keil 1968). In addition, many normally lithophile (silicate-loving) elements such as K, Na, Ca, Mg, Cr and Ti, display chalcophilic (sulfide-loving) behavior.

These meteorites are generally divided into two groups: *EH* (for high iron; containing abundant chondrules) and *EL* (low-iron, coarser-grained, with little or no evidence of chondrules). Neither group matches any of the asteroid spectral classes, although it is possible that some of the class M or P asteroids might be analogous to enstatite chondrites (Lipschutz et al. 1989).

Silicates comprise roughly 60 to 80% of an enstatite chondrite; mainly enstatite ($MgSiO_3$), Na-rich plagioclase, and one of the SiO_2 polymorphs (Keil 1968). Metallic Fe-Ni makes up roughly 20 to 25% of each type of enstatite chondrite, and contains an average of 3.2 wt% Si (EH) and 1.3% Si (EL). A less-common metallic phase, perryite $(Fe,Ni)_x(Si,P)_y$ is even higher in Si, with from 75 to 81% Ni, 12 to 15% Si, 3 to 6% Fe and 2 to 4% P (Wasson and Wai 1970). Up to 2 to 3% perryite is present in some EH chondrites (Easton 1986).

The bulk nitrogen content of the enstatite chondrites ranges from about 200 to 800 ppm, roughly 5 to 10 times higher than that in ordinary, CV and CO chondrites (Moore et al. 1969). This is due in part to the presence of two nitride minerals, osbornite TiN and sinoite Si_2N_2O. The latter is restricted to EL chondrites, which typically have a higher bulk nitrogen content than the EH chondrites.

Cr, Mn and Ti, three valuable metals usually found as minor constituents in a wide number of silicate minerals, are concentrated in a few distinct

sulfide phases in the enstatite chondrites. Three sulfide minerals, troilite FeS, daubreelite $FeCr_2S_4$ and ferroan alabandite (Fe,Mg,Mn)S contain an average of 58% (range 43–71%) of the total Ti, 82% (65–100%) of the total Cr and 75% (66–81%) of the total Mn in EL chondrites. In the EH chondrites, an average of 52% (31–66%) of the bulk Ti, 62% (42–100%) of the bulk Cr and 47% (21–73%) of the bulk Mn can be found in three sulfide minerals: troilite, daubreelite, and niningerite MgS (values were calculated using modal and mineral chemical analyses by Keil [1968] and bulk chemical analyses by Von Michaelis et al. [1969], Mason [1966], Wiik [1969], Kallemeyn and Wasson [1986], and Jarosewich and Mason [1968]). Several other Cr-bearing sulfides have been observed in EH chondrites. The modal abundances of these phases are not yet determined. Caswellsilverite $NaCr_2S_4$ is found in EH3 meteorites (El Goresy et al. 1988). Two hydrated Na-Fe-Cr sulfides, containing less than about 3% each of Na and Fe, and 10 and 20% O, have been observed in both EH3 and EH4 chondrites (Rambaldi et al. 1983; Ramdohr 1973). A third aqueous Fe-Cr sulfide, with about 20 to 21% Fe and 5% O, was recently reported by Lin et al. (1990). In addition, two poorly characterized Cr sulfides, a Na-Cr-Cu sulfide and a Na-Cu-Zn-Cr sulfide, have been observed in two EH3 chondrites (El Goresy et al. 1988). Besides being additional sources of Cr, the above phases and djerfisherite $K_3(Na,Cu)(Fe,Ni)_{12}Si_{14}$, which is restricted to EH chondrites, may provide an easily extractable source of alkali elements.

The enstatite chondrites have about 56 to 163 ppm of bulk V (Mason 1966). The mineral(s) containing this element has yet to be identified. However, the discovery of a V-rich daubreelite (0–1.2% V) in an enstatite-like clast in a carbonaceous chondrite (Kimura and El Goresy 1989), suggests that in enstatite chondrites, this element will also be concentrated in one of the sulfide phases. In general, the nonsilicate portion of the enstatite chondrite provides a valuable and potentially easily extractable source of many strategic materials.

C. Basaltic Achondrites

Most of the achondritic meteorites (Table III) are quite similar to lunar basalts in composition and mineralogy. The asteroidal and lunar basalts differ from terrestrial basalts in that they contain no ferric iron, a ubiquitous component of terrestrial rocks. As they are the products of differentiation, they have been drained of virtually all of their native iron and FeS content, and hence have severely depleted contents of siderophile and chalcophile elements. Their ferromagnesian silicates and feldspar are similar to those found in lunar basalts, reflecting an "ordinary" origin under moderately reducing conditions. Consolmagno and Drake (1977), based on both major-element and trace-element fractionation patterns in the eucrites, argue that the eucrites formed from a first melt in a material similar to L-group or C-group chondrites, but not actually the same as any known class of chondrites. McCord et al. (1970) found an excellent spectral match between basaltic achondrites and the aster-

oid Vesta, and McFadden et al. (1989) list three spectrally similar asteroids (3551 1983 RD, 3908 1980 PA and 4055 1985 DO2) among the Amor NEA population. Irrespective of their accessibility, however, their extremely low content of both volatiles and free metals make them clearly less attractive than chondritic material from a resource standpoint.

TABLE III
Achondrite Taxonomy

Achondrites
Very silicate-rich igneous objects (about 99% silicates)
Formed by density-dependant differentiation
Almost all solidified about 4.6 Gyr ago
Eucrites
Ca-rich end-member of basaltic achondrite association
Howardites
Basaltic achondrites; mixtures of eucrite and diogenite material
Diogenites
Ca-poor end-member of basaltic achondrite association
Ureilites
Ca-poor; C-rich; 1% shock-synthesized diamond
Enstatite Achondrites
Calcium-poor, dominated by $MgSiO_3$
Martian and lunar rocks
Non-asteroidal classes, including shergottites, nakhlites, chassignites

D. Lunar and Martian Meteorites

Eleven achondritic meteorites have so far been identified as lunar rocks based on their chemical, isotopic, and age similarity to lunar samples. Even more interesting are the shergottite, nakhlite and chassignite (SNC) classes of achondrite. They have crystallization ages under 1.4 Gyr, are enriched in alkalis and other volatiles, contain more oxidized iron than their asteroidal counterparts, and have weakly bound gases that mirror the chemical and isotopic composition of the atmosphere of Mars. The relevance of the SNC meteorites to our understanding of the geochemistry and petrology of Mars is discussed in the chapter by Singer et al.

E. Ureilites

Ureilites are a rather diverse group of achondrites dominated by olivine (about 90%), poor in plagioclase, with a high carbon content, including up to 1% by weight of tiny, black shock-produced diamonds. They exhibit strange patterns of abundances of the volatile elements that are simply not understood (see,

e.g., Berkley et al. 1978). Some workers have emphasized a possible genetic relationship between ureilites and shocked and differentiated CM chondrites, but any scenario for the formation of ureilites must either be very complex, or invoke unknown processes, or both. The visible and infrared spectra of ureilites are probably diverse but generally similar to carbonaceous chondrites. Ureilites have low albedos, militating against their detection.

F. Enstatite Achondrites

The enstatite achondrites, also known as aubrites, are metal-poor igneous rocks. Laboratory spectra of enstatite achondrites resemble those of E asteroids, located in the inner part of the main asteroid belt (Gaffey et al. 1989). At least one of the near-Earth asteroids (number 3103, 1982 BB) is known to be an E asteroid (McFadden et al. 1989).

On average, 90% (range 75-98%) of the volume of an enstatite achondrite consists of only one mineral: enstatite $MgSiO_3$. (All modal and compositional data given in this section are taken from Watters and Prinz [1979].) Three silicate minerals, Na-rich plagioclase, diopside $CaMgSi_2O_6$, and forsterite Mg_2SiO_4 dominate the remaining 10% of the meteorites. From 1 to 16% (average about 4%) of an enstatite achondrite consists of plagioclase, with an average composition of $ab_{93}an_4or_3$, where "ab" stands for albite $NaAlSi_3O_8$, "an" for anorthite $CaAl_2Si_2O_8$, and "or" for orthoclase $KAlSi_3O_8$. Relatively calcic plagioclase (e.g., $an_{23.8}$ in Khor Temiki) has been reported in a few of the enstatite achondrites. Forsterite comprises from 0.3 to 10% (average about 3%) and diopside makes up from 0.2 to 8.1% (average about 2%) of an enstatite achondrite.

Volumetrically small amounts (typically $<1\%$) of other phases are present in enstatite achondrites. On average, about 0.5% (from <0.1 to 7.1%) consists of Ti-rich troilite FeS, with Ti contents from 0.6 to 5.7% Fe-Ni metal (kamacite), containing from 0.12 to 4.02% Si, makes up from <0.1 to 3.7% (average about 0.2%) of an enstatite achondrite. Trace amounts of oldhamite CaS, daubreelite $FeCr_2S_4$, ferromagnesian alabandite (Fe,Mn,Mg)S and schreibersite $(Fe,Ni)_3P$ have also been observed.

From a resource standpoint, the enstatite achondrites can be considered as a good source for essentially monomineralic enstatite. This is a desirable feedstock for the production of rather pure silicon, magnesium and oxygen by (for example) a process such as carbothermal reduction.

G. Irons

Iron meteorites (Table IV) are composed principally of alloys of iron, nickel, cobalt and platinum-group metals, containing up to about 1% by weight of inclusions of troilite FeS, schreibersite $(Fe,Ni)_3P$, and several other trace minerals that are less universally distributed (Buchwald 1975). The spectral properties of meteoritic metal are well known but unfortunately not highly distinctive, featuring moderate total reflectivity (albedo), a slightly reddish slope in the visible region, and the absence of characteristic absorption bands.

A number of asteroids in the inner part of the asteroid belt, and one in near-Earth space (3554 Amun), satisfy these criteria (Gradie et al. 1989). In addition, the near-Earth asteroid 1986 DA has been shown convincingly to have the radar properties of metal (Ostro et al. 1991). However, chondritic material with a low FeO content and abundant metal would have a very similar spectrum, distinguished only by a broader variability of albedo.

TABLE IV
Iron Meteorite Taxonomy

Irons
About 99% metallic Fe-Ni-Co alloy
Inclusions of graphite, FeS, carbides, phosphides, silicates, etc.
Hexahedrites
<7% Ni; predominantly alpha-Fe (kamacite)
Octahedrites
>7% Ni; intergrowth of kamacite with gamma-Fe (taenite)
Ni-rich Ataxites
16 to 60% Ni

The metal in meteorites is made up of two minerals, kamacite and taenite. *Kamacite* is the metallurgist's α-Fe (which has body-centered cubic structure and a Ni content of less than 6%) and taenite is γ-Fe (which has face-centered cubic structure and a Ni content ranging up to about 60%). Iron meteorites that contain less than 6% Ni contain only large crystals of kamacite, and are called *hexahedrites*. In addition to the ubiquitous minerals, hexahedrites generally contain small amounts of daubreelite $FeCr_2S_4$, cohenite Fe_3C, and graphite C.

Most irons, however, contain 7 to 20% Ni, and consist of intergrowths of kamacite and taenite, which grew by exsolution from a common taenite phase during cooling from temperatures above 900°C. When these *octahedrite* irons are cut, polished, and etched with acid, the exsolution lamellae appear as striking patterns of "woven" bands, called *Widmannstaetten* patterns. A few irons with very high Ni contents (over 16 or 20%) show Widmannstaetten structure only under a microscope. They are called *ataxites*, a term that means "without structure." Most irons that are devoid of structure visible to the naked eye are indeed Ni-rich, but a few of more normal composition have been artificially heated since their arrival on Earth (both heated and hammered by blacksmiths!) and cooled so quickly that the Widmannstaetten structure is not visible. These severely altered "false" ataxites are obviously of very limited scientific value. The genuine Ni-rich ataxites commonly contain graphite, phosphates (no schreibersite), haxonite $(Fe,Ni)_{23}C_6$, and sphalerite ZnS, or sometimes only trace daubreelite. The presence of phosphates and absence of schreibersite further attest to the high degree of oxidation of the

Ni-rich ataxites.

In addition to the categorization of irons by their gross physical properties, detailed chemical analyses of hundreds of irons have revealed a complex and interesting, but still poorly understood, diversity. Sixteen major classes of chemically coherent irons have been identified to date (Wasson 1974). Interestingly, nearly 14% of the irons analyzed to date do not fit into any well-defined compositional cluster, and are termed "anomalous." The main classes of irons are shown in Fig. 2.

Some of the iron meteorite compositional groups are chemically reduced, with very low Ni contents. Among these the type IC irons are interesting: they make up 2% of all known iron meteorites, and have a Ni content of 6.1 to 6.8%. Their accessory minerals include daubreelite, cohenite, carlsbergite CrN and sphalerite. The type IIAB irons, which make up 10.8% of all iron meteorites, contain only 5.3 to 6.4% Ni and small amounts of daubreelite, cohenite and graphite. Some anomalous irons, such as the famous Tucson iron ring, contain more than 1% elemental Si dissolved in the metal.

Also of interest are the highly oxidized metal-bearing assemblages, such as the metal in higher petrologic grade carbonaceous chondrites or the Ni-rich ataxites. These irons have lost most of their metallic iron, and therefore contain much higher concentrations of the less readily oxidized metals, especially Ni, Co and the Pt-group. The IID and IVB irons contain 15 to 22.5% Ni. The IB iron San Cristobal has 25% Ni, and the compositionally similar Lime Creek, Twin Cities and Santa Catharina irons have 29.1, 30.1 and 33.6% Ni, respectively. The Dermbach iron contains 42.1% Ni, and the extraordinary Lafayette and Oktibbeha County irons have 60% Ni. The latter are as highly oxidized as the metal in CV and CO chondrites.

There are other occasional rarities among irons worthy of note. The IIE irons, notably Weekeroo Station, Colomera and Kodaikanal, contain large inclusions of potassium feldspar $KAlSi_3O_8$ with crystal sizes up to several centimeters, making them an obvious source of meteoritic alkali metals. The Butler anomalous iron (15.2% Ni) contains an astonishing 1970 ppm of germanium and 87 ppm of gallium, and the Reed City anomalous iron (7.35% Ni) contains 54 ppm of iridium. The IVB irons (17% Ni) contain 13 to 36 ppm of Ir, and the IIAs (5.5% Ni) range from 3 to 59 ppm Ir. Within many groups there is a clear anticorrelation of Ir with Ni, reflecting a strong partitioning of Ir into kamacite during magmatic processes (Wasson 1974).

The metal phase in chondrites, which has not undergone partial melting or fractional crystallization, preserves a positive correlation of the Pt metals with Ni. The Ni concentration in H-group chondrite metal ranges from 8 to 12%. The L chondrite metal contain 12 to 18% Ni, and the LL chondrites metal, 20 to 30% Ni. The ratios Co:Ni, Pt:Ni, Ir:Ni, etc. are all constant and equal to the corresponding cosmic abundance ratios. The composition of the metal phase in chondrites is compared to the composition groups of iron meteorites in Fig. 2.

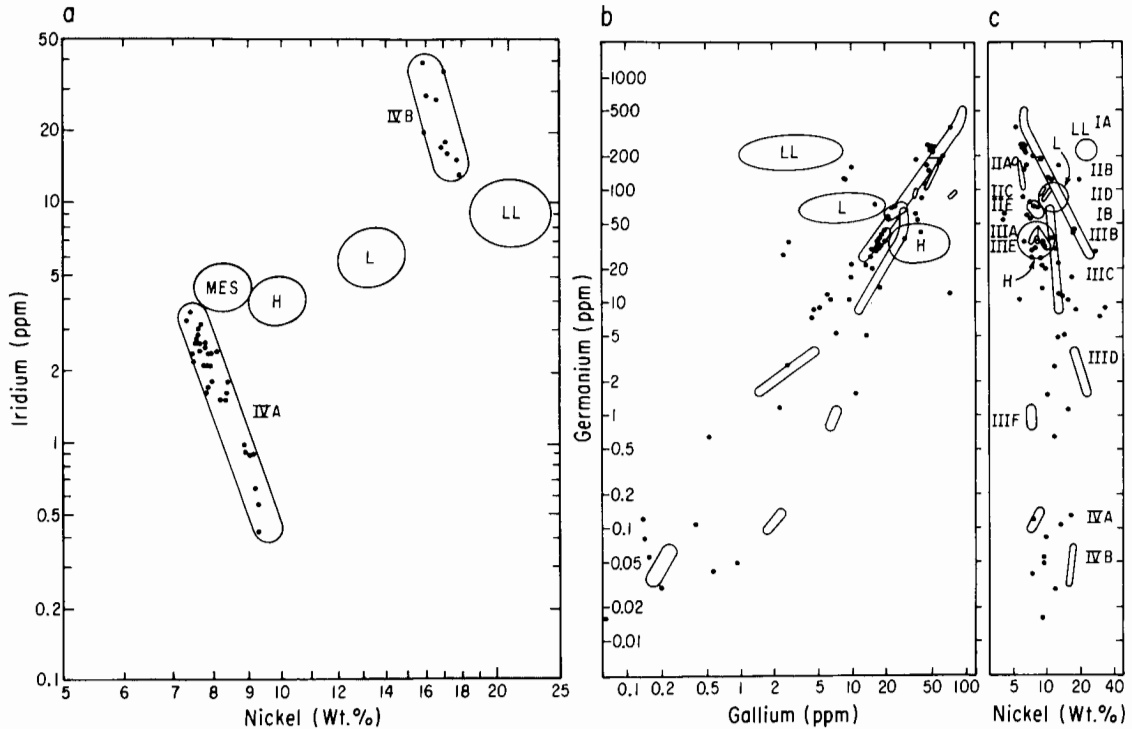


Figure 2. Compositional groups of meteoritic metal. (a) is a plot of Ir vs Ni; (b) shows Ge vs Ga; and (c) shows Ge vs Ni. The data on iron meteorite classes are from Wasson (1974) and the data on the metal phase from chondritic meteorites are from Rambaldi (1977*a*, *b*, *c*). The differences between the chondrite groups reflect nebular condensation processes. The differences between the groups of irons are due not only to oxidation state differences, but also to partitioning between kamacite, taenite, and other phases during magmatic processing. MES indicates the metal in mesosiderites.

H. Stony-Irons

The stony-irons, as the name implies, contain roughly equal amounts of native metal and silicates (Table V). The large majority of the stony-irons belong to two distinct classes. First are the *pallasites* (about 35 known), which are essentially made of olivine crystals embedded in a continuous metal matrix very similar to the IIIAB irons. The olivine and metal make up some 95% of the total mass, with about 2% of FeS, 1% of schreibersite, 0.4% chromite, and traces of pentlandite $(Fe,Ni)_9S_8$ and the phosphates whitlockite, stanfieldite $Ca_4(Mg,Fe)_5(PO_4)_6$ and farringtonite $Mg_3(PO_4)_2$ (Buseck 1977).

The rest of the stony-irons are *mesosiderites* (about 20 known), which resemble basaltic achondrites except for their much larger metal content and have abundant metal dispersed among silicate grains. The stony-irons present the resource opportunities of basaltic achondrites and irons in the same body, with no significant new features added. The difficulty of cutting and crushing pallasites places them more nearly in the category of irons, so far as the technical requirements for using them are concerned (Powell 1969).

TABLE V
Stony-Iron Meteorite Taxonomy

Stony-Irons
About 50% ferrous metal alloys, 50% silicates
Mesosiderites
Related to eucrite/howardite/diogenite family of achondrites with added metal
Pallasites
Related to IIIAB irons

III. RESOURCE UTILIZATION OPPORTUNITIES

A. Metals

Among the noncarbonaceous meteorite classes, the most attractive resource is native ferrous metal alloy. Iron and nickel are widely distributed in the terrestrial meteorite collection, being abundant in all ordinary and enstatite chondrites, all stony-irons, and all irons. These make up 91% of the total number of meteorite falls on Earth. Aside from a small bias toward higher Ni and Co concentrations (for use in corrosion-resistant and high-temperature alloys) and higher metal content, the most important consideration weighting our interest in different metal sources is the ease of mining and extracting the metal. The latter suggests that mining metal from a 10-km M-type asteroidal core might be extremely difficult compared to the option of extracting it magnetically from a chondritic asteroidal regolith. Meteoritic iron-nickel is extremely tough at temperatures characteristic of near-Earth space. At very

low temperatures, far below the daytime temperatures in near-Earth space or in the inner asteroid belt, where the M asteroids are found, kamacite metal undergoes a ductile-to-brittle transition that makes it easy to crush.

The other solutes and inclusions in meteorite metal include native elemental Si from enstatite achondrite metal, platinum-group and precious metals from all native metal sources, gallium, germanium, arsenic, etc., and the elements common in inclusions in native metal (especially S, C, P, N, Cr and Zn).

Direct melting and casting of native ferrous metals has the advantage of simplicity and familiarity, and requires no additional reagents. Unfortunately, this approach requires high process temperatures and also precludes separation of the solute and inclusion elements from iron, and may lead to alloys with undesirable physical and chemical properties, weakened by sulfur and vulnerable to corrosion. More complex, but offering direct recovery of unaltered inclusions and relatively simple routes to elemental separation of the solutes and low process temperatures, is gaseous carbonyl extraction of the metals. The only reagent required is gaseous CO, which is readily produced by heating almost every known type of meteoritic material. Precisely controlled mixtures of metals (especially Fe, Ni and Co) can be compounded from carbonyl process products for use in casting metal alloys. After these three elements and carbon, the terrestrial steelmaker looks next to Cr, Mn, V and Mo as alloying elements. These are treated in the section on sulfides below.

The residue from carbonyl extraction of native ferrous metal alloys is very rich in cobalt and platinum-group metals. The cobalt may in turn be separated from the platinum-group metals by very high-pressure extraction with CO, by extraction with $CO\text{-}H_2O$ mixtures as the carbonyl hydride, or by wet chemical techniques. The metal fraction of a typical LL chondrite contains 50 to 60 ppm of platinum-group metals, and the concentration in the metal grains in CV and CO chondrites should reach 100 to 200 ppm. In general, metals that are the result of partial melting or of fractional crystallization are not only enormously variable in composition, but also very rarely contain large amounts of Ir (the only Pt-group measured in most irons). Thus recovery of Pt for use as a catalyst, corrosion-resistant coating, or electrode material appears far easier from a chondritic metal feedstock.

B. Silicates and Oxides

The first obvious use of silicates is the electrolysis of molten ferromagnesian silicates to produce oxygen and metallic iron. The second target of choice would be reduction of high-purity enstatite from enstatite achondrites, which contain cm-sized $MgSiO_3$ crystals. The enstatite achondrites probably correspond to the E asteroids.

Magnetite is found only in the CI and CM chondrites, and ilmenite is a ubiquitous minor mineral in ordinary chondrites and basaltic achondrites. Autoreduction of the natural carbon-magnetite mixtures in CI and CM chondrites

will provide a CO-CO_2 mixture and metallic iron.

For nonferrous metals, Al and Ti are most attractive, but both must usually (except in the enstatite meteorites) be extracted from rare, extremely stable oxide minerals. Ilmenite, the only tractable titanium oxide mineral, would be best sought in the most coarsely crystalline basaltic achondrites, from which crushed material beneficiation of ilmenite seems a plausible possibility. Aluminum extraction from plagioclase is possible, but a clean mineral separate of feldspar seems feasible only from the massive K-spar crystals in the rare type IIIE irons.

C. Sulfides and Others

Several industrial uses of sulfur and its compounds, especially sulfuric acid, come to mind. Sulfides are, however, of interest mainly for their metal content. Virtually all meteorites (except achondrites and a few CI chondrites) contain abundant sulfides, usually FeS. Weak-acid leaching of sulfides can in general liberate hydrogen sulfide and an aqueous solution of ferrous iron salts. The best meteoritic sources of Mn, Cr and Ti, and possibly V, Zr, Mo, are the enstatite chondrites. Acid leaching of their sulfides also liberates potassium, sodium, and possibly lithium.

Nitrides are of great interest as sources of nitrogen gas. The minerals sinoite Si_2N_2O and osbornite TiN are found only in the enstatite meteorites, and CrN has been reported only in a few irons of types IC, IIIA and IIIE, for which crushing and leaching is a very unattractive option. The highest inorganic N concentration found in the EL chondrites, about 0.8%, is the same as the highest reported total nitrogen content in CM chondrites, and 60% of that found in the extremely volatile-rich CI chondrites. Interestingly, the isotopic compositions of these very nitrogen-rich meteorites are quite different, with the CI and CM chondrites clustered about a $\delta^{15}N$ value of +3.0%, whereas the EL chondrites are near −4.0%. The oxygen isotopic signatures of E and C chondrites also differ: thus any simple genetic relationship between the E and C chondrites, and hence the likelihood of deriving them from the same parent material (or finding them on the same parent body) seems extremely remote.

Phosphides and carbides are widely distributed in irons, although rarely making up more than a fraction of a percent of the total meteorite mass. The carbon concentration of the metal phase is typically close to 0.01%. Most of the carbon exsolved from the metal phase is found as distinct inclusions of graphite, cohenite and cliftonite. High total carbon concentrations are found only in ureilites and C chondrites. Acid leach of carbides commonly liberates kerosene-like liquid hydrocarbons, but the yield is unfortunately small. The last carbon resource of note is the industrial-quality diamond found in ureilites, a resource that would be required in such small quantities as to make its exploitation unrewarding. The problem of remote spectral identification of ureilite parent bodies was mentioned above.

IV. SCIENCE AND TECHNOLOGY NEEDS

In order to build an adequate scientific and technical foundation for the economic evaluation of asteroidal resources, several types of basic research are clearly required. First, there is a serious need for further acceleration of the search for near-Earth asteroids. The highly successful photographic search program of E. Helin and E. and C. Shoemaker, which has concentrated on wide-field coverage of bright objects, has achieved a discovery rate of about one new near-Earth asteroid per month. Gehrels, Rabinowitz, and Scotti have recently brought the Spacewatch program to operational status and, with the assistance of SERC funding, are now discovering 2 to 3 NEAs per month. Spacewatch has a much more limited sky coverage, but is much more sensitive to faint objects. It has already detected the smallest natural object known in heliocentric orbit, and also the one that passed closest to Earth, the Apollo asteroid 1991 BA. With a further expenditure of less than $80,000 it would be possible to acquire a new high-sensitivity CCD detector array that would expand the number of asteroids several-fold. Extension of the Spacewatch program to another site in India is already under way, and the addition of a third site in the Far East would permit nearly constant coverage of the night sky.

There is already a problem getting prompt spectral characterization of nearby asteroids. From the instant of discovery of an NEA to the end of the dark time in which it was discovered is usually a matter of only a few days. During the next dark of the Moon, the newly discovered asteroid is usually so distant and faint that its spectrum cannot be measured. Because of the intense competition for telescope time, it is usually impossible for an astronomer to find an available telescope of adequate size, transport himself and his equipment to it, and carry out the observation before the asteroid has faded from view. Usually the best that can be done is to plan ahead to the next apparition of the asteroid, two to four years in the future. More often than not, the next apparition is geometrically less favorable than the one in which the asteroid was discovered. Thus a dedicated telescope facility with appropriate permanently installed photometric instrumentation is highly desirable. As a measure of the importance of this problem, it is necessary only to note that a mere 20% of the NEAs discovered to date have any available spectral data to tell us about their compositions.

In support of the interpretation of these asteroid spectra, both laboratory study of simulants and natural meteoritic samples and theoretical efforts directed toward the interpretation of these spectra must be encouraged. Such a program requires only a few million dollars, cheap compared to the cost of even a single spacecraft instrument. But, in the longer term, advanced spacecraft missions cannot be effectively planned and executed without the kinds of data provided by these Earth-based studies. Only the first and simplest form of mission, examination of the surface of an asteroid at close range by a spacecraft flyby, can be planned and executed with presently available

data; indeed, a high-speed flyby of the asteroid Gaspra was achieved at on 29 October 1991 by the Galileo spacecraft. Analyses of the data returned after this encounter, hindered as it was by the disabled high-gain antenna, may permit planning of more advanced types of missions. Rendezvous, orbit, and landing missions are not much more difficult from the point of view of propulsion requirements than simple flybys. Indeed, missions to land on the Moon have more severe energy (delta V) requirements than missions to the most accessible 20% of the NEAs, higher required thrust and acceleration level, and far more stringent engine timing requirements during takeoff and landing. Sample return from most NEAs requires less propulsive energy than from the Moon, often less than 1 km s^{-1}, and sometimes as small as 60 m s^{-1} (compared to nearly 3000 m s^{-1} for return from the Moon). Thus many rendezvous and landing opportunities can very easily be extended to permit return of samples of asteroidal material to Earth.

Finally, the rich variety of meteorite materials available in our terrestrial laboratories today makes it possible to begin processing experimentation at once. Schemes oriented toward extraction of volatiles (H, C, N and O), metals (Fe, Ni, Co, Mn, Cr, Ti, etc.), reagents (O, S, HCl, Na, etc.) and refractories (Ca, Ti, Al, etc. oxides) from them can be tested readily in our laboratories on Earth, in most cases without serious limitations caused by rarity of materials.

Acknowledgments. The preparation of this chapter was supported by NASA funding provided to the University of Arizona/NASA Space Engineering Research for Utilization of Local Planetary Resources. The authors are indebted to M. J. Gaffey, W. van R. Boynton and M. Nelson for helpful comments.

REFERENCES

Berkley, J. L., Taylor, G. J., and Keil, K. 1978. Ureilites: Origin as related magmatic cumulates. *Lunar Planet. Sci.* IX:73–75 (abstract).

Buchwald, V. F. 1975. *Handbook of Iron Meteorites* (Berkeley: Univ. of California Press.)

Buseck, P. R. 1977. Pallasite meteorites—mineralogy, petrology, and geochemistry. *Geochim. Cosmochim. Acta* 41:711–740.

Consolmagno, G. J., and Drake, M. J. 1977. Composition and evolution of the eucrite parent body: Evidence from rare earth elements. *Geochim. Cosmochim. Acta* 41:1271–1282.

Dodd, R. T. 1981. *Meteorites: A Petrologic-Chemical Synthesis* (New York: Cambridge Univ. Press.)

Easton, A. J. 1986. Studies of kamacite, perryite and schreibersite in E-chondrites and aubrites. *Meteoritics* 21:79–93

El Goresy, A., Yabuki, H., Ehlers, K., Woolum, D., and Pernicka, E. 1988. Qingzhen and Yamato-691: A tentative alphabet for the EH chondrites. *Proc. NIPR Symp. on Antarctic Meteorites*, vol. 1, pp. 65–101.

Gaffey, M. J., Bell, J. F., and Cruikshank, D. P. 1989. Reflectance spectroscopy and asteroid surface mineralogy. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 98–127.

Gradie, J. C., Chapman, C. R., and Tedesco, E. F. 1989. Distribution of taxonomic classes and the compositional structure of the asteroid belt. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 316–335.

Jarosewich, E., and Mason, B. 1968. Chemical analyses with notes on one mesosiderite and seven chondrites. *Geochim. Cosmochim. Acta* 23:411–416.

Kallemeyn, G. W., and Wasson, J. T. 1986. Compositions of enstatite (EH3, EH4,5 and EL6) chondrites: Implications regarding their formation. *Geochim. Cosmochim. Acta* 50:2153–2164.

Keil, K. 1968. Mineralogical and chemical relationships among enstatite chondrites, *J. Geophys. Res.* 73:6945–6976.

Kerridge, J. F., and Matthews, M. S., eds. 1988. *Meteorites and the Early Solar System* (Tucson: Univ. of Arizona Press).

Kimura, M., and El Goresy, A. 1989. Discovery of E-chondrite assemblages, SiC, and silica-bearing objects in ALH85085: Link between E- and C-chondrites. *Meteoritics* 24:286 (abstract).

Lin, Y. T., Kimura, M., and El Goresy, A. 1990. Discovery of a new aqueous Fe-Cr sulfide in some enstatite chondrites. *Meteoritics* 25:379 (abstract).

Lipschutz, M. E., Gaffey, M. J., and Pellas, P. 1989. Meteoritic parent bodies: Nature, number, size and relation to present day asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 740–777.

Mason, B. 1966. The enstatite chondrites. *Geochim. Cosmochim. Acta* 30:23–39.

McCord, T. B., Adams, J. B., and Johnson, T. V. 1970. Asteroid Vesta: Spectral reflectivity and compositional implications. *Science* 168:1445–1447.

McFadden, L. A., Tholen, D. J. and Veeder, G. J. 1989. Physical properties of Aten, Apollo and Amor asteroids, In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 442–467.

Moore, C. B., Gibson, E. K. and Keil, K. 1969. Nitrogen abundances in enstatite chondrites. *Earth Planet. Sci. Lett.* 6:457–460.

Ostro, S. J., Campbell, D. B., Chandler, J. F., Hine, A. A., Hudson, R. S., Rosema, K. D., and Shapiro, I. I. 1991. Asteroid 1986 DA: Radar evidence for a metallic composition. *Science* 252:1399–1404.

Powell, B. N. 1969. Petrology and chemistry of mesosiderites. I. Textures and composition of nickel-iron. *Geochim. Cosmochim. Acta* 33:789–810.

Rambaldi, E. R. 1977*a*. Trace element content of metals from L-group chondrites. *Earth Planet. Sci. Lett.* 31:224–238.

Rambaldi, E. R. 1977*b*. The content of Sb, Ge, and refractory siderophile elements in metals of L-group chondrites. *Earth Planet. Sci. Lett.* 33:407–419.

Rambaldi, E. R. 1977*c*. Trace element content of metals from H- and LL-group chondrites. *Earth Planet. Sci. Lett.* 36:347–358.

Rambaldi, E. R., Housley, R. M., Rajan, R. S., Cirlin, E., El Goresy, A., and Wang, D. 1983. Unusual mineral assemblages and textures in Qingzhen enstatite chondrite. *Meteoritics* 18:380–381.

Ramdohr, P. 1973. *The Opaque Minerals in Stony Meteorites* (Berlin: Akademie-Verlag).

Urey, H. C., and Craig, H. 1953. The composition of stone meteorites and the origin

of the meteorites. *Geochim. Cosmochim. Acta* 4:36–82.

Van Schmus, W. R., and Wood, J. A. 1967. A chemical-petrologic classification for the chondritic meteorites. *Geochim. Cosmochim. Acta* 31:747–765.

Von Michaelis, H., Ahrens, L. H., and Willis, J. P. 1969. The composition of stony meteorites II. The analytical data and an assessment of their quality. *Earth Planet. Sci. Lett.* 5:387–394.

Wasson, J. T. 1974. *Meteorites: Classification and Properties* (Berlin: Springer-Verlag).

Wasson, J. T. and Wai, C. M. 1970. Composition of the metal, schreibersite and perryite of enstatite achondrites and the origin of enstatite chondrites and achondrites. *Geochim. Cosmochim. Acta* 34:169–184.

Watters, T. R., and Prinz, M. 1979. Aubrites: Their origin and relationship to enstatite chondrites. *Proc. Lunar Planet. Sci. Conf.* 10:1073–1093.

Wiik, H. G. 1969. On the regular discontinuities in the composition of meteorites. *Commentationes PhysicoMathematicae* 34:135–145.

VOLATILE PRODUCTS FROM CARBONACEOUS ASTEROIDS

CHARLES R. NICHOLS
Bose Corporation

Samples from the Moon and most near-Earth asteroids are depleted in the volatile elements carbon and hydrogen. Only one class of asteroids is rich in these elements: carbonaceous asteroids. No carbonaceous asteroid has yet been sampled directly. Our knowledge of their composition is based on the study of carbonaceous meteorites. A wide variety of evidence suggests that near-Earth carbonaceous asteroids are the source of most carbonaceous meteorites. As about one third of the classified near-Earth asteroids appear to be carbonaceous, we can expect to find an abundance of volatile resources among the near-Earth asteroids. For space missions in the past, it has been necessary to bring all supplies from Earth. For the manned lunar base and Mars exploration missions of the future, we have the option of "living off the land." For example, space flight would be much less expensive if propellant were available in orbital fuel depots. Spacecraft could then be launched with empty tanks, dramatically reducing the payload mass and launch cost. Carbonaceous ore could be processed into propellant identical to that used by the space shuttle's main engines. When heated, the ore releases large quantities of water, carbon monoxide, and carbon dioxide. A fuel production plant could extract and process the water to provide a steady supply of hydrogen and oxygen propellant. Once in operation, the fuel plant becomes part of the growing infrastructure of the space economy. If we invest wisely in this infrastructure, it will yield tremendous economic benefits for decades to come. This chapter begins by discussing the economics of asteroid mining. Some simple volatile products are then identified. Production processes are suggested, with an emphasis on existing industrial practice. Finally, the mining environment is detailed, using the moons of Mars as well-studied examples.

I. INTRODUCTION

Mankind's investment in space has grown steadily over the past thirty years. Space industrialization will become more desirable, and eventually an economic necessity, as this trend continues. These future space industries will need to find an economical source of materials. Carbonaceous asteroids will probably be the least expensive source of the volatile compounds needed for propellant and life support.

The utilization of native materials in space will require a large investment for the equipment which mines, processes, and transports them. Aerospace and government planners must decide when to begin investing in space-based

infrastructure. To do so, they will need information on what is known, what is possible, and what might be affordable.

A. The Market for Asteroidal Products

The space resources literature continues to mature in its recognition of the importance of economic perspective. Cutler and Hughes (1985) identified low initial capital investment and quick payback as the prime requirements for economic feasibility of space mining. They identified propellants and metal plate, delivered to low Earth orbit (LEO) as the best products. Cutler also identified the design and development of mining processes as the biggest time bind (Gertsch et al. 1985). Cordell and Steinbronn (1988) found space resources pivotal for large-scale human operations in space, concluded that volatiles should be mined even if lunar oxygen were already in production, and agreed that keeping initial costs down was vital.

The cost of placing a pound of goods into LEO currently exceeds $3000. Although designs for cheap launch services abound, they have not as yet been subjected to the rigors of the market. However, the economic viability of space resource utilization would probably survive the development of low cost launch technology. Leonard et al. (1987) concluded that a propellant mine at Phobos would be profitable as long as Earth-to-LEO launch cost exceeds $250 to $500 per pound.

Space-based industry can provide goods to LEO at a much lower cost than Earth-to-LEO under certain conditions. The cost of developing a new supply must be justified by a large demand (by weight) for the product. The cost of providing the product from space must be well below the cost of delivering it from Earth. Finally, the product must be easy to make from materials already present in space.

Space planners have attempted to predict the future needs of the space community, and have identified some volatile products which look promising. Volatiles in large quantities will be needed for life support for a permanently manned station. Propellants could be used for moving satellites, spacecraft, humans, and raw materials. Volatile reagents can be transported to other space factories and used to reduce the cost of their products, such as lunar oxygen and steel.

Volatile products for life support include water, oxygen, and nitrogen for humans, and carbon dioxide and fixed nitrogen for agriculture. Their availability allows a lifestyle in space characterized by abundance and self-sufficiency, as contrasted with the canned air, sponge bath, and reconstituted food lifestyle of the Apollo missions.

Cheap locally produced propellants could revolutionize the economics of space operations. Many space-derived propellant systems have been suggested. Medium-velocity systems include hydrogen, methane, kerosene, or methanol as fuel with oxygen as the oxidizer. A high-velocity option is superheated hydrogen monopropellant. The choice of medium- or high-velocity propellant depends on the mission's total velocity change (ΔV) before refuel-

ing. For missions with ΔV under 5 km s^{-1} (typical of missions to near-Earth destinations), hydrogen/oxygen is the preferred propellant option. Its advantages include a mature technology base and wide availability of water as feed stock. For cometary resource missions, super-heated hydrogen is the recommended propellant option. To obtain a useful fuel to payload ratio on these high ΔV missions, a very high exhaust velocity is required.

Volatile products which are potentially useful in metal processing are carbon monoxide, hydrogen and methane. Carbon monoxide is essential to the carbonyl process for low-temperature purification and deposition of iron and nickel. Hydrogen and carbon monoxide have been proposed as reducing agents for lunar ilmenite reduction. Methane has been suggested for the carbothermal reduction of lunar maria soils (Rosenberg et al. 1965).

B. Sources for Volatiles

While carbonaceous asteroids are a high-grade source of volatiles, other nonterrestrial sources must also be considered. The Moon and many noncarbonaceous asteroids have ample oxygen and sulfur. Mars has abundant oxygen, sulfur, carbon, nitrogen, hydrogen and argon (Meyer and McKay 1989). These resources are discussed in detail elsewhere in this book.

Carbonaceous asteroids can be categorized by their orbital location in the solar system: main-belt asteroids, near-Earth asteroids (NEAs), short-period comets, and the Martian moons, Phobos and Deimos. Inferred composition is strongly correlated to the type of orbit. Composition is inferred based on reflectance spectrum, using a variety of taxonomic schemes (Tholen and Barucci 1989; Zellner 1979). This chapter uses "carbonaceous" in its most general sense as carbon-bearing. Current asteroid classification would include types P, D, RD, T, F, G and B as subclasses of type C, and thus probably rich in carbon and volatiles.

Main-belt asteroids include bodies up to 1000 km in diameter and their collisional fragments. The large (>100 km, $N = 195$) main-belt asteroids are 75% type C (Zellner 1979). The usual interpretation is that these type C bodies are carbonaceous. The composition of the main belt varies systematically from the inner to the outer edge, with type S dominating the inner belt and type C dominating the outer belt (Gradie and Tedesco 1982). Within the type C population, we can expect the composition to vary from thermally altered C3/C4 compositions in the inner belt to relatively pristine C1/C2 types in the outer belt.

The NEAs are small bodies (<10 km) with a range of compositions spanning all common asteroid types. They are derived from a mixture of main-belt collisional fragments and burned-out short-period comets (Shoemaker et al. 1979). The total population of NEAs over 100 m diameter is estimated to be about 100,000 objects, with 150 or so currently known. A semi-automated telescope system at the University of Arizona called Spacewatch has considerably increased the discovery rate recently. The utility of NEAs as resources is heavily dependent on the economics of returning massive cargoes

from the candidate body. Dominating variables are fuel costs, launch window frequency, and round-trip time. If atmospheric braking can be used, many bodies can return cargo to Earth with extremely low fuel costs.

Short-period comets are small (<10 km) bodies which produce a characteristic tail of gas and dust when they are near the Sun. They (and the NEAs derived from them) may be composed primarily of volatiles. Short-period comet material is unlikely to be represented in meteorite collections due to its fragile structure. However, the volatile components are expected to include H_2O, CO_2, CO, NH_3, CH_4 and a wide variety of organic compounds. A comet would be a first-class candidate for volatile resources. Unfortunately, the economics of returning cargo from a comet to Earth are nearly prohibitive. The very high return ΔV produces an unacceptably low payload fraction, and is beyond the state of the art of atmospheric braking technology. For a few well-placed short-period comets, a combination of cheap propellant and atmospheric braking would allow economical recovery of high-value volatiles.

The Martian moons are black, very low density, small ($\approx$10 km), heavily cratered, and covered with dust. They have been well observed by four spacecraft with many types of instruments. Some question remains about whether they are actually carbonaceous. Although they resemble carbonaceous meteorites in density and color, they seem to contain little or no water of hydration at the surface (see Sec. VI.B). In any case, their proximity to Mars and excellent accessibility make them prime resource candidates. For the purposes of this review, they are considered to be carbonaceous asteroids.

C. Target Selection

On Earth, mining sites are chosen based on simultaneous minimization of cost and risk. The same should be true in space. Ultimately, all estimates of cost and risk depend on having accurate knowledge of the site, and an accurate model of the operations to be located there.

The most economical way to estimate the cost and risk of mining on proposed ore bodies is to gather data through an extensive Earth-based observation program. Many bulk properties of asteroids can be determined by telescopic measurements. Among them are size, shape, mineral composition, orbit, rotation rate, orientation of the rotational axis, presence of a pulverized surface deposit, reflective albedo, and limited surface detail. From this data can be derived information about mass, density, surface gravity, probable composition, transportation costs, surface environment, and expected yield of volatile products.

To date, very few asteroids have been well characterized according to these criteria. The NEAs, which are among the most accessible asteroids, are also among the most poorly characterized. With NEAs, time is of the essence, because they can only be observed during rare close approaches to Earth. The solution is to dedicate a telescope with the necessary instruments to the task of immediately following up new discoveries. A large number of accessible asteroids should be characterized. The best candidates could then be selected

and exhaustively studied.

D. Prospecting for Resources

When the target characterization and selection process reaches a consensus, it is time to send a prospecting mission. Prospecting at such remote bodies is so expensive that it will probably be confined to a single mission. The prospecting craft should include an orbiter, a lander, and penetrators. The orbiter allows detailed surface mapping and the selection of candidate sites. The lander allows tests of various means of ore acquisition and processing, plus evaluation of ore composition. A cluster of small penetrator probes allows subsurface sampling, plus soil depth and rock strength experiments. With this data in hand, a go/no-go decision can be made for the body, a site can be selected, and equipment design can begin.

II. DESIRED VOLATILE PRODUCTS

In this section, "volatile products" refers to simple compounds of hydrogen, oxygen, carbon, sulfur and nitrogen, which could be produced from carbonaceous ore (see Table I). "Desired" indicates that, in the author's judgement, they are likely to be useful in large quantities for the support of some ongoing space project.

TABLE I

Properties of Volatile End Products and Byproducts[a]

Molecule	mp, °C	bp, °C	Liq g/cc	Gas g/L	Primary Use
H_2	−259	−253	0.0700^{bp}	0.0899^{0}	Fuel
N_2	−210	−196	0.808^{bp}	1.251^{0}	Air
CO	−199	−192	0.814^{bp}	1.250^{0}	Metallurgy
O_2	−218	−183	1.149^{bp}	1.429^{0}	Propellant, Air
CH_4	−182	−164	0.466^{bp}	0.717^{0}	Fuel
CO_2	-57^{5atm}	-78^{subl}	1.101^{-37}	1.977^{0}	Agriculture
H_2S	−85	−60	0.993^{bp}	1.539^{0}	Metallurgy
NH_3	−78	−33	0.682^{bp}	0.771^{0}	Agriculture
SO_2	−73	−10	1.50^{-20}	2.92^{0}	Refrigerant
$Ni(CO)_4$	−25	+43	1.32^{17}		Metallurgy
SO_3	(+17)	+45	1.97^{20}		To make H_2SO_4
CH_3OH	−94	+65	0.791^{20}		Fuel
NH_4OH	−77	(+100)	0.90^{25}		Agriculture
H_2O	+0	+100	1.000^{mp}		Life support
$Fe(CO)_5$	−21	+103	1.457^{21}		Metallurgy
H_2O_2	−0	+150	1.407^{25}		Oxidizer
H_2SO_4	+10	+290	1.841		Metallurgy

[a] Data from Weast 1983; Windholz 1983. Superscript gives temperature at which the measurement was made. NH_4OH is 29% NH_3 (by weight) in water.

A. Water

Water seems ideal as a first product for industry in space. It is produced by low temperature pyrolysis of C1 or C2 ore, along with carbon dioxide, from which it is easily separated. Water is useful in large quantities for life support. On manned missions, it is an efficient shield against solar flares and cosmic rays. It can be converted (by electrolysis or thermolysis) into hydrogen and oxygen, which are also among the most useful of products. It is also a useful solvent and reagent in many industrial processes. Water can be easily transported as a liquid or as a solid. Large solid blocks could remain exposed to vacuum for years with minor loss, thus avoiding the cost of containers.

B. Hydrogen

Hydrogen's potential usefulness has been stifled on Earth, due to the danger of ignition in an oxygen-rich atmosphere. In space it is likely to find many new uses. Its extremely low molecular weight makes it ideal for a number of applications. For example, it has the highest mean velocity of any molecule at a given temperature, making it the ideal monopropellant for thermal rockets. It also has by far the highest specific heat of any gas, making it a good candidate for a heat transfer medium. Hydrogen is also a useful reagent, especially for lunar production of oxygen and metals (Ness et al. 1991). The hydrogen in such reactions is converted to water, which can be efficiently recovered and recycled.

Hydrogen is as unwieldy as it is useful. Gaseous hydrogen is uneconomical to store, because the container far outweighs the gas. Liquid hydrogen is somewhat cheaper to store, but requires continuous refrigeration to prevent boil-off, even in the space environment. Hydrogen is also expensive to liquefy, due to its extremely low boiling point, its very low inversion temperature (above which Joule-Thompson expansion produces a rise in temperature, instead of lowering it), and the need to convert ortho-hydrogen to para-hydrogen (Neary 1983). Hydrogen should therefore be produced at the site of use whenever possible, to avoid storage.

When it must be transported, hydrogen should first be converted to a carrier compound such as water, methane, or ammonia. The carrier compound should be easy to produce, easy to recover hydrogen from, have a high boiling point, have a high weight of hydrogen per volume, and contain only useful elements available in space. For comparison, hydrogen storage density in each liquid at its boiling point is: $H_2 \approx 0.070$, $H_2O \approx 0.112$, $CH_4 \approx 0.117$, and $NH_3 \approx 0.121$ g cm^{-3}.

Water is the perfect hydrogen carrier for hydrogen/oxygen propellant. The ratio of hydrogen to oxygen in water is nearly ideal for propellant use. The liquefaction of hydrogen could be avoided entirely by designing rockets which electrolyze water, and immediately burn the gases. The tiny rocket engine would produce a steady low thrust, with only water and electricity as inputs.

Methane is a more appropriate carrier than water if the end-use site is

the Moon, because the Moon already has plenty of oxygen, but is deficient in carbon. During low-thrust spaceflight, methane can be decomposed to hydrogen, which is burned immediately, and graphite, which is stored (see Sec. III.F). When a high-thrust burn is required, the methane can be burned directly.

Ammonia is the best hydrogen carrier in terms of hydrogen storage density. Its use is hampered by the low abundance of nitrogen in typical carbonaceous asteroids. Ammonia may occur in abundance in comet nuclei and burned-out comets, however.

C. Oxygen

Oxygen is useful both for propulsion and for life support. As an oxidizer for propulsion, there is no good alternative. The other oxidizers used in rocketry (chlorates, nitric acid, nitro compounds and nitrogen tetroxide) have lower performance and contain elements that are scarce in carbonaceous meteorites. Hydrogen peroxide is sometimes used as an oxidizer, but also has low performance, and is hard to make in space. Water seems an ideal noncryogenic carrier for oxygen. It is simple to make as well as to decompose, is 90% oxygen by weight, and requires minimal tankage.

D. Carbon

Carbon is useful as a reagent and as a refractory material. Its use in steelmaking is well established on Earth. It is likewise used in refining high-energy elements such as silicon, aluminum, magnesium and calcium. Its use as a reducing agent in space will require improvements in carbon dioxide recycling technology, however. Carbon recycling has not been developed on Earth due to the very low cost of industrial carbon. A means of reducing CO_2 to CO using stabilized zirconia at 900°C has been demonstrated in the laboratory (Erstfeld et al. 1979). On the Moon, carbon can be converted to carbon monoxide using locally produced oxygen. Carbon monoxide has been suggested as a reducing agent for oxygen production from lunar ilmenite (Zhao and Shadman 1991). It could also be used to win the 0.2 to 0.5% of metallic iron in lunar regolith (Agosto 1981) via the carbonyl process.

E. Carbon Monoxide

Carbon monoxide is a possible export from a carbonaceous to a non-carbonaceous asteroid. The choice of whether to export carbon or carbon monoxide is a matter of economics. Carbon monoxide is 57% dead weight if oxygen is available at the destination. If the transportation cost is small, or the conversion of carbon to the monoxide at the site of use is inconvenient, then carbon monoxide is the preferred export.

Carbon monoxide is a cryogenic gas, and therefore costly to store. Metal carbonyls are excellent carrier compounds, being solid below −40°C. Iron and nickel carbonyls contain 70% carbon monoxide by weight, and are easily produced and decomposed.

Most types of noncarbonaceous asteroids are expected to be rich in nickel-iron (NiFe) metal. Carbon monoxide is a nearly ideal reagent in the processing of NiFe, via the carbonyl process. The process works at low temperatures, requiring little power or cooling equipment. It works at modest pressures, thus allowing an inexpensive reaction vessel. Best of all, it produces finished metal goods from unrefined NiFe in a single process step (Lewis and Nozette 1983; Lewis et al. 1988).

The carbonyl process consumes very little carbon monoxide. The only loss is the tiny percentage remaining in the ore and the metal product. Thus a very large foundry could operate with a small input of makeup gas. This is actually an argument against mining asteroids for carbon monoxide alone, because the amount needed could be brought from Earth at reasonable cost. However, because carbon monoxide will be a by-product of other processes, its production is essentially free.

F. Miscellaneous Fuels

A wide variety of fuels for propulsion can be produced from carbonaceous ore. This subsection will briefly mention some of the more promising fuels. Hydrogen has already been discussed in Sec. II.B.

Methane is a high specific impulse fuel which is rarely used currently. As methane and oxygen boil only 20°C apart, the two gases can be liquefied by the same equipment.

Kerosene is a fuel with a long history in rocketry. It has a specific impulse somewhat lower than methane, and is more difficult to produce, but does not require refrigeration.

Methanol has a still lower specific impulse, but is rather easier to produce than kerosene. Like kerosene, it requires no refrigeration.

Carbon monoxide has been suggested as a fuel in the context of propellant production on Mars, assuming no water is available (French 1989). Carbon monoxide/oxygen rockets have a low specific impulse compared to conventional systems. However, if a carbonaceous asteroid contains carbon but no water, CO/O_2 could become the propellant system of choice for local use.

Magnesium metal and magnesium hydride are very high energy fuels. They could be made from epsomite ($MgSO_4 \cdot 7H_2O$), which is easy to extract and purify, and comprises 5 to 15% of C1 ore (Gaffey and McCord 1979). The disadvantage in using these fuels is that they produce high molecular weight (and therefore low mean velocity) oxides. The specific impulse of these fuels could therefore be improved by adding water, hydrogen, or methane to the exhaust (Mul et al. 1990).

G. Rare Gases

A number of gases which are present in low or trace quantity in carbonaceous matter should be briefly mentioned because of their high desirability as by-products. These rare gases include nitrogen, the halogens, and the noble

gases.

Nitrogen comprises 0.1 to 0.3% of C1 and C2 ore (Moore 1971) and 1 to 3% of the organic matrix material (see Table II). It is released primarily as N_2 gas during pyrolysis of these materials (Wszolek et al. 1973), and could be recovered as a by-product. Cometary materials are likely to contain much higher percentages of nitrogen.

The primary market for nitrogen is as an inert filler gas in breathing air. While other gases (e.g., helium) have been used for such purposes, no other inert gases are available in like quantity in near-Earth space. As it is likely that any substitution would be noticed by the residents, nitrogen is the natural choice if it is available.

If an abundant source of nitrogen were available (from a comet for example) nitrogen-based propellants would be economically viable. Hydrazine (N_2H_4) is a widely used monopropellant which is catalytically decomposed to generate thrust on demand. Nitrogen tetroxide (N_2O_4) is a noncryogenic storable oxidizer, useful in rocketry as a substitute for liquid oxygen.

The halogens chlorine and fluorine are present in carbonaceous meteorites to the extent of ~0.03% and ~0.02%, respectively (Reed 1971). The similarity in concentration from C1 to C4 indicates that moderate heating does not volatilize halogens. Polymeric carbon extracted from carbonaceous meteorites by HF/HCl digestion contains ~1% each of F and Cl (Hayes 1967), indicating that most of the halogen resides in this fraction and could be released during high-temperature pyrolysis. The halogens are of interest mainly as reagents.

The noble gases helium, neon and argon represent a highly variable resource. Solar-wind implantation seems to be the primary source of the noble gases in gas-rich meteorites (Heymann 1971). Their uses include pressurizing gas for liquid propellants, cryogenic coolant, and inert atmospheres for high temperature processing.

The lunar regolith can be used as a rough guide to the upper limits of solar-wind implantation in carbonaceous asteroids. These atoms are implanted at a shallow depth ($< 100\mu$m), and should be released by moderate heating. The maximum concentrations seen in lunar regolith are N$\approx$200 ppm, Cl$\approx$200 ppm, F$\approx$400 ppm, He$\approx$100 ppm, Ne$\approx$10 ppm and Ar$\approx$10 ppm (Haskin and Warren 1991). The finest material is the richest in solar wind gases.

III. LESSONS FROM INDUSTRIAL CHEMISTRY

Before any large-scale engineering project can begin, it must receive financial approval. Investors require proof that the project is a good investment, and part of the proof is a comprehensive plan for minimizing risk. Accountants must be pessimistic precisely because engineers cannot help being optimistic. Every new technology has unforseen bugs that need to be worked out. A project which relies heavily on new technology will limp through life and die in shame.

TABLE II

Elemental Composition of Ash-free Organic Phases[a]

	C	H	O	N	S	Tot.	C/H	Reference
C1 meteorite	73.6	4.6	10.3	1.66	7.23	97.4	15.9	Hayes 1967
C2 meteorite	78.	3.1	13.	1.7	4.2	100.	25.	Zinner 1988
Oil Shale, KY	82.0	7.4	6.3	2.3	2.0	100.	11.1	Smith and Jensen 1987
Oil Shale, UT	80.5	10.3	5.8	2.4	1.0	100.	7.8	Smith and Jensen 1987
Bituminous, PA	82.4	5.6	9.0	1.6	1.3	100.	14.6	Smoot 1979
Anthracite	89.4	3.8	4.6	1.1	1.1	100.	23.6	Smoot 1979
Heavy Petroleum	85.3	11.0	0.5	0.4	2.8	100.	7.7	
Light Petroleum	85.7	12.7	0.3	0.2	1.1	100.	6.8	

[a] All compositions are weight percent. C1 is Orgueil, water- and solvent-insoluble fraction only. (Content includes 1.28% Cl, 1.31% F; may be contamination.) C2 is Murchison. C1 and C2 organic phases are obtained by HCl/HF digestion of inorganic phases. Some insoluble mineral matter remains, so the above data have been recalculated on an ash-free basis. For a discussion on the nature and origins of the organic matter in carbonaceous meteorites, see Cronin et al. (1988). Petroleum data are for comparison only; derived by author from data in Perry and Chilton (1973) and Grayson and Eckroth (1982).

Asteroid mining seems at first glance to have no established technological base to build upon. But while the physical environment is unprecedented, the raw materials bear a useful similarity to some earthly raw materials. The chemical engineering for processing asteroids should draw from existing technology, and thereby reduce the overall project risk.

On Earth, the technology for converting raw materials into volatile fuels is well developed. Coal can be gasified to $H_2 + CO$, hydrogenated to a mixture of hydrocarbons, or pyrolyzed to coal tar and coke. Petroleum is distilled to remove light hydrocarbons, and the residue is either cracked to yield more light hydrocarbons, or pyrolyzed to yield heavy hydrocarbons plus coke. Oil shale is pyrolyzed to give a mixture of light and heavy hydrocarbons, leaving a substantial carbonaceous residue which is discarded (Allred 1982).

The choice of technology is dominated by the carbon:hydrogen ratio (see Table II). A feedstock with a low ratio will be easily distilled. A high ratio will produce little or no volatile fraction. For petroleum C/H≈6–8 (by weight), for oil shale C/H≈8–13, and for coal and carbonaceous meteorites C/H ≈13–25 (see Table III). Experiments on carbonaceous meteorites confirm that little organic vapor is released during vacuum pyrolysis. Furthermore, pyrolysis produces a complex mixture of compounds, including many with limited utility in space. Thus oil shale and petroleum technology are generally of little use in processing carbonaceous asteroids.

TABLE III

C/H Ratio vs Yield of Volatiles[a]

C/H	Volatiles	Source of Coal
12.7	44.8%	Church Mine, UT, USA (bituminous)
13.7	39.9%	Illinois, USA (bituminous)
14.6	36.6%	Pittsburgh, PA, USA (bituminous)
16.4	28.9%	Pittsburgh (high-volatile bituminous)
19.6	16.3%	Sewell (medium-volatile bituminous)
23.6	8.8%	(low-volatile anthracite)
C/H	**Volatiles**	**Source of Oil Shale**
7.8	69. %	Garfield County, CO, USA
8.4	60. %	Glen Davis, Australia (Cannel Seam)
9.8	53. %	Ermelo, Transvaal, South Africa
11.1	33. %	New Albany, KY, USA
11.5	26. %	Glen Davis, Australia (Top Seam)
12.8	13. %	Pictou County, Nova Scotia, Canada

[a] C/H ratios and % Volatiles are by weight. Oil shale "volatiles" are the percentage of organic carbon converted and recovered as oil during Fischer assay. Coal "volatiles" are the percentage of the coal's weight (as mined) which is released as vapor during vacuum pyrolysis. Water is not included. Coal data are from Smoot (1979). Oil shale data are from Smith and Jensen (1987).

A more useful model for processing carbonaceous material is coal gasification technology. This technology should be tested on carbonaceous meteorite material. Of course, coal and carbonaceous chondrite have significant differences in mineral content. Also, economic constraints are different at a mining site in space. We can therefore expect some processes to require significant modification or even be totally unworkable. These processes are presented only as a starting point for a more extensive investigation.

An excellent introduction to modern coal technology can be found in *Riegel's Handbook of Industrial Chemistry*. The chapter on coal technology (Funk 1983) covers the conversion of carbon plus water to carbon monoxide plus hydrogen, and then to synthetic petrochemicals. The chapter on industrial gases (Neary 1983) discusses the production, purification, and liquefaction of hydrogen. The chapter on synthetic organic chemicals (Haberstroh and Collins 1983) details the synthesis of methanol from carbon monoxide and hydrogen.

The remainder of this section is a review of coal processing reactions which appear to be both useful and workable. A summary of the reactions is given in Table IV.

TABLE IV

Industrial Processes to be Tested for Use in Space

1. Synthesis gas via coal gasification (O_2 is also added)
2. The watergas shift, converting CO to H_2 (iron oxide catalyst)
3. Methanol production from synthesis gas (zinc oxide catalyst)
4. Alkenes via Mobil's M-gasoline process (zeolite catalyst)
5. Methane production from synthesis gas
6. Conversion of methane to hydrogen and carbon (Ni catalyst)

(1) $H_2O + C \xrightarrow{1400°C,10\ atm} CO + H_2$

(2) $H_2O + CO \xrightarrow{450°C} CO_2 + H_2$

(3) $2H_2 + CO \xrightarrow{350°C,300\ atm} CH_3OH$

(4) $nCH_3OH \xrightarrow{400°C,2\ atm} C_nH_{2n} + nH_2O$

(5a) $2H_2 + 2CO \longrightarrow CH_4 + CO_2$
(5b) $3H_2 + CO \longrightarrow CH_4 + H_2O$

(6a) $CH_4 + CO_2 \xrightarrow{900°C,1\ atm} 2CO + 2H_2$
(6b) $2CO \longrightarrow CO_2 + C$

A. Synthesis Gas

Coal gasification is widely used to convert coal into useful industrial gases. The coal reacts with water, at 1000 to 1800°C and 1 to 100 atm pressure. Oxygen is added to generate the process heat which drives the reaction. The exit gases are cooled, separated from liquid and solid by-products, and scrubbed to remove acid gases. The product is a mixture of carbon monoxide and hydrogen gas, known as synthesis gas.

If hydrogen alone or carbon monoxide alone is the desired product, the gases can be separated cryogenically or by scrubbing. More commonly, the gas mixture is modified by the watergas shift, and used to synthesize other chemicals. The three most common end-products are methane, methanol and synthetic petroleum.

B. The Watergas Shift

The watergas shift is an important subsidiary process in reactions involving synthesis gas. It is used to optimize the carbon monoxide to hydrogen ratio for increased yields. Water is added to the synthesis gas just downstream of the gas generator. The amount of water added must be carefully controlled to obtain the desired $CO:H_2$ ratio. The water reacts exothermically with the carbon monoxide to produce carbon dioxide and more hydrogen. The carbon dioxide is then removed in the scrubber with the other acid gases.

C. Synthetic Methanol

Nearly all methanol is currently produced from synthesis gas. The synthesis gas is compressed to 300 atm, heated to 350°C, and reacted in contact with a zinc oxide/chromium oxide catalyst. Work is proceeding on a reduced-pressure process, which would allow a much less massive reactor.

D. Synthetic Gasoline

Mobil's M-gasoline process converts methanol into high-octane gasoline. The gasoline retains about 95% of the methanol's fuel energy, but only 44% of its weight. The product, a mixture of low-boiling alkenes, can be used without fractional distillation.

E. Synthetic Methane

While most of the world's methane supply is natural, significant quantities are produced from coal via synthesis gas. A plant in Beulah, North Dakota produces 700,000 tons per year of synthetic methane from low-rank coal. The highly exothermic reaction produces methane, carbon dioxide and water. The water is removed by condensation. The carbon dioxide is removed by scrubbing with water at 30 atm pressure. The last traces of each are removed by molecular sieves if the methane is to be liquefied.

F. Methane Decomposition

If methane is used as a hydrogen carrier, a means must be devised to recover

the hydrogen. A reaction used for making synthesis gas from methane can be pressed into service. Carbon dioxide is reacted with the methane at 900°C to produce carbon monoxide and hydrogen. The exit gases are cryogenically separated. The hydrogen is stored or used immediately, and the unreacted methane and carbon dioxide are recycled. The carbon monoxide is heated with a catalyst, causing it to disproportionate into graphite, which is stored, and carbon dioxide, which is returned to the first step. The net reaction is $CH_4 \longrightarrow 2H_2 + C$.

IV. HOW TO PROCESS A CARBONACEOUS ORE

This section proposes some processes for use with carbonaceous ore, based on current industrial practice. These processes should be studied, tested and improved, as discussed in the next section.

Process 1: Water Production. The crushed ore is placed in a closed vessel and heated to 400°C (Hashimoto et al. 1979) with steam at 1 atm pressure. The vapor is removed, cooled to 10°C, and solids and gases are separated from the liquid water. The water (some of which is recycled to the first step) is outgassed briefly in vacuum to remove dissolved gases, and placed in storage.

Process 2: Hydrogen/Oxygen Production. Water from Process 1 is electrolyzed or thermolyzed. The two resulting gas streams are cooled to condense water vapor, then dried further with molecular sieves. The gases are liquefied and placed in storage. The process should be located near the point of use. Many designs for this process have appeared in the literature (Bock and Fisher 1978; Heald et al. 1978; Schuster et al. 1986).

Process 3: Carbon Monoxide Production. The crushed ore is placed in a vessel with carbon monoxide gas. The gas serves as a nonreactive heat transfer medium to more quickly heat the ore. The vessel is heated until gas begins to evolve. The resulting gas phase (H_2O, CO_2, and organics) is then continuously flushed out with preheated CO to prevent it from reacting with the fixed carbon. When gas ceases to evolve, the temperature is raised to 1000°C. The fixed carbon and magnetite in the ore then react to form carbon monoxide and iron metal (Hashimoto et al. 1979; Gooding 1975). The exit gases are cooled to −150°C, and the condensed impurities (S_8, SO_2, and Na_2O) are removed. The CO is then liquefied and placed in storage.

Process 4: Methane/Oxygen Propellant Production. Crushed ore is heated with steam to 1000°C in a closed vessel. The volatiles are drawn off and worked up to give purified H_2O, H_2 and CO. The water is electrolyzed as in Process 2, with only the oxygen placed in cryogenic storage. The hydrogen and the carbon monoxide are sent to a methanator. The resulting methane is purified, liquefied and placed in storage.

V. TESTING A PROPOSED PROCESS

This section discusses techniques for testing the processes proposed for car-

bonaceous ores. A new process for use on Earth can be tested at lab-scale and pilot-scale with real ore samples. Unfortunately, the small supply of C1/C2 meteorites and the high cost of increasing the supply make this testing approach impractical for carbonaceous ore.

Thorough testing of a process is not only much more difficult with asteroidal resources. It is also much more important. The cost of adding to or modifying the equipment, once delivered, will be prohibitive. Therefore, every effort must be made to fully utilize existing process simulation methods, to develop effective new methods, and to design intelligent and flexible hardware.

A. Computer Simulation of Ore Processing

Computer modeling of carbonaceous ore processing has just begun recently (Ganguly and Saxena 1989; Bose and Ganguly 1991). It would be desirable to use such methods to predict the products of pyrolyzing or gasifying carbonaceous ore. Researchers must simultaneously refine their models of carbonaceous mineralogy and thermodynamic parameters. Once these milestones are achieved, however, computer modeling will allow rapid testing, comparing, and optimization of processes. This area of study is vital to the future of space resources.

B. Processing of Carbonaceous Ore Simulant

Some form of simulant is clearly desirable for pilot-scale testing. Lunar simulants have been developed (Weiblen et al. 1990) and successfully employed in processing experiments (see, e.g., Ness et al. 1991; Garvey and Magoffin 1991). Carbonaceous meteorite simulants do not currently exist, and may not be possible with current technology. Carbonaceous materials are microcrystalline to amorphous, contain many nonsilicate phases, and are not in equilibrium. A realistic near-term research goal would be the use of computer models to design limited simulants. These materials would be able to mimic the response of carbonaceous ore to a limited range of processing conditions.

C. Micro-Scale Experiments on Carbonaceous Meteorites

Experiments on carbonaceous meteorites have concentrated on scientific, rather than engineering, problems. The most useful for present purposes involve pyrolysis. Stepwise pyrolysis with gas chromatography (GC)/mass spectroscopy (MS) was performed on C3 Allende (Levy et al. 1970) and on C3 Allende, C2 Murchison and C2 Murray (Studier et al. 1972). Slow pyrolysis experiments with MS include C3 Allende (Gooding 1975), C2 Murchison (Hashimoto et al. 1979), C3 Allende and C2 Murchison (Simoneit et al. 1973), and C2 Murray (Wszolek et al. 1973). Extraction with organic solvents was also used with GC/MS (Studier et al. 1972) and with absorption and fluorescence spectrometry (Hodgson and Baker 1969).

An important goal of future experiments is to provide verification of computer models. Experiments should be designed to exercise a wide range

of reaction conditions. Temperature should be varied from −200°C to 1600°C. Pressure should be varied from vacuum to 100 atm. Aqueous pH should vary from 0 to 14. Strong oxidizing and reducing, as well as neutral, environments should be tested. The goal is to provide reference experiments which create confidence that all the model parameters have been thoroughly tested.

More specifically, experiments should measure the composition of: water or acid extracts as a function of temperature, pressure and contact time; products of contact with a hydrogen, oxygen, steam, carbon monoxide, or carbon dioxide atmosphere as a function of temperature, pressure and contact time; and the reduced metal phase as a function of pyrolysis temperature and pretreatment conditions. This is far from an exhaustive list.

D. Pilot-Plant Testing

As the process approaches implementation, scale-up will be necessary. A pilot plant should be built and tested in Earth orbit. At first it will process simulants. When it is working reliably with simulants, it is time to feed it real ore. A large-scale sample return is probably not economically practical. Therefore, it will be necessary to send the pilot plant to the mining site. The delivered pilot plant should be highly reprogrammable and reconfigurable so that the process can be adapted to unforeseen problems. If Earth-based mining experience is any guide, there will be plenty of unforeseen problems.

VI. THE MOONS OF MARS AS ORE BODIES

It may seem unwarranted to include the satellites of Mars in a discussion of carbonaceous asteroids, as they are not technically asteroids. They are, however, quite asteroid-like, in size, shape, and rotational rate (Veverka and Thomas 1979). They may even have been formed in the asteroid belt, and later captured by Mars (Hartmann 1987). In any event, they are potential sources of volatiles, and seem to be carbonaceous.

The Martian moons, Phobos and Deimos, are of special interest as resources because of their location. In addition to their value as bases for the exploration of Mars, they are more accessible more often than any known asteroids (O'Leary 1985). Their proximity to Mars has also allowed them to be studied by four spacecraft at close range. Even after the upcoming asteroid flyby missions, they will remain the best understood small bodies in the solar system.

This section presents both data and speculation on the nature of the mining environment. Hard data has been obtained not only from the numerous spacecraft encounters, but also from Earth-based telescopes. The speculation is informed by observations of the other bodies in the solar system, the study of meteorites, and computer models of the micro- and macroscopic processes involved in the solar system's formation.

The level of our current knowledge is much less than a mining geologist would like, but much more than might be expected. The loss of PHOBOS 1

and 2, the only two spacecraft ever dedicated to the study of an "asteroid," is deeply felt.

A. Summary of Observations

The data from the two Viking Orbiters provided many details on the Martian moons (for a review, see Veverka and Thomas 1979). The two moons are similar in color, size, shape, surface gravity, density and orbit. Their color is black, with 4% reflectance and a nearly featureless spectrum. Their size is 21 km (Phobos) and 13 km (Deimos) in mean diameter. Their shape is roughly ellipsoidal ($1.5 \times 1.3 \times 1.0$), and their surfaces are saturated with craters. Their surface gravities are ~0.0005 Earth gravity. Their density is ~2 g cm^{-3}. Their orbits are circular in the plane of Mars' equator. Craters on Phobos are distinct, while craters on Deimos are often nearly obliterated by ejecta.

B. Composition

The composition of the moons is not firmly established. The author favors the hypotheses of Hartmann (1987,1990). He proposes that an icy carbonaceous asteroid formed in the cold outer belt, and was subsequently captured by Mars. An early impact split the body, creating two moons whose orbits then evolved separately. Since then, the surface material has suffered many impacts, causing loss of volatiles. The surface material has been extensively reworked, because most impact ejecta go into orbit around Mars and are eventually redeposited. The near-surface ice has been lost, both by impact vaporization (Grimm and McSween 1989) and by slow sublimation (Fanale and Salvail 1990). Ice remains at depths of ~500 m at the equator and ~50 m at the poles.

Based on laboratory and spacecraft measurements of spectral reflectance, the proposed meteorite analogs are C1 or C2 (Pang et al. 1978; Pollack et al. 1978; Veverka 1978) and shock-blackened ordinary chondrite. It now appears that no known material matches the moons if spectra in the ultraviolet, visible and infrared are included (Britt and Murchie 1991).

C1 and C2 meteorites are composed primarily of water-bearing phyllosilicates (clays), which show a characteristic 3 μm absorption feature in the laboratory. The 3 μm feature has also been observed on many type C asteroids, indicating that they are hydrated. This feature is weak or absent in spectra of Phobos' surface obtained by the Phobos 2 spacecraft (Bibring et al. 1989). Thus C1 and C2 do not seem to be good model compositions for Phobos.

The arguments against a type C3 or ordinary chondrite composition hinge on their high density. Phobos has a mass of $1.08(\pm.01) \times 10^{19}$ g (Avanesov et al. 1989) and a volume of $5.68(\pm.25) \times 10^{18}$ cm^3 (Duxbury, personal communication). Its calculated density is therefore 1.90(±.09) g cm^{-3}. By comparison, densities are typically 2.2 for C1, 2.6 to 2.9 for C2, 3.4 for C3, and 3.3 to 3.9 for ordinary chondrites (Mason 1966). If Phobos were made of ordinary chondrite, the bulk density would require that nearly the entire

body consist of uncompacted rubble, with 40–54% mean void volume. Logic argues against such a high void content in a compaction environment characterized by constant bombardment and 100 to 600 g cm^{-2} (Thomas et al. 1986) internal gravitational pressure.

It should be noted that such high void content does occur naturally under special conditions. Regolith at the lunar surface and well-sorted sediments on Earth have similarly high void content. However, the density of lunar regolith rises rapidly with depth (due to compaction processes), and there is no plausible mechanism for creating well-sorted rubble on Phobos.

Observations also argue against such a high void content. "On Phobos, patterns of grooves and ridges suggest a moderately competent internal structure," rather than a rubble pile (Thomas et al. 1986). Also, the near-surface material is strong enough in many places to produce large blocks of coherent crater ejecta.

Having found difficulties with all of the proposed meteorite analogs, we must consider the possibility that Phobos is composed of some material not represented in meteorite collections. The material is similar in density and spectral reflectance to C1 and C2 chondrite but contains little or no hydrated silicate. Either the silicates never hydrated, or they did hydrate and then were dehydrated.

Theoretically, the silicates in an icy asteroid which has remained forever frozen should be anhydrous, while silicates in contact with liquid water for ~1000 yr should be completely hydrated (Grimm and McSween 1989). This claim is supported by recent observations that most type C asteroids in the warm inner belt show absorption at 3 μm, while those in the cold outer belt do not (Lebofsky et al. 1990; Jones et al. 1990).

The author therefore proposes C1 plus ~10% water ice as the initial composition of the Martian moons. This prediction is consistent with the observed density of Phobos. Its high content of volatiles would cause it to break up quickly in an encounter with Earth's atmosphere, explaining the lack of corresponding meteorites on Earth. The surface dust would be low in volatiles due to repeated impact, but would retain fixed carbon. The interior would be cracked rather than shattered, and would retain most of its primordial ice.

C. The Mining Environment

Phobos and Deimos have been observed from Mariner 9 (1971), Viking Orbiters I and II (1976), and the Soviet craft Phobos 2 (1988) as well as from Earth. The results from these observations which are relevant to mining operations are summarized here.

The rotational axes of Mars and its moons are inclined 25° to the ecliptic. The moons each keep one hemisphere permanently facing Mars (Burns 1978). Sunlight on the inner faces is blocked by Mars once per orbit near the equinoxes, but not near the solstices. Phobos is eclipsed by Mars more frequently and during more of the year than Deimos.

Sunlit surfaces are exposed to solar-wind protons at 700 km s^{-1}. During eclipses, the surface is exposed to Mars' plasma tail of atomic oxygen at 150 km s^{-1} (Lundin et al. 1989; Rosenbauer et al. 1989). The entire surface is subjected to micrometeorites and cosmic rays. It should be noted that the Moon is subjected to similar particle fluxes.

The effective gravitational force on the moons' surfaces is highly variable with position. Four factors contribute to the field: gravitational attraction to the moon's center of gravity, gravitational anomalies due to the moons' nonspherical shape, centrifugal force of their rotation and their orbit around Mars, and tidal forces from Mars. On Phobos, all four factors are of the same order of magnitude (Davis et al. 1981; Dobrovolskis and Burns 1980).

The surface temperature of both moons has been modeled by computer (Kührt and Giese 1989; Giese and Kührt 1990). The model's predictions were verified by the Phobos 2 spacecraft (Ksanfomality et al. 1989). Surface temperature extremes are +40°C at the equator and −210°C at the poles. The axial tilt causes large annual temperature swings as a function of latitude. The orbital eccentricity also causes large annual temperature swings, as a function of solar distance. The direction of axial tilt when the Sun is closest causes maximum temperatures at the south poles to be higher than at the north poles. The minimum temperature on the Mars-facing hemispheres is higher than on the anti-Mars faces, due to thermal radiation from the Martian surface.

The temperature in the interior throughout history has also been modeled. The model was used to predict whether a conjectured icy component could have survived to the present. The conclusion was that primordial ice could still be present at depths of 10 m near the poles and 100 to 1000 m near the equator (Fanale and Salvail 1990).

Viking photography provides complete coverage of both bodies at a resolution of 200 m, and of some areas at a resolution of 10 m (Duxbury 1978). The shape of Phobos has been described in the form of a topographic map (Turner 1978), a sixth-order mathematical surface (Duxbury 1989), a globe (Turner 1978), and a computer model of radius versus latitude and longitude (Stooke 1989).

The surfaces of both moons are seen to be covered with a regolith of powdered rock. Regolith depth is difficult to estimate from photographs. Craters on Deimos seem to contain 6 to 12 m while craters on Phobos contain very little (Thomas and Veverka 1980). The regolith depth in valleys may be 200 m or more due to the downslope movement of surface deposits (Veverka and Thomas 1979).

Regolith particle size has been determined by two independent methods. The thermal inertia parameter of Phobos indicates "mean grain sizes tens to hundreds of micrometres" (Ksanfomality et al. 1989). Spectrally, bright material on Deimos matches C2 powder $<40\,\mu$m in size, while dark material matches 75 to 150 μm C2 powder (French et al. 1988).

Both moons exhibit many large (10 to 200 m) blocks on the surface, usually near large craters (Lee et al. 1986). Large blocks are especially

prominent near Stickney, the largest crater on Phobos. The degree to which the moons have been internally disrupted remains controversial.

A model of the interior of Phobos offers some insight into the relationship between external appearance and internal structure (Fujiwara and Asada 1983). Clay ellipsoids were fragmented by high-velocity projectiles to determine the effect of a near-catastrophic impact such as Stickney. The model was shattered and dispersed near the impact, and cracked into large chunks far from the impact. The resulting distribution of cracks is similar to the grooves observed on Phobos.

D. Mining Economics

Leonard et al. (1987) designed a commercial mining operation for producing water on Phobos or Deimos, for export to LEO. Their stated market was for propellant, needed to carry out the National Commission on Space scenario for a lunar base and a manned Mars mission. They proposed an automated mining, processing, and transport system for delivering water to LEO. The ice barges had a payload of 200 tons and a transit time of about 15 months. The ion rockets were powered by 10 MW electric plants. Their system cost estimate was $1 billion to 5 billion.

VII. CONCLUSIONS

The National Commission on Space (NCS) proposal for a manned lunar base and a manned exploration of Mars has created a context for the space program of the next century. If that proposal is implemented, the resources of near-Earth space could make the project much less expensive and much more self-sufficient. The timetable for designing an asteroid mine could probably match the NCS timetable.

The fundamental research and process development outlined in this review could be completed in about 6 yr. The prospecting spacecraft design and the asteroid selection could proceed in parallel with the research and development program. Allowing 1 yr for the prospecting mission, 5 yr for the pilot plant to be designed, tested and delivered, and 3 yr for the design and delivery of the scaled-up plant, the total time required comes to about 15 yr.

Recent history shows that few government programs survive that long. Thus, we might conclude that a substantial increase in long-term thinking is needed before space resources can play a leading role in public policy. Alternatively, the mining of asteroids could take the path of slow but steady progress, following rather than leading the breakout into space. While this author would prefer to see a National Asteroid Mining Initiative, he would gladly settle for a program based on sound economics. A reasonable precondition for building a $3 billion asteroid mine is to have clear evidence of an economic advantage worth more than the cost. In space as on Earth, investments in infrastructure must somehow pay for themselves, whether they are publicly or privately funded.

Assuming that these preconditions can be met in the future, several near-term problems must be solved to keep the asteroid mining option open. Now that Spacewatch has broken the logjam of asteroid discovery, a new bottleneck has formed in the follow-up area. A list of 1000 new NEA discoveries is of no use if they are characterized only by orbital elements. High-quality spectral reflectance data on every NEA are most urgently needed.

The design of a plant for mining and processing asteroids presents a novel challenge. This may be the first design for which trial and error is completely unaffordable, and there is no established technology. Much development is needed in process simulation. This is an area with obvious spin-off potential. Every mining engineer would love to have a means of trying out a dozen variants on his process without bending any metal. Improved simulation requires improved models of mineralogy and better thermodynamic data. Once developed for asteroid mining, the rest of the world will benefit from it.

More experimental work is needed on processing of the asteroidal material we already have: meteorites. To date, nothing has been published on how carbonaceous material actually behaves in even the simplest proposed process. This is a fruitful area in economic terms. A small input of time and equipment can multiply our knowledge a hundred-fold.

Clever and adaptable robotics technology will be needed for the pilot plant. When it is needed, it probably will not be ready. This is a big future bottleneck waiting to happen.

Developing the new frontier in space will be slow, but it is guaranteed to be an exciting adventure. This is the first frontier in history which will be opened by engineers and scientists. It offers us an opportunity to learn to be responsible visionaries.

REFERENCES

Agosto, W. N. 1981. Beneficiation and powder metallurgical processing of lunar soil metal. In *Space Manufacturing 4, Proc. of the Fifth Princeton/AIAA Conf.*, eds. J. Grey and L. Hamdan (New York: AIAA), pp. 365–370.

Allred, V. D. 1982. Oil shale retorting phenomenology. In *Oil Shale Processing Technology*, ed. V. Allred (East Brunswick, N. J.: The Center for Professional Advancement), pp. 55–66.

Avanesov, G. A., Bonev, B. I., Kempe, F., Basilevsky, A., Boycheva, B., Chikov, K., Danz, M., Dmitrov, D., Duxbury, T., Gromatikov, P., Halmann, D., Head, J., Heifets, V., Kolov, V., Kostenko, V., Kottsov, V., Krasavtsev, V., Krasikov, V., Krumov, A., Kuzmin, A., Losev, K., Lumme, K., Mishev, D., Möhlmann, D., Muinonen, K., Muravev, V., Murchie, S., Murray, B., Neumann, W., Paul, L., Petkov, D., Petuchova, I., Pössel, W., Rebel, B., Shkuratov, Yu., Simeonov, S., Smith, B., Totev, A., Uzunov, Yu., Fedotov, V., Weide, G.-G., Zapfe, H., Zhukov,

B., and Ziman, Ya. 1989. Television observations of Phobos: First results. *Nature* 341:585–587.

Bibring, J.-P., Combes, M., Langevin, Y., Soufflot, A., Cara, C., Drossart, P., Encrenaz, Th., Erard, S., Forni, O., Gondet, B., Ksanfomality, L., Lellouch, E., Masson, Ph., Moroz, V., Rocard, F., Rosenqvist, J., and Sotin, C. 1989. Results from the ISM experiment. *Nature* 341:591–593.

Bock, E. H., and Fisher, J. G. 1978. In-space propellant processing using water delivered as a shuttle contingency payload. Paper 78-941, presented at the 14th AIAA/SAE Joint Propulsion Conf., July 25–27.

Bose, K., and Ganguly, J. 1991. Kinetics of volatile extraction from carbonaceous chondrites: Dehydration of talc. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 13.

Britt, D., and Murchie, S. L. 1991. The composition of Phobos: Meteorite analogs based on KRFM and VSK spectral data from the Phobos 2 spacecraft. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 24.

Burns, J. A. 1978. The dynamical evolution and origin of the martian moons. *Vistas in Astronomy* 22:193–210.

Cordell, B., and Steinbronn, O. 1988. An analysis of possible advanced space strategies featuring the role of space resource utilization. Paper presented at the 39th Intl. Astronautical Congress. IAA-88-587.

Cronin, J. R., Pizzarello, S., and Cruikshank, D. P. 1988. Organic matter in carbonaceous chondrites, planetary satellites, asteroids and comets. In *Meteorites and the Early Solar System*, eds. J. F. Kerridge and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 819–857.

Cutler, A. H., and Hughes, M. L. 1985. Transportation economics of extraterrestrial resource utilization. In *Space Manufacturing 5: Engineering with Lunar and Asteroidal Materials*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 233–244.

Davis, D. R., Housen, K. R., and Greenberg, R. 1981. The unusual dynamical environment of Phobos and Deimos. *Icarus* 47:220–233.

Dobrovolskis, A. R., and Burns, J. A. 1980. Life near the Roche limit: Behavior of ejecta from satellites close to planets. *Icarus* 42:422–441.

Duxbury, T. C. 1978. Spacecraft imaging of Phobos and Deimos. *Vistas in Astronomy* 22:149–161.

Duxbury, T. C. 1989. The figure of Phobos. *Icarus* 78:169–180.

Erstfeld, T. E., Mullins, O., Jr., and Williams, R. J. 1979. Carbon dioxide electrolysis using a ceramic electrolyte. In *Space Manufacturing Facilities 3*, eds. J. Grey and C. Krop (New York: AIAA), pp. 83–88.

Fanale, F. P., and Salvail, J. R. 1990. Evolution of the water regime of Phobos. *Icarus* 88:380–395.

French, J. R. 1989. Rocket propellants from Martian resources. *J. British Interplanet. Soc.* 42:167–170.

French, L. M., Veverka, J., and Thomas, P. 1988. Brighter material on Deimos: A particle size effect in a carbonaceous material? *Icarus* 75:127–132.

Fujiwara, A., and Asada, N. 1983. Impact fracture patterns on Phobos ellipsoids. *Icarus* 56:590–602.

Funk, J. E. 1983. Coal technology. In *Riegel's Handbook of Industrial Chemistry*, ed. J. A. Kent (New York: Van Nostrand Reinhold), 8th ed., pp. 66–129.

Gaffey, M. J., and McCord, T. B. 1979. Mineralogical and petrological characterizations of asteroid surface materials. In *Asteroids*, ed. T. Gehrels (Tucson: Univ. of Arizona Press), pp. 688–723.

Ganguly, J., and Saxena, S. K. 1989. Theoretical predictions of volatile bearing phases

and volatile resources in some carbonaceous chondrites. In *Space Manufacturing 7: Space Resources to Improve Life on Earth*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 97–105.

Garvey, J., and Magoffin, M. 1991. Lunar resource processing using solar energy: A research project status report. In *Space Manufacturing 8: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 143–149.

Gertsch, R. E., Barnes, R., Cutler, A. H., and Gaffey, M. 1985. Space resources. In *Proc. of the Fourth Annual L5 Space Development Conf.*, ed. F. Hecker, AAS 85–746 to 85–749 (San Diego: Univelt), pp. 63–90.

Giese, B., and Kührt, E. 1990. Theoretical interpretation of infrared measurements at Deimos in the framework of crater radiation. *Icarus* 88:372–379.

Gooding, J. L. 1975. A High-Temperature Study on the Vaporization of Alkalis from Molten Basalts Under High Vacuum: A Model for Lunar Volcanism. M. S. Thesis, Univ. of Hawaii.

Gradie, J., and Tedesco, E. 1982. Compositional structure of the asteroid belt. *Science* 216:1405–1407.

Grayson, M., and Eckroth, D., eds. 1982. Petroleum—survey of refinery process. In *Kirk-Othmer Encyclopedia of Chemical Technology*, vol. 17, 3rd ed. (New York: Wiley), p. 189.

Grimm, R. E., and McSween, H. Y. 1989. Water and the thermal evolution of carbonaceous chondrite parent bodies. *Icarus* 82:244–280.

Haberstroh, W. H., and Collins, D. E. 1983. Synthetic organic chemicals. In *Riegel's Handbook of Industrial Chemistry*, ed. J. A. Kent (New York: Van Nostrand Reinhold), 8th ed., pp. 66–129.

Hartmann, W. K. 1987. A satellite-asteroid mystery and a possible early flux of scattered C-class asteroids. *Icarus* 71:57–68.

Hartmann, W. K. 1990. Additional evidence about an early flux of C asteroids and the origin of Phobos. *Icarus* 87:236–240.

Hashimoto, A., Kumazawa, M., and Onuma, N. 1979. Evaporative metamorphism of primitive dust material in the early solar nebula. *Earth Planet. Sci. Lett.* 43:13–21.

Haskin, L., and Warren, P. 1991. Lunar chemistry. In *Lunar Sourcebook—A User's Guide to the Moon*, eds. G. H. Heiken, D. T. Vaniman and B. M. French (Cambridge: Cambridge Univ. Press), pp. 434–448.

Hayes, J. M. 1967. Organic constituents of meteorites—A review. *Geochim. Cosmochim. Acta* 31:1395–1440.

Heald, D. A., Blatt, M. H., Bock, E. H., Bradley, R. E., Drowns, R. E., and Leonhard, K. E. 1978. Orbital Propellant Handling and Storage Systems for Large Space Programs. NASA Contractors Rept. JSC-13967 (Houston: Johnson Space Center).

Heymann, D. 1971. The inert gases: He, Ne, Ar, Kr and Xe. In *Handbook of Elemental Abundances in Meteorites*, ed. B. Mason (New York: Gordon and Breach), pp. 29–44.

Hodgson, G. W., and Baker, B. L. 1969. Porphyrins in meteorites: Metal complexes in Orgueil, Murray, Cold Bokkeveld, and Mokoia carbonaceous chondrites. *Geochim. Cosmochim. Acta* 33:943–958.

Jones, T. D., Lebofsky, L. A., Lewis, J. S., and Marley, M. S. 1990. The composition and origin of the C, P, and D asteroids: Water as a tracer of thermal evolution in the outer belt. *Icarus* 88:172–192.

Ksanfomality, L.V., Moroz, V.I., Bibring, J. P., Combes, M., Soufflot, A., Ganpantzerova, O. F., Goroshkova, N. V., Zharkov, A. V., Nikitin, G. E., and Petrova, E. V. 1989. Spatial variations in thermal and albedo properties of the surface of

Phobos. *Nature* 341:588–591.
Kührt, E., and Giese, B. 1989. A thermal model of the martian satellites. *Icarus* 81:102–112.
Lebofsky, L., Jones, T. D., Owensby, P. D., Feierberg, M. A., and Consolmagno, G. J. 1990. The nature of low-albedo asteroids from 3-μm multicolor photometry. *Icarus* 83:16–26.
Lee, S. W., Thomas, P., and Veverka, J. 1986. Phobos, Deimos, and the moon: Size and distribution of crater ejecta. *Icarus* 68:77–86.
Leonard, R. S., Blacic, J. D., and Vaniman, D. T. 1987. The economics of mining the Martian moons. In *Space Manufacturing 6: Nonterrestrial Resources, Biosciences, and Space Engineering*, eds. B. Faughnan and G. Maryniak (New York: AIAA), pp. 380–395.
Levy, R. L., Wolf, C. J., Grayson, M., Gilbert, L., Updegrove, W. S., Zlatkis, A., and Oró, J. 1970. Organic analysis of the Pueblito de Allende meteorite. *Nature* 227:148–150.
Lewis, J. S., and Nozette, S. 1983. Extraction and purification of iron-group and precious metals from asteroidal feedstocks. In *Space Manufacturing 1983*, eds. J. D. Burke and A. S. Whitt (San Diego: Univelt), pp. 351–353.
Lewis, J. S., Jones, T. D., and Farrand, W. H. 1988. Carbonyl extraction of lunar and asteroidal metals. In *Engineering, Construction and Operations in Space*, eds. S. Johnson and J. Wetzel (New York: American Soc. of Civil Engineers), pp. 111–122.
Lundin, R., Zakharov, A., Pellinen, R., Borg, H., Hultquist, B., Pissarenko, N., Dubinin, E. M., Barabash, S. W., Liede, I., and Koskinen, H. 1989. First measurements of the ionospheric plasma escape from Mars. *Nature* 341:609–612.
Mason, B. 1966. The carbonaceous chondrites. *Space Science Rev.* 1:642.
Meyer, T. R., and McKay, C. P. 1989. The resources of Mars for human settlement. *J. British Interplanet. Soc.* 42:147–160.
Moore, C. B. 1971. Nitrogen. In *Handbook of Elemental Abundances in Meteorites*, ed. B. Mason (New York: Gordon and Breach), p. 96.
Mul, J. M., Korting, P. A. O. G., and Schoyer, H. F. R. 1990. A search for new storable high-performance propellants. *ESA Journal* 14:253–270.
Neary, R. M. 1983. Industrial gases. In *Riegel's Handbook of Industrial Chemistry*, ed. J. A. Kent (New York: Van Nostrand Reinhold), 8th ed., pp. 607–630.
Ness, R. O., Jr., Sharp, L. L., Runge, B. D., Knudsen, C. W., and Gibson, M. A. 1991. Hydrogen reduction of lunar simulants for the production of oxygen in a continuous fluid-bed reactor. In *Space Manufacturing 8: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 325–330.
O'Leary, B. 1985. Phobos and Deimos (PhD): Concept for an early human mission for resources and science. In *Space Manufacturing 5: Engineering with Lunar and Asteroidal Materials*, ed. B. Faughnan and G. Maryniak (New York: AIAA), pp. 41–48.
Pang, K. D., Pollack, J. B., Veverka, J., Lane, A. L., and Ajello, J. M. 1978. The composition of Phobos: Evidence for carbonaceous chondrite surface from spectral analysis. *Science* 199:64–66.
Perry, R. H., and Chilton, C. H. 1973. *Chemical Engineers' Handbook*, 5th ed. (New York: McGraw-Hill), pp. 8–11.
Pollack, J. B., Veverka, J., Pang, K., Colburn, D., Lane, A. L., and Ajello, J. M. 1978. Multicolor observations of Phobos with the Viking lander cameras: Evidence for a carbonaceous chondritic composition. *Science* 199:66–69.
Reed, G. W., Jr. 1971. Fluorine and chlorine. In *Handbook of Elemental Abundances in Meteorites*, ed. B. Mason (New York: Gordon and Breach), pp. 103–106;

143–147.

Rosenbauer, H., Shutte, N., Apathy, I., Galeev, A., Gringauz, K., Grunwaldt, H., Hemmerich, P., Jockers, K., Kiraly, P., Kurova, G., Livi, S., Marsch, E., Richter, A., Riedler, W., Remizov, T., Schwenn, R., Schwingenschuh, K., Steller, M., Szego, K., Verigin, M., and Witte, M. 1989. Ions of martian origin and plasma sheet in the martian magnetosphere: Initial results of the TAUS experiment. *Nature* 341:612–614.

Rosenberg, S., Guter, G., and Miller, F. 1965. Manufacture of oxygen from lunar materials. *Annals of the New York Academy of Science* 123:1106–1122.

Schuster, J. R., Bennett, F. O., Alton, T. J., Johns, W. A., Kim, I., Liggett, M. W., Simon, M. C., and Torre, C. N. 1986. Long Term Cryogenic Storage Facility Systems Study, Interim Report, Part A: Executive Summary. NASA GDSSD-SP-86-038, prepared for Marshall Space Flight Center by General Dynamics Space Systems Division, San Diego, Ca.

Shoemaker, E. M., Williams, J. G., Helin, E. F., and Wolfe, R. F. 1979.Earthcrossing asteroids: Orbital classes, collision rates with Earth, and origin. In *Asteroids*, ed. T. Gehrels (Tucson: Univ. of Arizona Press), pp. 783–806.

Simoneit, B. R., Christiansen, P. C., and Burlingame, A. L. 1973. Volatile element chemistry of selected lunar, meteoritic, and terrestrial samples. *Proc. Lunar Sci. Conf.* 4, *Geochim. Cosmochim. Acta Suppl.* 2:1635–1650.

Smith, J. W., and Jensen, H. B. 1987. Oil shale. In *Encyclopedia of Science and Technology*, 6th ed. (New York: McGraw-Hill) 12:333–340.

Smoot, L. D. 1979. General characteristics of coal. In *Pulverized-Coal Combustion and Gasification*, eds. L. Smoot and D. Pratt (New York: Plenum Press), pp. 123–132.

Stooke, P. J. 1989. Sizing up Phobos. *Sky & Telescope* 1989:477.

Studier, M. H., Hayatsu, R., and Anders, E. 1972. Origin of organic matter in the early solar system: V. Further studies of meteoritic hydrocarbons and a discussion of their origin. *Geochim. Cosmochim. Acta* 36:189–215.

Tholen, D. J., and Barucci, M. A. 1989. Asteroid taxonomy. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 298–315.

Thomas, P., and Veverka, J. 1980. Downslope movement of material on Deimos. *Icarus* 42:234–250.

Thomas, P., Veverka, J., and Dermott, S. 1986. Small satellites. In *Satellites*, ed. J. A. Burns (Tucson: Univ. of Arizona Press), pp. 802–835.

Turner, R. 1978. A model of Phobos. *Icarus* 33:116–140.

Veverka, J. 1978. The surfaces of Phobos and Deimos. *Vistas in Astronomy* 22:163–192.

Veverka, J., and Thomas, P. 1979. Phobos and Deimos: A preview of what asteroids are like? In *Asteroids*, ed. T. Gehrels (Tucson: Univ. of Arizona Press), pp. 628–645.

Weast, R., ed. 1983. In *Handbook of Chemistry and Physics*, 64th ed. (Boca Raton, Fl.: CRC Press), pp. B:65–158, C:65–576.

Weiblen, P. W., Murawa, M. J., and Reid, K. J. 1990. Preparation of simulants for lunar surface materials. In *Engineering, Construction, and Operations in Space II: Proc. of Space 90*, vol. 1, eds. S. Johnson and J. Wetzel (New York: American Soc. of Civil Engineers), pp. 98–106.

Windholz, M., ed. 1983. In *The Merck Index*, 10th ed. (Rahway, N. J.: Merck & Co.); see alphabetical listings.

Wszolek, P. C., Simoneit, B. R., and Burlingame, A. L. 1973. Studies of magnetic fines and volatile-rich soils: Possible meteoritic and volcanic contributions to lunar carbon and light element chemistry. In *Proc. Lunar Sci. Conf.* 4, *Geochim.*

Cosmochim. Acta Suppl. 2:1693–1706.

Zellner, B. 1979. Asteroid taxonomy and the distribution of the compositional types. In *Asteroids*, ed. T. Gehrels (Tucson: Univ. of Arizona Press), pp. 783–806.

Zhao, Y., and Shadman, F. 1991. Reaction engineering for materials processing in space: Reduction of ilmenite by hydrogen and carbon monoxide. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 12.

Zinner, E. 1988. Interstellar cloud material in meteorites. In *Meteorites and the Early Solar System*, eds. J. F. Kerridge and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 956–983.

SHORT-PERIOD COMETS

PAUL R. WEISSMAN
Jet Propulsion Laboratory

and

HUMBERTO CAMPINS
University of Florida

The spacecraft flybys of comet Halley in 1986 confirmed Whipple's icy conglomerate model for cometary nuclei and showed that comets are far richer in volatiles than any other class of solar system bodies. Water is the most abundant volatile, comprising roughly 80% of the gas flowing out from the nucleus. CO is next with a content of 15% relative to water, though with at least half of that coming from an extended source in the cometary coma, most likely hydrocarbon dust grains. The detection of large numbers of hydrocarbon grains was one of the more significant discoveries of the Halley flybys, as was the groundbased observation that CN occurs in jets, again suggesting an extended source. Estimates of the total dust-to-gas ratio for Halley range as high as 2:1, indicating that a substantial fraction of the volatiles may be tied up in solid hydrocarbons rather than ices. The role of clathrates (if any) in trapping more volatile ices is not yet understood. If Halley can be taken to be representative of all short-period comets, which to first order we think it is, then the short-period comets provide a significant source of volatiles in near-Earth space. This resource is more difficult to reach dynamically than the near-Earth asteroids, but the high volatile content may justify the additional effort necessary. In addition, there is considerable evidence that at least some fraction of the near-Earth asteroids are extinct cometary nuclei which have evolved into asteroidal orbits, and which may contain significant volatiles buried beneath an insulating lag deposit crust of nonvolatiles. Our knowledge of comets will be greatly enhanced in the near future by the Comet Rendezvous Asteroid Flyby mission now under development by NASA, and by the proposed Rosetta mission, a joint effort by ESA and NASA to return a cometary sample to Earth for detailed analysis.

I. INTRODUCTION

Comets are the most volatile-rich bodies in the solar system. The 1986 spacecraft flybys of comet Halley dramatically confirmed Whipple's (1950) icy conglomerate model for cometary nuclei, in which the nuclei are composed of water and other volatile ices, and interstellar dust grains in an intimate, heterogeneous mixture. In addition, the spacecraft flybys demonstrated the existence of a third major component, hydrocarbon grains, which plays a

major role in coma chemistry and kinetics. Compositional analyses from *in-situ* instruments showed that the outflowing Halley coma materials were even closer to solar composition than the most primitive carbonaceous chondrite meteorites. All of this is consistent with the expectation that comets are icy planetesimals from the outer-planets region or beyond, preserved remnants from the solar nebula that contain a cosmochemical record of that earliest period of solar system history.

Comets are typically classified as either long- or short-period, depending on whether their orbital periods are greater than or less than 200 yr, respectively. The orbits of the long-period (LP) orbits are randomly oriented on the celestial sphere, whereas the short-period (SP) comets are generally confined to direct orbits with inclinations less than $\sim 35°$. Most of the known SP orbits have periods between 5 and 20 yr, while the LP orbits range up to 10^7 yr. Short-period comets are often denoted by the designation "P/" ahead of their names.

The long-period comets have been stored over the history of the solar system in distant orbits in the Oort cloud (Oort 1950), a huge swarm of several times 10^{12} to 10^{13} comets surrounding the planetary system and extending halfway to the nearest stars (for a review of the Oort cloud, see Weissman [1990]). The Oort cloud is a fairly benign environment, providing an effective "cold storage" for the cometary nuclei, though a variety of physical mechanisms have now been identified which slowly process the near-surface layers of the nuclei (Weissman and Stern 1993). Nevertheless, the LP comets are still the most pristine objects in the solar system.

Passages of LP comets through the planetary region occur at random intervals and cannot generally be predicted. In addition, their near-parabolic energies and typically high inclinations make them very difficult targets with which to match orbits. In contrast, the SP comets are in well-determined orbits with modest eccentricities and inclinations. This makes them a possible resource for space developments. Although the SP comets are less accessible energetically than near-Earth asteroids, their rich volatile content may in fact make them a better resource economically for space mining. In addition, the likely possibility that some of the Earth-approaching asteroids may be extinct cometary nuclei means that this volatile-rich resource could be even more accessible than currently thought.

In this chapter, we review the physics and dynamics of SP comets. In particular, we will review the wealth of compositional data gained from the 1986 spacecraft flybys of comet P/Halley 1986 III and supporting groundbased observations. We will also review our knowledge of other SP cometary nuclei and the possible evolution of SP comets to dormant, asteroidal objects. Finally, we discuss several new spacecraft missions that are currently being developed and/or planned to SP comets.

The technologies involved in space mining and resource recovery are discussed in a number of other chapters in this book; we will not consider them here. The reader is referred to those other chapters for a description of

the hardware and methods being contemplated.

II. DYNAMICS OF SHORT-PERIOD COMETS

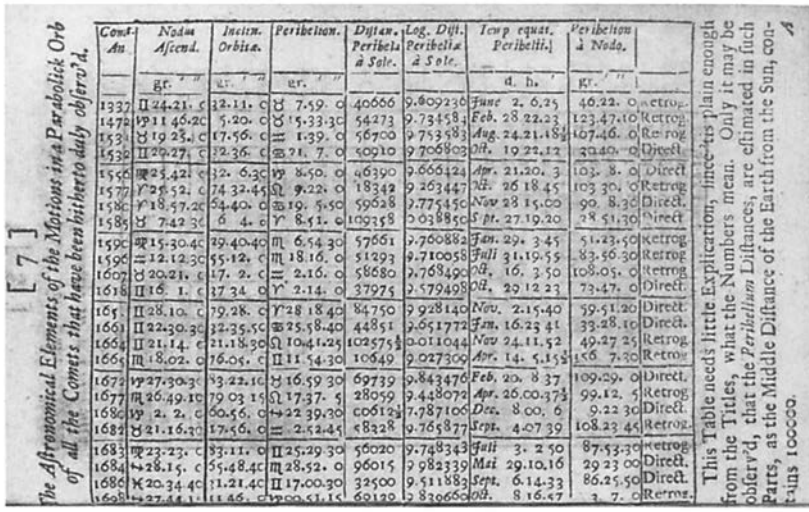
Marsden's (1992) Catalog of Cometary Orbits lists 170 known SP comets, of which 67 have only been observed on one perihelion passage. Of the 67 one-apparition comets, 47 were not yet expected back at the time of publication of the catalog. Marsden also lists eight low-inclination LP comets whose orbits are poorly determined, and which may in fact be SP comets.

The perihelion distances of the known SP comets range from 0.126 AU for comet Machholz to 5.772 AU for comet Schwassmann-Wachmann 1. Orbital periods range from 3.28 yr for comet Encke to 187 yr for comet Wilk. However, 149 of the 170 known SP comets have periods <20 yr. Only 5 of the known SP comets are in retrograde orbits (inclinations >90°); 153 of the 170 have inclinations less than 35°. The most observed object is comet Encke, seen on 55 (of 63) returns since its discovery in 1786.

The best known SP comet is comet Halley which has been seen on 30 consecutive perihelion passages since 240 B. C. Halley is an unusual object as it is far more active than any of the other SP comets, with typical water production rates near perihelion of 10^{30} molecules s^{-1} or more, 1 to 2 orders of magnitude greater than other bright SP comets. One reason for this greater activity is its relatively small perihelion distance of 0.5871 AU. It is also one of only five known SP comets in a retrograde orbit, having an inclination of 162°. In addition, its orbital period of 76 yr makes it among the longer period SP comets, with an aphelion beyond the orbit of Neptune. All of the retrograde SP comets have relatively longer orbital periods, a fact which may be related to their origin (see below).

Edmond Halley was the first to recognize that some comets were periodic and returned at regular intervals to the planetary region. Halley (1705) calculated the orbits of 24 bright comets; his first catalog of cometary orbits is shown in Fig. 1. Halley noticed that the orbits of the comets of 1531, 1607 and 1682 were remarkably similar, and that the appearances were separated by intervals of ~76 yr. He predicted that the comet would return again in 1758. As a result of this successful prediction, the comet now bears his name. The next successful prediction of a cometary return was by the German mathematician Johann Encke, who derived the orbit of a comet discovered by Pons in 1818, and matched it with ones seen in 1786, 1795 and 1805.

The evolution of SP comet orbits is controlled by planetary perturbations, primarily by Jupiter. Virtually all SP comets are in Jupiter-crossing or Jupiter-approaching orbits with the notable exception of comet Encke which has an aphelion distance of only 4.0 AU. A sample of SP comet orbits is shown in Fig. 2. Frequent encounters between the comets and Jupiter lead to a chaotic evolution of their orbits, often with large changes in their perihelion distances and orbital periods. Typical dynamical lifetimes of SP comets among the



[7]

The Astronomical Elements of the Motions in a Parabolick Orb of all the Comets that have been hitherto duly observ'd.

Cons. An	Nodus Ascend.	Inclin. Orbitæ.	Perihelion.	Distan. Perihel. à Sole.	Log. Dist. Perihelii à Sole.	Temp equat. Perihelii.	Perihelion à Nodo.	
	gr.	gr.	gr.			d. h.	gr.	
1337	♊ 24.21. 5	32.11. 0	♌ 7.59. 0	40666	9.609236	*June* 2. 6.25	46.22. 0	Retrog.
1472	♍ 11 46.20	5.20. 0	♎ 15.33.30	54273	9.734581	*Feb.* 28 22.23	123.47.10	Retrog.
1531	♉ 19 23. 0	17.56. 0	♊ 1.39. 0	56700	9.753583	*Aug.* 24.21.18½	107.46. 0	Retrog.
1536	♊ 29.27. 0	32.36. 0	♋ 21. 7. 0	50910	9.706803	*Oct.* 19 22.12	30.40. 0	Direct.
1556	♐ 25.42. 0	32. 6.30	♍ 8.50. 0	46390	9.666424	*Apr.* 21.20. 3	103. 8. 0	Direct.
1577	♈ 25.52. 0	74 32.45	♌ 9.22. 0	18342	9.263447	*Nov.* 26 18.45	103 30. 0	Retrog.
1580	♈ 18.57.20	54.40. 0	♎ 19. 5.50	59628	9.775450	*Nov* 28 15.00	90. 8.30	Direct.
1585	♉ 7.42.30	6 4. 0	♈ 8.51. 0	109358	0.038850	*Sept.* 27.19.20	28 51.30	Direct.
1590	♐ 15.30.40	29.40.40	♏ 6.54.30	57661	9.760882	*Jan.* 29. 3.45	51.23.50	Retrog.
1596	♊ 12.12.30	55.12. 0	♏ 18.16. 0	51293	9.710058	*Juli* 31.19.55	83.56.30	Retrog.
1607	♉ 20.21. 0	17. 2. 0	♊ 2.16. 0	58680	9.768490	*Oct.* 16. 3.50	108.05. 0	Retrog.
1618	♊ 16. 1. 0	17 34. 0	♈ 2.14. 0	37975	9.579498	*Oct.* 29 12 23	73.47. 0	Direct.
165.	♊ 28.10. 0	79.28. 0	♈ 28 18 40	84750	9 928140	*Nov.* 2.15.40	59.51.20	Direct.
1661	♊ 22.30.30	32.35.50	♋ 25.58.40	44851	9.651772	*Jan.* 16.23 41	33.28.10	Direct.
1664	♊ 21.14. 0	21.18.30	♌ 10.41.25	102575½	0.011044	*Nov* 24.11.52	49.27 25	Retrog.
1665	♏ 18.02. 0	76.05. 0	♊ 11 54.30	10649	9.027309	*Apr.* 14. 5.15½	156 7.30	Retrog.
1672	♈ 27.30.30	33.22.10	♉ 16.59 30	69739	9.843476	*Feb.* 20. 8 37	109.29. 0	Direct.
1677	♏ 26.49.10	79 03 15	♌ 17.37. 5	28059	9.448072	*Apr.* 26.00.37½	99.12. 5	Retrog.
1680	♈ 2. 2. 0	60.56. 0	♑ 22 39.30	006123	7.787106	*Dec.* 8.00. 6	9.22 30	Direct.
1682	♉ 21.16.30	17.56. 0	♊ 2.52.45	58328	9.765877	*Sept.* 4.07 39	108.23 45	Retrog.
1683	♐ 23.23. 0	83.11. 0	♊ 25.29 30	56020	9.748343	*Juli* 3. 2 50	87.53.30	Retrog.
1684	♑ 28.15. 0	55.48.40	♏ 28.52. 0	96015	9.982339	*Mai* 29.10.16	29 23 00	Direct.
1686	♐ 20.34 40	71.21.40	♊ 17.00.30	32500	9.511883	*Sept.* 6.14.33	86.25.50	Direct.
1698	♑ 27.44 1	11.46. 0	♑ 00.51.15	69129	9.839660	*Oct.* 8 16.47	3. 7. 0	Retrog.

This Table needs little Explication, since it is plain enough from the Titles, what the Numbers mean. Only it may be observ'd, that the *Perihelium* Distances, are estimated in such Parts, as the Middle Distance of the Earth from the Sun, contains 100000.

Figure 1. Original catalog of 24 cometary orbits from Halley's *A Synopsis of the Astronomy of Comets* (1705). The comets of 1531, 1607 and 1682 are the return of Halley's comet, the first one to be recognized as periodic.

inner planets are $\sim 10^4$ to 10^6 yr (Wetherill 1975), with dynamical ejection by Jupiter being the most likely end-state.

One of the more extreme examples of the chaotic changes possible for SP comet orbits was comet Lexell which was discovered when it approached to within 0.015 AU (2.2×10^6 km) of the Earth in 1770, the closest cometary encounter with Earth on record. Integration of the orbit backward in time showed that the comet had been thrown into its Earth-crossing orbit by a close encounter with Jupiter in 1767, at a Jovicentric distance of 0.020 AU. Prior to the Jupiter encounter, Lexell was in a low eccentricity orbit with a perihelion distance of 3 AU, and had never been observed. The new orbit had a period of 5.6 yr which was close to half of Jupiter's period, leading to a second, even closer Jupiter encounter in 1779 at 0.0015 AU (about 3 Jupiter radii, or half the distance of Io's orbit). The comet was thrown into a long-period ellipse with perihelion just outside the orbit of Jupiter, aphelion of about 80 AU, and a period of over 260 yr. A somewhat similar evolutionary path was experienced by comet Oterma earlier this century, involving a 3:2 resonance with Jupiter, though that comet never became an Earth-crosser (Kazimirchak-Polonskaya 1972).

In addition to planetary perturbations, nongravitational forces play an important role in the evolution of cometary orbits. Nongravitational forces result from the jetting of volatile ices on the sunlit hemisphere of the cometary nucleus, creating a net thrust on the nucleus, $\sim 10^{-4}$ times the solar gravitational force for typical LP comets, and $\sim 10^{-5}$ for SP comets. Observations of SP comets show that they often return either earlier or later than predicted.

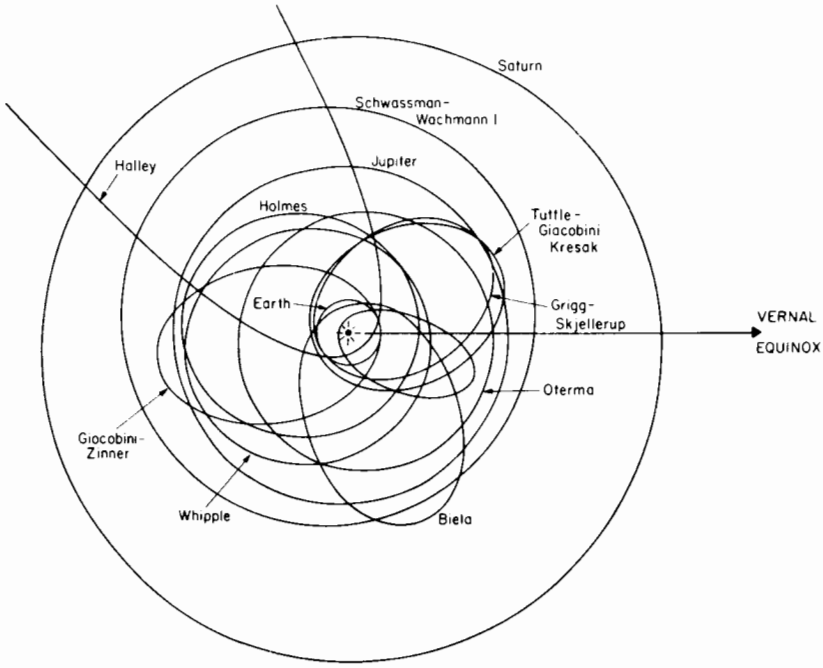


Figure 2. Representative orbits of some SP comets, as compared to the orbits of the Earth, Jupiter and Saturn. The smallest ellipse passing closest to the Sun is the orbit of periodic comet Encke.

For example, comet Halley typically returns about four days late on each orbit, as compared with that predicted by a purely gravitational orbit. The ability to explain nongravitational forces was an important factor in the early acceptance of Whipple's (1950) icy-conglomerate model for cometary nuclei.

The existence and nature of the nongravitational forces were definitively established in a classic series of papers by Marsden and colleagues, culminating with Marsden et al. (1973). They showed that the forces on the comets could be explained by a model based on water-ice sublimation rates versus solar distance given by

$$g(r) = \alpha(r/r_o)^{-2.15}[1 + (r/r_o)^{5.093}]^{-4.6142} \quad (1)$$

where α is a normalization constant = 0.111262, so that $g(1) = 1$, and $r_o = 2.808$ AU, the approximate solar distance at which water sublimation rates fall sharply towards zero. The function behaves as an approximately inverse square law inside of 2 AU (where most of the incident solar energy goes into sublimating surface ices) and then falls off sharply beyond that distance. The form of Eq. (1) was derived by A. H. Delsemme based on estimates of water ice sublimation rates versus heliocentric distance.

Using this model, Marsden et al. (1973) were able to reduce significantly orbit residuals for 34 SP comets and for 7 LP comets. Nongravitational

forces are more difficult to estimate in LP comets because they are only observed on one apparition. However, in the cases where they have been successfully applied to LP comets, it has been possible to show that some apparently hyperbolic comets are in fact Oort cloud members, neglect of the nongravitational forces making the orbits appear more eccentric than they actually were. For the SP comets, the nongravitational force model has made it possible to predict their returns more accurately, and to reduce ephemeris errors for targeting of spacecraft missions to comets.

An improvement on Marsden's model was recently provided by Yeomans and Chodas (1989) who showed that improved orbit fits could be obtained by offsetting the peak of the nongravitational forces (as given by Eq. [1]) from perihelion. They showed that these empirically derived offsets were well correlated with the observed pre- versus post-perihelion activity of each comet.

The chaotic nature of SP comet dynamics occasionally makes it difficult to recognize the return of a previously observed comet, because of changes to its orbit from strong planetary perturbations, usually a close approach (<1 AU) to Jupiter. However, an approximate constant of the motion in the 3-body problem (Sun-Jupiter-comet) is the Tisserand criterion

$$T = a_J/a + 2\sqrt{(a/a_J)(1-e^2)}\cos i \qquad (2)$$

where a, e and i are the semimajor axis, eccentricity and inclination, respectively, of the comet's orbit, and a_J is the semimajor axis of Jupiter's orbit (or the perturbing planet if it is not Jupiter). SP comets evolve along lines of constant T in (a, e, i) space and generally have values of $T \leq 3$.

Although Jupiter is the dominant perturber of the SP comets, the other giant planets also play a role in their chaotic orbital evolution (Lowery 1973). This was further expanded upon by Everhart (1978) who constructed a Monte Carlo simulation model of SP comets as perturbed by the giant planets. By far the most complete study of SP comet dynamics has been a ±400 yr integration of the orbits of all known SP comets by Carusi et al. (1985). They confirmed much of the previous work and found many cases of temporary capture of comets into orbital resonances. Perhaps most interesting was their finding that many SP comets are temporarily captured into distant satellite orbits around Jupiter. Unfortunately, Carusi et al. did not include nongravitational forces in their calculations. Thus, their results are interesting with regard to the nature and statistics of SP comet dynamics, but do not accurately reproduce or predict the dynamical histories of specific comets.

Two unusual SP comets which avoid Jupiter encounters are Arend-Rigaux and Neujmin 1. Marsden (1970) studied the motion of these comets and showed that they were trapped in resonances which kept their minimum approach distances to Jupiter >0.9 AU for almost 10^3 yr. Interestingly, these two comets are among the least active SP comets known and have often been suggested as transitional to dormant, asteroidal-appearing objects. It appears

that the orbital resonances have greatly lengthened the dynamical lifetimes of these two comets, allowing physical rather than dynamical mechanisms to determine their eventual end-states (see additional discussion below).

Twenty-five of the known SP comets are in Earth-crossing orbits. Of those 25, 15 have been seen more than once. Of the remaining 10 one-apparition comets, 5 are not yet expected back. Some comets, P/Giacobini-Zinner, for example, were Earth-crossing in the recent past but have been perturbed to larger perihelia by Jupiter in the relatively brief period that they have been under observation. Comet Lexell (described above) was Earth-crossing for two orbits in the 18th century, but is now in an orbit beyond Jupiter. Another interesting object is comet Biela which was observed on several perihelion passages. It was seen to have split in 1846, returned as a double comet in 1852, and was never observed again. An additional 26 SP comets have perihelia $1.0<q<1.3$ AU, 17 of them having been observed on more than one apparition.

The Earth-crossing SP comets have a mean encounter velocity with the Earth of 37.5 km s^{-1}, though the most probable velocity is 28.9 km s^{-1}, the mean being heavily biased by Halley and a few other comets (Weissman 1982,1991*a*). The mean Earth impact probability for the known SP comets is 6.6×10^{-9} per perihelion passage, or 8.2×10^{-10} yr^{-1}. Impacts of comets and asteroids on the Earth have been proposed as initiators of global biological extinction events observed in the fossil record, in particular the Cretaceous-Tertiary extinction 65 Myr ago (Alvarez et al. 1980).

All-Sky maps of data from the IRAS satellite showed peculiar linear featuresmainsection which were identified by Sykes et al. (1986) as dust trails in the orbits of SP comets. These trails are the precursors of cometary meteor streams, and are composed of 0.1 to 10 mm particles moving very slowly away from the parent comet. Sykes and Walker (1992) showed that 8 of 17 identified trails were associated with known SP comets, and that all of the identified comets were near perihelion at the time of the IRAS observations. This led them to infer that all SP comets had such dust trails, but that they were only observable (by IRAS) when they were relatively close to the Sun. No trails were associated with any of the known asteroids, or with any LP comets.

The existence of 9 additional dust trails that do not have any known comets associated with them suggests that there are still a large number of undiscovered SP comets in the inner solar system. Based on the statistics of Sykes and Walker (1992), the total number of SP comets is at least twice the currently discovered number of 170, or possibly more. Incompleteness in comet discovery was studied extensively by Everhart (1967) who showed that only about 5% of all LP comets brighter than absolute magnitude $H_{10} = 11$ and with perihelia <4 AU are discovered as they pass through the planetary region.[a] Discovery opportunities are obviously greater for SP comets because

[a] The cometary absolute magnitude H_{10} assumes a cometary brightness law varying

of their frequent returns. However, there are no detailed calculations of their total population that take all these effects into consideration.

Discovery rates for SP comets have been fairly constant at about 5 or 6 new SP comets per year throughout the late 1980s. Most of the new SP comets have perihelia beyond 1.5 AU, some as large as 3 or 4 AU, though several new Earth-crossers have also been discovered in the past decade (Marsden 1992).

III. ORIGIN OF SHORT-PERIOD COMETS

It has generally been thought that the SP comets are LP comets which have random-walked to small semimajor axis orbits as a result of perturbations by Jupiter and the other planets. Capture of comets to short-period in a single planetary encounter is highly unlikely (Newton 1893) and could not explain the observed number of SP comets. Planetary perturbations are larger for direct, low-inclination orbits and it was suggested that this acted as a selection mechanism for producing the direct, low inclination SP comet population. However, evolution from an Oort cloud orbit to a short-period one would be expected to take approximately 400 returns (Weissman 1979), and it is most likely that a comet would be ejected over that time, or destroyed by one of several poorly understood physical mechanisms (e.g., disruption, sublimation).

A dynamical mechanism which results in improved efficiency was proposed by Everhart (1972) who suggested that comets with perihelia among the outer planets random walk in orbital energy without ever coming close to the Sun, and then are dumped into small perihelia orbits late in their dynamical evolution. In this manner, the SP comets can be spared some of the physical loss associated with solar heating or passage through the more densely populated regions of the planetary system. This mechanism also greatly increases the planetary system's cross-section for capturing LP comets to SP orbits, thus better supplying the observed number of SP comets.

Estimates of the number of SP comets produced by planetary perturbation from the Oort cloud have varied considerably (Joss 1973; Delsemme 1973), with some suggestion that the above mechanisms can not produce the observed number of SP comets. An alternative suggested source of SP comets is a ring of comets beyond the orbit of Neptune (Kuiper 1951; Whipple 1964; Fernandez 1980), which may be up to 300 times more dynamically efficient than repeated perturbation of LP comets from the Oort cloud. Fernandez (1980) pointed out that some massive comets $\sim 10^3$ km in diameter would be required in the distant comet ring, to perturb other comets slowly back into the planetary region, providing the SP comet flux.

---

as the inverse fourth power of heliocentric distance and the inverse square of geocentric distance: $V = H_{10} + 10 \log r + 5 \log \Delta$, fit to the total observed brightness of the cometary coma.

Duncan et al. (1988) compared the two possible sources and found that LP comets tended to preserve their inclinations as they evolved inward to SP orbits. Thus, if SP comets were initially LP comets from the Oort cloud, far more high inclination SP comet orbits would be expected. Duncan et al. (1988) showed that the observed inclination distribution was more consistent with a low-inclination trans-Neptunian population of comets as the source of the SP comets. They called this population the Kuiper belt in honor of Gerard Kuiper who first suggested its existence in 1951. Subsequently, Torbett (1989) showed that planetary perturbations would lead to chaotic motion in a disk of comets beyond Neptune, throwing those with perihelia near 30 AU into Neptune-crossing orbits in only 10^7 yr. However, Torbett and Smoluchowski (1990) showed that the motion would not be chaotic for Kuiper belt orbits with perihelia >45 AU. Levison (1991) confirmed this by studying the evolution of orbits with perihelia >30 AU and aphelia <100 AU, and showed that many can persist over the age of the solar system, in particular those with initially low eccentricity orbits beyond 45 AU.

A number of questions have been raised with regard to the Kuiper belt scenario (Stagg and Bailey 1989; Bailey and Stagg 1990). First, physical loss mechanisms (e.g., splitting, loss of all volatiles) might remove high inclination LP comets during their longer evolution inward from the Oort cloud to SP comet orbits (longer because of the smaller average planetary perturbations for high inclination orbits), leading to a predominantly low inclination population of observed SP comets. This would remove the objection raised by Duncan et al. (1988) to this source for the SP comets. Second, it was suggested that the combined mechanisms (LP comet evolution from the Oort cloud plus the Kuiper belt) might produce too many SP comets, and that this possibly implied a lower than estimated population for the Oort cloud or the Kuiper belt, or both. Finally, it was noted that Duncan et al. (1988) had increased the planetary masses by a factor of 40 to speed the numerical integrations, a common technique in celestial mechanics simulations but one that could lead to erroneous results. However, Quinn et al. (1991) repeated the integrations with the mass enhancement factor reduced to 10 and obtained identical results for the Kuiper belt comet evolution. In addition, Wetherill (1991) was able to recreate the same results with no mass enhancement, using a simpler Öpik-type integrator for simulating the cometary evolution.

The correlation between high inclinations and relatively longer orbital periods for some SP comets, such as Halley, suggests that these latter objects might be evolved LP comets from the Oort cloud, whereas the majority of SP comets, those in shorter period, low inclination orbits, are from the Kuiper belt. It would be interesting if these two groups exhibited systematic differences in either behavior or composition. No such differences have been noted as yet. Given expectations of a low temperature gradient in the outer solar nebula, it is not clear that recognizable differences would actually occur, though they have been suggested (Delsemme 1991).

Duncan et al. (1988) estimated that 0.02 Earth masses ($M_\oplus$) of comets

are required in the Kuiper belt to maintain the current population of SP comets. Duncan et al. used an average cometary nucleus mass of 3.2×10^{17} g, implying a population of 3.8×10^8 objects. However, using the more carefully determined average nucleus mass of 3.8×10^{16} g found by Weissman (1991*a*), the same number of Kuiper belt comets would have a total mass of $\sim 0.0025\ M_\oplus$.

Observational searches for Kuiper belt comets have so far been negative. Levison and Duncan (1990) set an upper limit of less than one object brighter than magnitude $V = 22.5$ per square degree based on a search of 4.9 square degrees of sky near the ecliptic (a visual magnitude of 22.5 corresponds to a 60 km radius cometary nucleus at 50 AU from the Sun, assuming an albedo of 5%). This translates to $N \leq 10^{10}$ objects if the Kuiper belt comets are in orbits between 35 and 60 AU from the Sun. Hamid et al. (1968) and Yeomans (1986) set an upper limit on the unknown mass just beyond the orbit of Neptune of $\sim 1\ M_\oplus$, based on a failure to observe gravitational perturbations on the orbit of comet Halley. An earlier limit of $5\ M_\oplus$ was found by Anderson and Standish (1986) based on tracking of the Pioneer 10 spacecraft. Future tracking of the Voyager and Pioneer spacecraft can be expected to further refine these limits, and may provide the first confirmation of the existence of material in the Kuiper belt.

IV. COMETARY NUCLEI: OBSERVATIONS AND PHYSICAL PROPERTIES

The direct observation of cometary nuclei using Earth-based telescopes is very difficult. When a comet is near the Sun, the nucleus is almost always masked by the bright gas and dust coma. On the other hand, when a comet is at large heliocentric distances and there is presumably little or no coma, the nucleus is often too faint to be observed in detail. This situation is changing as observational detector technology improves. Nevertheless, the basic physical properties of cometary nuclei—their size, shape, albedo, color and rotation period—have only been measured for a few comets to date. Table I presents known values for 8 comets, along with references to the original observations. The most detailed picture of a cometary nucleus has resulted primarily from the spacecraft studies of comet Halley. The nuclei of the other 7 comets in Table I were all observed from the ground.

Admittedly, this sample is very limited and one must be cautious when generalizing because of the possibility of large selection effects. In particular, the intensive study of the 1986 apparition of comet Halley (Fig. 3), including the only spacecraft missions ever to encounter a cometary nucleus, could bias our observations of other comets, or lead to generalizations that are not really supported. However, all indications are that Halley in fact is, to first order, a fairly typical comet (Weissman 1987*a*).

Only eight years ago, none of the measurements in Table I existed. Now, certain trends appear to be emerging. General properties that apply

Figure 3. Composite photograph of the nucleus of comet Halley as recorded by the Giotto spacecraft March 14, 1986. The Sun is at left. The spacecraft approached the nucleus from the dark side, and is viewing the sunrise terminator. Discrete active areas and dust jets are visible on the sunlit hemisphere. The dark, irregular nucleus is silhouetted against the bright dust coma of the comet. The nucleus is approximately 16 km long by 8 km wide. Photo courtesy of H. U. Keller, Max-Planck-Institut für Aeronomie, Lindau.

to all nuclei of SP comets include: very low reflectivities, i.e., geometric albedos of 0.02 to 0.05 being typical; elongated shapes, with axial ratios generally >1.5; moderately slow rotation periods, typically about 10 hr or longer; and low gas production per unit surface area. There is a range of effective radii, though a typical value would be near 5 km. Some of these characteristics do not apply, or have not been measured for object 2060 Chiron (originally classified as an outer solar system "asteroid"; see Sec. VII), comet Schwassmann-Wachmann 1, and comet IRAS-Araki-Alcock 1983 VII (an LP comet). These three objects are dynamically distinct from typical SP comets and not as relevant to the issues of near-Earth resources. However, they could be precursors to Earth-approaching SP comets (see discussion below).

Visible albedos are directly measured only for S comets Neujmin 1, Arend-Rigaux and Tempel 2 (from simultaneous photometry and radiometry), Schwassmann-Wachmann 1 (photometry and radiometry but not simultaneous), and Halley (spaceborne photometry of the spatially resolved nucleus). All five nuclei are dark and the four best determined values are among the darkest values for any solar system objects. Constraints on the albedo and

TABLE I

Well-Studied Cometary Nuclei*

Comet	R_{eff}^{a} (km)	Axial Ratio	p_v	Color	P_{rot} (hr)	$Q(r)^{d}$ (s^{-1})	r (AU)	Active fraction–% Isothermal	Pole-on
Halley (references)	6.0[b] (1,2,3)	2.0[b]	0.04 (4,5)	neutral, red (6,25)	53?, 89?, 177? (7)	6×10^{29}	0.8	30.0[f]	7.0[f]
Arend-Rigaux	5.2 (8,9,10,11)	> 1.6 (8,12,13)	0.03 (8,9)	neutral, red (8,13)	13.5 (8,12,13)	2×10^{26} (8)	1.58	0.08	0.016
Neujmin 1	10.4 (14)	> 1.65 (12,13)	0.02 (14)	very red (14,15)	12.7 (12,13)	2×10^{26} (14)	1.68	0.1	0.023
Schwassmann-Wachmann 1	40.? (16)	small?	0.13? (16)	red in near IR (26)					
Tempel 2	5.6 (23)	1.9 (13,22,23)	0.02 (23)	very red (13,17)	8.9 (13,22,23)	2×10^{27} (24)	1.71	0.9	0.13
Encke	<2.2, 1.? (18,19)	> 2.0?[c] (20)			22.4?[c] (20)	6×10^{28}	0.76	17.0	3.7

IRAS-Araki-Alcock	4.0 ? (20,21)			48–72? (30)	2×10^{28}	1.03	3.0	0.8
Chiron	<186., 84.? (27,28)	>0.027, 0.14? (27,28)	neutral (28)	5.92 (29)				

* Table from Weissman et al. (1989) and expanded to include object 2060 Chiron.

[a] $R_{eff} = \sqrt{(ab)}$ where a and b are the projected semi-axes at maximum light.

[b] Actual Halley nucleus dimensions: $16 \times 8 \times 7$ km.

[c] See also discussion by Sekanina (1990).

[d] Production of H_2O based on Festou's (1981) vectorial model using observational data given in the references.

[e] Fraction of surface which must be "active" for vaporization of water in equilibrium to provide the observed gas assuming 5% albedo and either (a) incident heat distributed uniformly around an isothermal nucleus, or (b) incident heat in local equilibrium for a nucleus either pole-on to the Sun, or rotating very slowly.

[f] Directly observed value from Giotto (Keller et al. 1986): 10–15% of total surface or 20–30% of sunlit hemisphere.

Table I references:

1. Möhlmann et al. 1986
2. Wilhelm et al. 1986
3. Sagdeev et al. 1986*a*
4. Sagdeev et al. 1986*b*
5. Delamere et al. 1986
6. Belton et al. 1987
7. Sekanina 1990
8. Millis et al. 1988
9. Veeder et al. 1987
10. Brooke and Knacke 1986
11. Tokunaga and Hanner 1985
12. Wisniewski et al. 1986
13. Jewitt and Meech 1988
14. Campins et al. 1987
15. Hartmann et al. 1987
16. Cruikshank and Brown 1983
17. Spinrad et al. 1979
18. Campins 1988
19. Kamoun et al. 1982
20. Goldstein et al. 1984
21. Hanner et al. 1985
22. Wisniewski 1988
23. A'Hearn et al. 1988
24. A'Hearn et al. 1989
25. Thomas and Keller 1989
26. Hartmann et al. 1982
27. Sykes and Walker 1991
28. Lebofsky et al. 1984
29. Bus et al. 1989
30. Harmon et al. 1989

radius of 2060 Chiron are discussed below.

The spectral reflectance of cometary nuclei, i.e., the variation of albedo with wavelength, is poorly known because the observations are very difficult. This is due to the difficulty of separating out the contribution from scattering and thermal emission by dust, and from weak emission by gas in the coma. Overall the cometary nuclei are reddish but there also appear to be significant differences in color among the comets.

The spectral reflectances for 5 cometary nuclei are shown in Fig. 4 (from Weissman et al. 1989). All the observations are normalized to unity at 0.56 μm, and broadband infrared measurements have been converted to reflectances using the calibration of solar colors by Campins et al. (1985). Visual colors have been converted to reflectance using a variety of methods. Although these reflectance spectra are from several sources and were obtained under a variety of circumstances, most of the data are thought to refer to the nuclei. This is particularly important because recent studies of comet Tempel 2 show that the color of the dust coma is significantly different from the color of the nucleus (Jewitt and Luu 1989; A'Hearn et al. 1989). The Halley reflectance observations of Cruikshank et al. (1985) likely include significant contamination by the dust coma. Although these data were obtained when the comet was still far from the Sun, Halley's nucleus already appeared to be active.

The spectral reflectances in Fig. 4 vary markedly from one comet to another. In particular, the nuclei of Neujmin 1 and Tempel 2 are significantly redder than the other three nuclei. Furthermore, comparison with the reflectance of near-Earth asteroids (see Weissman et al. 1989, Fig. 6; McFadden et al. 1989, Fig. 2) shows that the cometary reflectance spectra do not match well any single asteroid type. There are even features in some of the cometary spectra that do not appear in any asteroid spectra. For example, observations of comets Schwassmann-Wachmann 1 (Cochran et al. 1982) and Tempel 2 (Spinrad et al. 1979; A'Hearn et al. 1989) suggest that the reflectance increases shortward of 0.4 μm. No asteroid, with the possible exception of 2201 Oljato (an extinct comet candidate; McFadden et al. 1984), shows this effect. In fact, most asteroid types exhibit sharply decreasing reflectance shortward of 0.4 μm.

As with albedo, radii have been directly measured for the same five comets: Schwassmann-Wachmann 1, Neujmin 1, Arend-Rigaux, Tempel 2 and Halley. If we assume that all cometary nuclei have low albedos (comparable to these five), then the infrared radiometry alone is sufficient to determine the size of any nucleus that can be separated from its associated coma. Comets Encke and IRAS-Araki-Alcock were observed at thermal infrared wavelengths, although in the case of Encke the radiometric size estimate is only an upper limit (see references in Table I). The radius and albedo for comet Schwassmann-Wachmann 1 (Cruikshank and Brown 1983) are somewhat questionable. Although it was thought that the observations were of a bare nucleus, Jewitt and Luu (1990) have since shown that this comet is

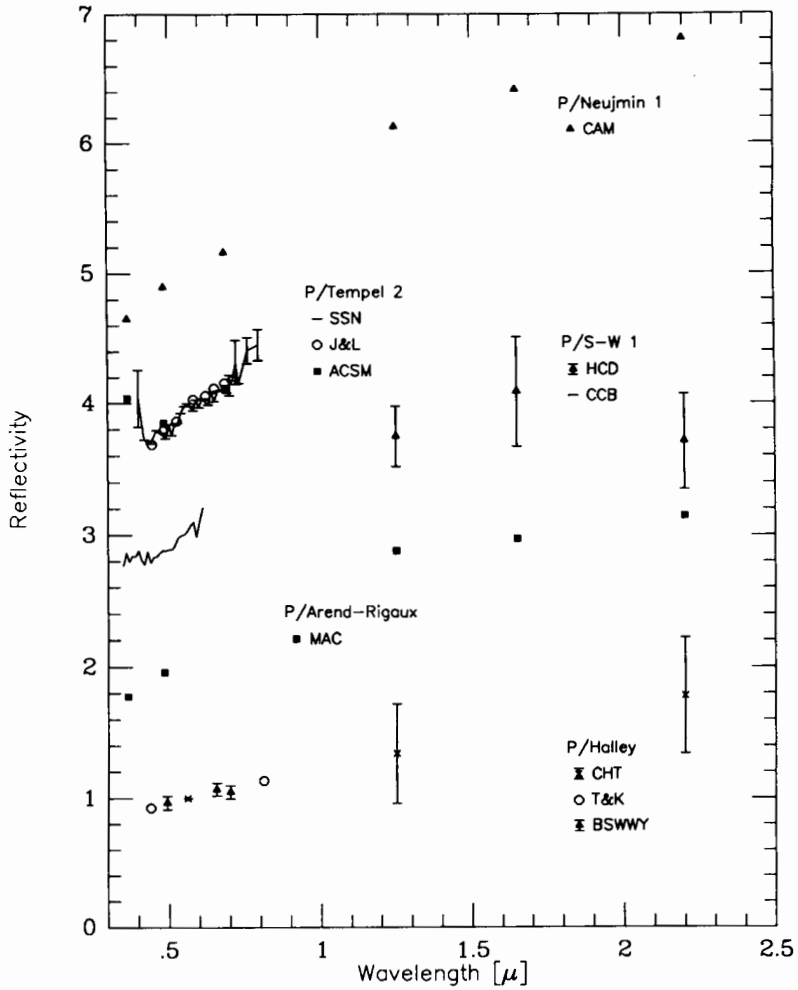


Figure 4. Spectral reflectance of cometary nuclei; the data is normalized to unity at 0.56 μm. Data for Arend-Rigaux, Schwassmann-Wachmann 1, Tempel 2 and Neujmin 1 have been offset vertically by values of 1.0, 2.0, 3.0 and 4.0, respectively. References for the data: Neujmin 1: CAM = Campins et al. (1987); Tempel 2: ACSM = A'Hearn et al. (1989), SSN = Spinrad et al. (1979), J&L = Jewitt and Luu (1989); Schwassmann-Wachmann 1: CCB = adapted from Cochran et al. (1982), HCD = Hartmann et al. (1982); Arend-Rigaux: MAC = Millis et al. (1988); Halley: BSWWY = Belton et al. (1987), T&K = Thomas and Keller (1989), CHT = Cruikshank et al. (1985).

always active, even during its apparently "quiescent" state.

Sykes and Walker (1991) set an upper limit to the diameter of Chiron of 372 km and a lower limit to the geometric albedo of 0.027, based on the nondetection of Chiron at 60 μm by the IRAS spacecraft, and using

the "isothermal latitude" thermal model (ILM). Lebofsky et al. (1984) used the standard thermal model (modified by the lower value of the beaming parameter in the new standard thermal model of Lebofsky and Spencer 1989) to determine a diameter of 167 km and albedo of 0.14 for Chiron. However, thermal modeling of comet Halley at large heliocentric distances by Weissman and Kieffer (1981) clearly shows that the ILM model is more applicable to icy bodies in the outer solar system (see their Fig. 1). Therefore, the constraints by Sykes and Walker appear to be more significant. In any case, if one assumes a "typical" albedo of 5 to 10%, the diameter of Chiron would be in the 300 to 200 km range. Thus, Chiron may be the largest known cometary nucleus.

Radar observations have made interesting contributions to this field of study and have the potential to become a very powerful diagnostic tool. Radar detections of the nuclei of comets Encke and IRAS-Araki-Alcock yield estimates of the radar cross-section times the radar albedo. These radar size estimates are consistent with the radiometric ones. In the case of comet IRAS-Araki-Alcock, the radiometric diameter was used to constrain the radar reflectivity to that of moderately packed snow. Moreover, the Doppler bandwidth of the radar data indicates a rotation period of 2 to 3 days (Harmon et al. 1989). In addition, a broadband radar echo, indicative of a cloud of relatively large particles, has been detected from comets IRAS-Araki-Alcock (Shapiro et al. 1983; Goldstein et al. 1984; Harmon et al. 1989) and Halley (Campbell et al. 1989). The size of these particles is relatively well constrained between one and a few centimeters. The existence of these large particle clouds, which appear to be the predecessors to dust trails and meteor streams, should be considered both as a potential source of material and as a hazard to spacecraft.

Axial ratios of cometary nuclei have been derived primarily from the amplitude of photometric lightcurves except for comet Halley where they are known by direct imaging. In most cases, the observed amplitude of the photometric variations yield a lower limit to the true axial ratio. However, in some cases, the details of the geometry between the rotation axis, and the directions to the Earth and the Sun, are such that the photometric variations are not just due to changes in cross-section but are also the result of variations in solar illumination (Sekanina 1991, see his Fig. 5). Hence, the photometric lightcurves of cometary nuclei must be interpreted with this limitation in mind, and estimates of the actual axial ratios will only be possible for extensively observed nuclei such as comet Tempel 2 (Sekanina 1991).

In a recent review, A'Hearn (1988) gives a complete discussion of the different methods used for the determination of rotational periods for comets. In this chapter we limit ourselves to a brief discussion of the periods given in Table I. These values were determined directly from photometric variations of the nucleus brightness in all cases, except for Halley and IRAS-Araki-Alcock. A period corresponding to a double-peaked lightcurve has been adopted. Interestingly enough, comet Halley's period is the most controversial. The values of 2.2, 3.7 and 7.4 days listed have been debated extensively, and it is

still not clear which is the correct one. It is curious that Halley's rotational period appears to be considerably longer than those of the other SP comets. However, the limited sample prevents us from commenting further on this difference.

We also consider the question of the gas production rate of cometary nuclei and a comparison with the values expected from equilibrium sublimation of ices. Here, we follow the discussion in Weissman et al. (1989). The water production rate for each comet is derived from observations of OH in the ultraviolet. This production rate can be compared with the expected water ice sublimation rate to determine the fraction of the nucleus surface that must be active. Because the sublimation rate of an ice surface depends critically on the detailed geometry, two simple, extreme examples are considered. The first case, which minimizes the gas production rate per unit area, is an isothermal (fast rotator) nucleus. In the second case, the opposite extreme, sublimation occurs from an area around the subsolar point on a nonrotating nucleus, or from an area around the sunward pole of a slowly rotating, highly inclined nucleus. In both cases, the assumed albedo is 0.05. Dividing the observed water production rate by the equilibrium sublimation rate yields the effective active area for each case. Dividing this active area by the surface area of the appropriate spheroid (assumed to be prolate) yields the fraction of the nucleus surface that must be active, as given in Table I.

These values are all uncertain by a factor of 2 or more, due to long-and short-term temporal variations in observed cometary activity and due to model dependence. The value for Halley refers to the time of the Giotto encounter; the range of active fractions includes the result of direct imaging by Giotto that 20 to 30% of the sunlit surface (10 to 15% of the total surface) was active at that time. More detailed thermal models of Halley (Weissman 1987b) show that the simple calculation used for the Table I entries yields appropriate values. On the other hand, Sekanina's (1991) model for Tempel 2 yields an active fraction of 3.8%; considerably larger than the value in Table I. This difference is likely not a fundamental one; the detailed topography assumed in Sekanina's model accounts for the discrepancy. He models the active area as three vent-like depressions with outgassing occurring only from the sunlit floor of these vents; this allows for larger areas of exposed ices that are only active a fraction of the time.

All the nuclei that have been studied appear to have most of their surface area covered by an inert crust (see discussion below). The active fraction of Halley is significantly larger than that of all others; only comet Encke may have an active fraction comparable to that of Halley, but the large uncertainty in Encke's size prevents a proper comparison.

These observations (particularly the Giotto images of the Halley nucleus) are consistent with an inert crust covering an ever increasing fraction of the surface as the comets age, rather than building up uniformly over the whole nucleus surface. As a larger fraction of the nucleus is covered by the crust, the activity is reduced to the very low levels observed in comets Neujmin 1 and

TABLE II

Abundance of Probable Parent Molecules in the Coma of Comet Halley*

Molecule	Relative Abundance	Comments
Comet P/Halley		
H_2O	100	Remote and *in-situ* detections[1]
CO	~ 7	Direct (native) source[2]
	~ 8	Distributed source[2]
H_2CO	0–5	Variable[3]
CO_2	3	Infrared (Vega 1 IKS)
CH_4	<0.2–1.2	Groundbased infrared[4]
	0–2	Giotto IMS[5]
NH_3	0.1–0.3	Variable.[6] Based on NH_2
	1–2	Giotto NMS[7]
HCN	0.1	Variable. Groundbased radio
	< 0.02	Giotto IMS
N_2	~ 0.02	Groundbased N_2^+ emission
SO_2	< 0.002	Ultraviolet (IUE)
H_2S	—	Giotto IMS[8]
CH_3OH	~ 1	Giotto NMS and IMS[9]
Other Comets		
CO	20	West (1976 VI)
	2	Bradfield (1979 X)
	1–3	Austin (1990 V)
CH_4	<0.2	Levy (1990 XX)[10]
	1.5–4.5	Wilson (1987 VII)[11]
CH_3OH	1–5	Variable[12]
H_2CO	0.1–0.04	If a parent species[13]
HCN	0.03–0.2	Several comets[14]
H_2S	0.2	Austin (1990 V); Levy (1990XX)[15]
S_2	0.025	IRAS-Araki-Alcock (1983 VII)[16]

* Table adapted from Mumma et al. (1993).

[1] Water was detected directly by infrared spectroscopy (Mumma et al. 1986; Combes et al. 1986) and by mass spectroscopy (cf. Krankowsky et al. 1986).

[2] CO was detected directly at ultraviolet wavelengths (Feldman et al. 1987; Woods et al. 1986,1987) and in neutral mass spectra (Eberhardt et al. 1987). A tentative detection at infrared wavelengths (Combes et al. 1988) provided production rates in agreement with the native source.

[3] The H_2CO abundance is variable, relative to water. The largest value found for comet Halley was 4.5%±0.5%, measured by both IKS and Giotto NMS (also IMS, see text), but at other times the production rates were 10 times smaller (Mumma and Reuter 1989). The values retrieved for comets Austin and Levy were much smaller than the values found in comet Halley.

[4] Retrieved from a single spectral line of CH_4. The range reflects the uncertainty in rotational temperature for cometary CH_4 (50–200 K). The extrapolation from a single line to the ensemble production rate is therefore highly uncertain. See Kawara et al. (1988).

TABLE II (cont.)

[5] The production rate retrieved from the neutral mass spectra on Giotto is highly model dependent, and could be zero in comet Halley (Allen et al. 1987; Boice et al. 1990). Brooke et al. (1991*a*) recently found CH_4<0.2% in comet Levy (1990c).
[6] Assuming that NH_2 is produced solely from NH_3 (Magee-Sauer et al. 1989; Mumma et al. 1990; Krasnopolsky and Tkachuk 1991; Wyckoff et al. 1991).
[7] Allen et al. (1987). Boice et al. (1990) re-analyzed the IMS spectra, retrieving 1%, while Ip et al. (1990) retrieved 0.5% from the Giotto IMS data.
[8] Marconi et al. (1990). An incorrect lifetime was used in deriving the abundance of H_2S (see Crovisier et al. 1991).
[9] See Geiss et al. (1991) and Eberhardt et al. (1991).
[10] Based on groundbased infrared spectroscopy (Brooke et al. 1991*a*).
[11] Based on airborne infrared spectroscopy (Larson et al. 1989).
[12] The value in comet Levy was ~1%, but in comet Austin it was about 5% (Bockelée-Morvan et al. 1991; Bockelée-Morvan, personal communication); Hoban et al. 1991).
[13] 0.1% in comet Austin (1990 V), and 0.04% in comet Levy (1990 XX). The production rate would be about ten-fold larger if formaldehyde were a daughter product (Colom et al. 1992).
[14] In P/Brorsen-Metcalf (1989 X), Austin (1990 V), Levy (1990 XX); compare Bockelée-Morvan et al. 1990.
[15] Crovisier et al. (1991).
[16] cf. Kim et al. (1990).

Arend-Rigaux. These latter two comets have such low activity that they are often identified as asteroidal in appearance. Eventually, the crust may cover the whole surface and only gases will leak through the possibly porous crust. This may be the case for comet Neujmin 1, which shows no detectable dust but significant gas.

V. COMPOSITION

The composition of cometary nuclei is inferred from studies of the three coma components, namely neutral gas, plasma and dust. We will consider each component separately before discussing the bulk composition of the nucleus.

A. The Gas Composition

Groundbased studies of coma gases were, until the 1986 apparition of comet Halley, the principal source of information about the composition of comets, and they remain the basis for comparing comet Halley with other comets. The production rates of several molecular species (initially CN and C_2, and later C_3, OH, O, NH and NH_2) have been determined in the comae of more than 100 comets (A'Hearn et al. 1991; Osip et al. 1991; Newburn and Spinrad 1989; Schleicher et al. 1987; Cochran 1987; Cochran and Barker 1987). These studies indicate that comets are, to a significant extent, a chemically homogeneous set of objects. The "normal" production rate ratios are defined by roughly 50 to 70% of the objects, and comet Halley falls in this group (Cochran 1987,1989).

Most gaseous species accessible to groundbased observations of the coma are photo-dissociation (and to a lesser extent, chemical reaction) products of parent molecules that sublimate from nucleus ices, and/or that come from hydrocarbon grains in the coma. Until the application of improved infrared and radio spectroscopy techniques in 1985, the composition of these ices and hydrocarbons could only be inferred from coma observations of daughter products. Hence, the coordinated *in situ* and remote observations of comet Halley have brought about a dramatic improvement in our knowledge of the ice and dust composition. The two most recent reviews of this subject are by A'Hearn and Festou (1990) and by Krankowsky (1991).

Water vapor was identified spectroscopically for the first time in comet Halley, and was found to be the most abundant volatile, at 80% by number (Mumma et al. 1986; Weaver et al. 1986; Combes et al. 1986). It had previously been inferred from indirect evidence, such as the high observed OH production rates and the "turn-on" of comets inside 3 AU, that water ice was the major fraction of nucleus ices in practically all cometary nuclei (see, e.g., Delsemme 1991). Other molecules, which have been identified with a reasonably high degree of confidence in comet Halley are listed in Table II, along with their production rates relative to water.

The only species clearly present in comet Halley at more than 1% of H_2O were CO (Eberhardt et al. 1986,1987), CO_2 (Combes et al. 1988) and H_2CO. The H_2CO abundance appeared to vary, possibly correlated with rotational phase, with a value of $4.5\% \pm 0.5\%$ during the encounters with the Giotto (Krankowsky 1991) and Vega (Combes et al. 1988) spacecraft, and about a factor of 10 less abundant at other times. CO has only been observed in the brightest of comets because its emission bands have low fluorescence efficiencies. Its abundance varies greatly from comet to comet, ranging from about 1% to 30% (see, e.g., A'Hearn and Festou 1990); since the A'Hearn and Festou review, a detection of CO in comet Austin 1989V has been reported by DiSanti et al. (1990). At least half of the CO measured in comet Halley came from an extended source in the coma, probably hydrocarbon grains (Eberhardt et al. 1986,1987). Consistent with this result is the observation that CO is apparently more abundant in dusty comets (A'Hearn and Festou 1990). CO_2 has not been detected in any comet other than Halley, mainly due to practical limitations with groundbased observations. However, CO_2^+ has been detected in many comets (see below). Because of their high volatility, CO and CO_2 may be responsible for outbursts in Halley and in other comets at large heliocentric distances (Feldman et al. 1987; Weissman 1991*b*). The apparent coexistence of oxidized (CO, CO_2) and reduced (CH_4, NH_3) species in comet Halley and other comets is remarkable.[b]

The equilibrium condensation of a gas with cosmic abundances (Lewis

[b] So far there is no direct and unambiguous detection of NH_3 in any comet; CH_4 may have been detected in comets Halley and Wilson, but at least in comet Halley the CH_4 abundance is very uncertain. See Table II.

1974) predicts reduced compounds. The molecular abundances observed in comet Halley indicate that its ices formed in nonequilibrium conditions. In general, cometary compositions are similar to those observed in dense molecular cloud cores, and thus appear to have undergone very little reprocessing in the primordial solar nebula (Mumma et al. 1993). The abundance of CH_4 relative to CO may provide a key test for the manner in which cometary ices formed (Boice et al. 1990; Mumma et al. 1993, and references therein). Detailed descriptions of the current chemical theories on the origin of comets can be found in recent reviews (Delsemme 1991; Yamamoto 1991; Mumma et al. 1993).

Five other molecules have been observed in Halley or in other comets with a lower degree of confidence. S_2 was discovered and positively identified in comet IRAS-Araki-Alcock (A'Hearn et al. 1983) with an abundance of about 0.025% relative to OH (Kim et al. 1990). There is also a possible identification of S_2 in Halley (Kim and A'Hearn 1991). S_2 is not expected to have formed as a condensate from the gaseous phase. Its presence in comets seems to be the result of irradiation and surface reactions on interstellar grains (A'Hearn and Feldman 1984), which must have been preserved at temperatures below 30 K (Greenberg et al. 1986). Some objections to this proposed origin of S_2 are discussed by Mumma et al. (1993).

CH_3CN was reported in comet Kohoutek 1973 XII by Ulich and Conklin (1974) but never confirmed. OCS has been tentatively identified in comet Halley (Combes et al. 1988). Two new molecules, H_2S and CH_3OH, have been positively identified in millimeter observations of comets Austin 1990 V and Levy 1990 XX (Bockelée-Morvan et al. 1990; Crovisier 1991). These authors also observed HCN, which had previously been detected in comets Kohoutek 1973 XII (Huebner et al. 1974) and Halley 1986 III (Crovisier and Schloerb 1991, and references therein), and H_2CO, previously identified in comet Halley, in the two comets mentioned above, and in periodic comet Brorsen-Metcalf. H_2CO has also been reported in comet Machholz 1988 XV by Snyder et al. (1990); however, strong objections to this detection are discussed by Bockelée-Morvan and Crovisier (1992). CH_3OH has also been tentatively identified in the infrared spectrum of comet Austin 1990 V (Hoban et al. 1991).

A clear deviation from normal abundances was first noticed in periodic comet Giacobini-Zinner (Schleicher et al. 1987; Cochran and Barker 1987). Since then, several comets with C_2, C_3, CN and NH depletions have been identified. The most extreme example is periodic comet Wolf-Harrington with an abundance of C_2 and C_3 less than 0.05 times normal (Schleicher et al. 1991). Also, Fink (1991) has recently obtained spectra of comet Yanaka 1988 XXIV which shows *no* evidence of either C_2 or CN. These observations suggest that some fraction of comets may have formed in radically different regions of the solar nebula, in giant planet sub-nebulae, or were somehow processed to remove and/or destroy their hydrocarbon molecules. Analysis of many unpublished observations is still in progress (Osip et al. 1991).

B. The Plasma Composition

A number of ions have been identified in the comae and tails of comets, ten of them before 1985 (A'Hearn 1984) and as many as nine more by spacecraft mass spectrometers (Huebner et al. 1991). Although less directly than the neutral gas or the dust, ions provide clues about parent molecules and thus about the composition of the nucleus. For example, the detection of species such as H_2O^+ and CO_2^+ were indicative of the presence of H_2O and CO_2 in a number of comets, before these two neutral species were identified in comet Halley. Also, observations of H_3O^+ in comet Halley confirmed predictions of chemical models for cometary atmospheres and ionospheres, and further constrained these models (Huebner et al. 1991).

As with the CH_4/CO ratio mentioned above, the relative abundance of N_2 and NH_3 may be indicative of the environment where comets formed (Prinn and Fegley 1989). However, N_2 has never been identified spectroscopically in comets. Nevertheless, the observed ratios N_2^+/CO^+ and CO/H_2O have been used to estimate the abundance of N_2 relative to H_2O (Table II; Wyckoff and Theobald 1989).

C. The Dust Composition

Direct information on the composition of solid cometary particles was very scarce prior to *in-situ* analysis of particles in Halley's coma in 1986. Indirect information was available from radiometry of cometary dust comae (Hanner and Tokunaga 1991, and references therein), spectroscopic studies of meteoroids (Millman 1977), and from laboratory analysis of interplanetary dust particles (Brownlee 1985), at least some of which are believed to be of cometary origin.

Three solid-state spectral features had been detected in comets prior to the 1986 Halley apparition. An absorption feature near 3 μm attributed to water ice in coma grains was observed in two dynamically new LP comets with large perihelia: Bowell 1982 I (Campins et al. 1983) and Cernis 1983 XII (Hanner 1984). A broad emission feature from 8 to 13 μm attributed to silicates was discovered in comet Bennett (Maas et al. 1970) and was observed along with another silicate emission near 20 μm in four other bright comets (Ney 1982, and references therein). Comet Halley prominently displayed the 10 and 20 μm silicate features (Hanner et al. 1987; Ryan and Campins 1991). Structure discovered in Halley's 10 μm silicate emission (Bregman et al. 1987) has been attributed to crystalline olivine particles (Campins and Ryan 1989), and has also been observed in comet Bradfield 1987 XXIX (Hanner et al. 1990).

Also discovered in the spectrum of comet Halley was an emission feature near 3.4 μm (Combes et al. 1986). This feature is believed to arise from C-H molecular stretch bands in complex organic compounds. However, the composition of the specific emitting material is still unidentified. The width, intensity and structure of this emission band suggest that at least several and perhaps many molecular species are involved (Encrenaz and

Knacke 1991, and references therein). Encrenaz and Knacke (1991) point out that most or all of the 3.4 μm emission has to come from small cometary grains and/or large molecules. Hoban et al. (1991) identified CH_3OH as the progenitor of a feature at 3.52 μm in comet Austin 1990 V, and suggested that it should also contribute significantly to the 3.4 μm "cometary organics" feature. Reuter (1992) modeled the CH_3OH contribution to the 3.4 μm emission and concluded that methanol could have contributed 10 to 30% of the flux at 3.4 μm in recent comets.

The 3.4 μm feature has also been seen in comets Wilson 1987 VII (Brooke et al. 1989; Allen and Wickramasinghe 1987), Bradfield 1987 XXIX (Brooke et al. 1989), Okazaki-Levy-Rudenko 1989 XIX (Brooke et al. 1991*b*), and periodic comet Brorsen-Metcalf (Brooke et al. 1991*b*). Examples of this feature in each of these comets are shown in Fig. 5. A series of low-level features were discovered in the airborne spectra of comet Halley at 6.8 μm (Campins et al. 1986; Bregman et al. 1987), and at 24, 28, 35 and 45 μm (Herter et al. 1987; Glaccum et al. 1987). The four features between 24 and 45 μm have been tentatively attributed to crystalline olivine (Campins and Ryan 1989). The 6.8 μm feature can be attributed either to C-H deformation mode in organic compounds, or emission from carbonates (Bregman et al. 1987).

In-situ measurements of solid particles in the coma of comet Halley were carried out by the PIA and PUMA experiments onboard the Giotto and Vega spacecraft, respectively. These observations are complex and are still being analyzed; initial results have been presented in a number of papers (Kissel et al. 1986*a*, *b*) and have been summarized most recently by Grün and Jessberger (1990) and by Jessberger and Kissel (1991). It was known from groundbased observations that many elements in comets are present at their cosmic abundances. The *in-situ* measurements showed this to hold true for most elements. The average atomic abundances of the elements, relative to magnesium, in Halley's dust grains and in the whole comet (dust plus ice; see discussion below) are given in Table III. For comparison, the abundances of CI carbonaceous chondrite meteorites and the solar system (also known as cosmic) abundances from Anders and Grevesse (1989), are also presented in Table III. The abundances of all the elements heavier than oxygen (rock-forming elements) are within a factor of 2 of the CI and solar system values. The light elements, hydrogen, oxygen, carbon, nitrogen and sulfur are more abundant than in CI chondrites, which are considered to be the most chemically unaltered meteorites.

One of the unexpected results of the *in-situ* experiments was the discovery that a large fraction of the dust is composed of organic material. This organic component was termed CHON because it contained only the light elements carbon, hydrogen, oxygen and nitrogen. There appear to be two end-member particle types, CHON and silicates. Most Halley grains are mixtures of these two types (Jessberger and Kissel 1991). A silicate-core and CHON-mantle structure of the dust particles has been inferred by Krueger and Kissel (1987),

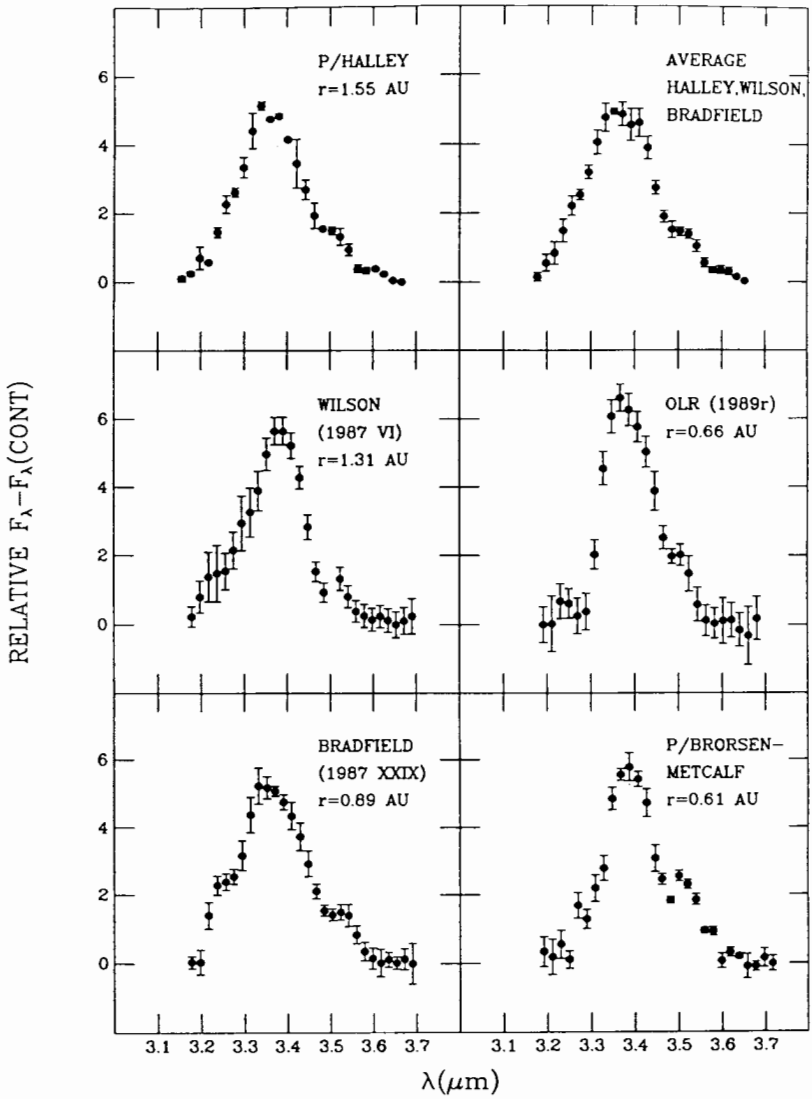


Figure 5. The "organic grain" feature in five recent comets. Comets Wilson and Okazaki-Levy-Rudenko were both dynamically new, while Bradfield was an older long-period comet and both Halley and Brorsen-Metcalf are short-period (from Brooke et al. 1991*b*).

as had been proposed by Greenberg (1982). The carbon that was "missing" from the gaseous fraction when compared with solar system abundances, has been found in the CHON material. Most of the oxygen is in the CHON material and not in the silicates. The *in-situ* studies have provided some

TABLE III

Average Atomic Abundances of Elements in Halley Dust Grains* [1]

Element	Halley		Solar System	C I-Chondrites
	Dust	Dust and Ice		
H[2]	2, 025	4, 062	2, 600, 000	492
C	814	1, 010	940	70.5
N	42	95	291	5.6
O	890	2, 040	2, 216	712
Na	10	10	5.34	5.34
Mg	= 100	= 100	= 100	= 100
Al	6.8	6.8	7.91	7.91
Si	185	185	93.1	93.1
S	72	72	46.9	47.9
K	0.2	0.2	0.35	0.35
Ca	6.3	6.3	5.69	5.69
Ti	0.4	0.4	0.223	0.223
Cr	0.9	0.9	1.26	1.26
Mn	0.5	0.5	0.89	0.89
Fe	52	52	83.8	83.8
Co	0.3	0.3	0.21	0.21
Ni	4.1	4.1	4.59	4.59

* Table from Jessberger et al. (1988).

[1] For comparison, the new solar system abundances and the CI-chondrite composition are also given (Anders and Grevesse 1989). The solar photospheric abundances of the listed elements are practically indistinguishable from the solar system abundances, with the exception of Fe, which has a photospheric abundance ratio of 123 (Anders and Grevesse 1989).

[2] From *short* spectra only.

information on the molecular composition of the organic dust, which consists of complex unsaturated polycondensates according to Krueger et al. (1990). Information on the mineralogy of particles is also limited and based principally on analogies with interplanetary dust particles (IDPs). Cometary silicates resemble most closely anhydrous silicate IDPs (Brownlee et al. 1987; Bradley 1988), in agreement with infrared observations (Campins and Ryan 1989).

The average composition of comet Halley's nucleus can be determined from the composition of the gas and dust components if the dust-to-gas ratio is known. However, this value has proven to be difficult to establish, even in comet Halley. Estimates range from 0.5 to 2 (McDonnell et al. 1991), or more. A value of 2 is adopted by Jessberger et al. (1988), and used to calculate the abundances in the "Dust and Ice" column of Table III.

Sykes and Walker (1992) estimated total dust production rates for eight SP comets, from estimates of the amount of material seen in the IRAS dust trails. Comparing these results with gas production measurements for the same comets, they find dust-to-gas ratios of 1.2 to 4.6, with a mean value of

~3.

The internal structure of cometary nuclei is very uncertain (see reviews by Keller 1990 and Rickman 1991). A variety of models have been suggested (Whipple 1950; Donn et al. 1985; Weissman 1986*a*; Gombosi and Houpis 1986) and several of these are illustrated in Fig. 6. The current consensus is that nuclei are weakly bonded, fractal assemblages of smaller icy-conglomerate planetesimals, possibly "welded" into a single nucleus by thermal processing and sintering. Dynamical scattering of planetesimals by the outer planets may have resulted in a mixing of bodies formed at different heliocentric distances in the protosolar nebula, and hence, in different temperature regimes. These different regimes may be reflected in compositional differences between the fragments making up each cometary nucleus.

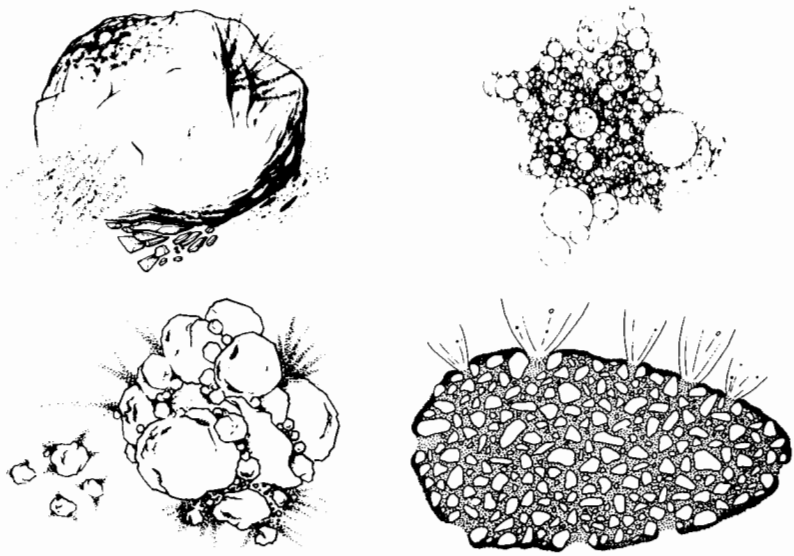


Figure 6. Four suggested models for the structure of cometary nuclei: (a) the icy conglomerate model (Whipple 1950; drawing from Weissman and Kieffer 1981); (b) the fractal model (Donn et al. 1985); (c) the primordial rubble pile (Weissman 1986*a*); and (d) the icy-glue model (Gombosi and Houpis 1986). All but (d) were suggested prior to the Halley spacecraft encounters in 1986.

The compositional similarities between new and periodic comets is one of the arguments used by Delsemme (1982,1991) to suggest that comet nuclei are not differentiated. However, small scale heterogeneity and differentiation is apparently present. For example, crust formation produces material rich in refractory elements and compounds, probably similar in composition to carbonaceous meteorites. In addition, gas jets rich in certain compounds such as H_2CO reported in comet Halley (Mumma et al. 1990), and the temporal variability of several chemical species in comet Encke (A'Hearn et al.

1983,1984; Campins 1988), are suggestive of some heterogeneity within the nuclei of these two comets.

The ratio of mass to volume yields the average density of cometary nuclei. In practice, the density is very difficult to determine because both mass and volume are so uncertain. For comet Halley, estimates based on modeling of nongravitational forces range from 0.2 to more than 1 g cm^{-3} (Rickman 1986,1989; Sagdeev et al. 1987; Peale 1989). Another method is to estimate the density of the nucleus based on the density of the solid particles in the coma. These estimates for Halley range from 0.3 to 0.6 g cm^{-3} (Greenberg and Hage 1990). On the other hand, the measured densities of somewhat larger though still modest sized bodies in the outer solar system, Pluto, Charon and Triton, are all $\sim$2.0 g cm^{-3}. The relatively high dust-to-gas estimates of Sykes and Walker (1992), described above, tend to support the possibility that these higher densities may also be present in comets.

The cometary mass distribution is also very difficult to determine. One suggested distribution by Weissman (1991*a*), shown in Fig. 7, is based on the distribution of absolute magnitudes H_{10} as found by Everhart (1967), who corrected for observational selection effects. For comets brighter than $H_{10} = 11$ (corresponding to a radius of 1.2 km and a mass of 4.0×10^{15} g), Weissman found that the average cometary nucleus had a mass of 3.8×10^{16} g and a radius of 2.5 km (assuming a density of 0.6 g cm^{-3}). Bailey and Stagg (1988) derived a very similar mass distribution, though with a steeper slope such that nucleus mass decreases more sharply with increasing absolute magnitude.

D. Isotopic Abundances

Isotopic abundances from measurements of the volatile fraction of comet Halley have been reported for hydrogen, carbon, nitrogen, oxygen and sulfur, and are given in Table IV (adapted from Krankowsky 1991; see also Vanysek 1991). Within the errors, the isotopic ratios are consistent with a solar system origin for the ices in Halley. The low $^{12}C/^{13}C$ value of 65$\pm$9 reported by Wyckoff et al. (1989), has been revised upward to 100$\pm$15 and is now consistent with the solar system value of 89.9 (Kleine et al. 1991). Jessberger and Kissel (1991) have reported $^{12}C/^{13}C$ values ranging from 10 to several times 10^3 in some Halley dust particles, possibly reflecting material from a variety of processing environments in dense molecular cloud cores, where large isotopic variations are often observed, and in the solar nebula. The large range in carbon isotope ratios is similar to that seen in primitive meteorites, and is indicative of the microscopic heterogeneity of cometary materials.

The deuterium/hydrogen ratio in Halley is significantly enriched relative to that of Jupiter and Saturn, which likely obtained their hydrogen reservoirs directly from the solar nebula gas. It is similar to the values found in solar system objects which probably accreted from solid grains in the nebula (i.e., meteorites), and/or those bodies which may have acquired their volatile reservoirs through asteroid and comet bombardment (e.g., Earth, Titan). Even

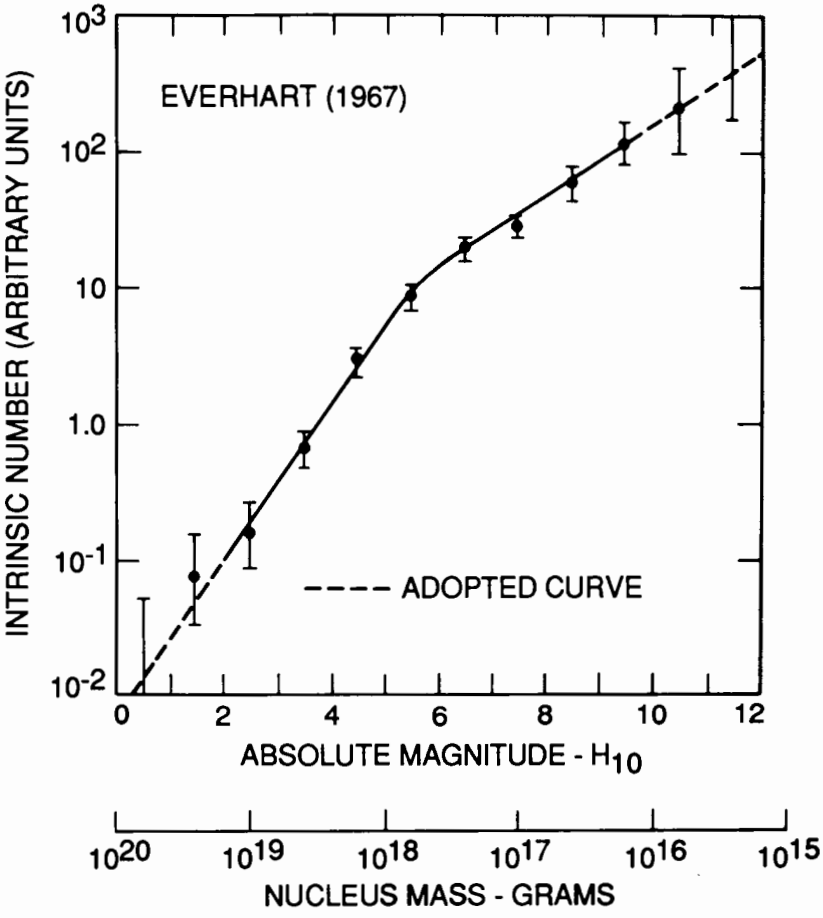


Figure 7. Estimated mass distribution for the long-period comets, based on the distribution of absolute magnitudes H_{10}, corrected for observational selection effects, as found by Everhart (1967). The average cometary nucleus has a radius of 2.5 km and a mass of 3.8×10^{16} g (Weissman 1991*a*).

larger D/H enrichments are observed in some molecules in dense molecular cloud cores, the birthplace of stars.

VI. PHYSICAL EVOLUTION OF COMETARY NUCLEI

A variety of processes act to modify the surfaces and interiors of cometary nuclei over their long lifetimes. These have been reviewed by Weissman (1986*b*) and Weissman and Stern (1993) and include: irradiation by cosmic rays and solar wind protons (Johnson et al. 1987), heating by supernovae and passing stars (Stern and Shull 1988), competing erosion and accretion by interstellar dust grains (Stern 1986,1990), conversion from amorphous to

TABLE IV
Isotope Ratios Measured in Halley Coma Volatiles*

Element	Isotope Ratio	Chemical Form Measured
D/H	$0.6 \times 10^{-4} - 4.8 \times 10^{-4}$	Water, protonated ions[1]
$^{12}C/^{13}C$	100 ± 15	Cyanide radical[2]
$^{14}N/^{15}N$	≥ 200	Cyanide radical[2]
$^{18}O/^{16}O$	0.0023 ± 0.0006	Water, protonated ions[1]
$^{34}S/^{32}S$	0.045 ± 0.010	Sulfur ion[3]

* Table adapted from Krankowsky (1991); see also Vanysek 1991.
[1] Eberhardt et al. (1987*a*).
[2] Wyckoff et al. (1989): $^{12}C/^{13}C$ since revised to near-solar system value (Kleine et al. 1991).
[3] Krankowsky et al. (1986).

crystalline ice (Smoluchowski 1981; Prialnik and Bar-Nun 1988), and solar heating which results in sublimation of surface ices, internal mobilization of volatile ices, and formation of cometary surface crusts (Brin and Mendis 1979; Weissman and Kieffer 1981; Fanale and Salvail 1984). All of these processes are fairly modest when compared with typical planetary processes such as giant impacts or differentiation and core formation. The low degree of processing is evidenced by the high volatile content of comets.

Of particular interest with regard to space exploitation of comets is the formation of nonvolatile crusts on the nucleus surfaces. As discussed above, all nuclei studied to date show evidence that a substantial fraction of their surfaces are covered by an inert crust. In essence, these crusts greatly extend the physical lifetimes of the comets by preventing the sublimation of cometary ices. In addition, the crust may eventually totally cover the nucleus surface, making the comet appear to be a dormant, asteroidal object. This leads to the interesting possibility that some of the Earth-crossing and Earth-approaching asteroids may in fact be extinct cometary nuclei which have been totally covered over by a nonvolatile crust. A number of theoretical models have been developed in which an inert crust, a few centimeters to perhaps a meter in thickness, builds up on the surface of the nucleus, reducing or stopping the sublimation of ices (Mendis and Brin 1977; Brin and Mendis 1979; Horanyi et al. 1984; Fanale and Salvail 1984; Rickman and Fernandez 1986). Despite these models, some of which are fairly complex, many details about crust formation remain unclear. Does the irradiation of nuclei by cosmic rays while in the Oort cloud lead to a heavily processed surface layer tens of centimeters thick, which serves as the basis of the nucleus crust even before the comet enters the planetary region? Is the crust simply a lag deposit of unbonded large particles, too heavy to be entrained in the evolving gas, or does the large fraction of organics revealed by the Halley missions play a role, acting as a cometary "glue" to cement nonvolatile grains together? Is the crust porous, allowing evolved gases to diffuse through it, or do the organic glue and the finer particles eventually seal the pores, creating a fairly impervious,

high-strength crust? As high pressure builds up beneath the crust, does it fail gradually or catastrophically, on small scales or large? How deep beneath this crust can ice be found?

Answers to these questions are difficult without direct observations of cometary surface processes, as would be afforded by the Comet Rendezvous Asteroid Flyby mission (see Sec. VIII). Observations of the Halley nucleus during the spacecraft flybys confirmed the existence of the dark, inactive crust, but provided little information about its physical characteristics. A more immediate source of information are the very interesting series of experiments that have been conducted by the KOSI group (Grün et al. 1991) in which icy-conglomerate cometary analogs have been constructed and instrumented, and exposed to simulated sunlight in large vacuum chambers usually used for spacecraft testing. These experiments have confirmed the existence of a number of interesting processes, including crust formation, and different depths of sublimation for different volatile ices. Continued work in this area has great potential to add to our understanding of cometary processes.

Evolution to an extinct, asteroid-like object is only one possible end-state for cometary nuclei. It is possible that the cometary ices continue to sublimate until nothing is left of the nucleus, except perhaps a meteoroid stream of small debris (Fernandez 1988). Or the nucleus may disrupt more spectacularly, forming a double comet as in the case of comet Biela, before totally disintegrating. None of these processes are well understood, and the fraction of comets that arrive at each end-state is unknown.

VII. EVOLUTION OF COMETS INTO ASTEROIDS

The Earth-crossing and Earth-approaching asteroids provide a readily available source of materials for space development. In some cases, they are available at less energy expenditure than material on the lunar surface, due to their low encounter velocities with the Earth and their lack of a significant gravity well. Perhaps most intriguing is the possibility that a significant fraction of the near-Earth asteroids may be volatile-rich extinct cometary nuclei, dormant remnants of SP comets.

The origin of the near-Earth asteroids has been the subject of an ongoing debate among solar system scientists for several decades now. Because they are in planet-crossing orbits, these objects have dynamical lifetimes of only ~30 Myr (Wetherill 1975). Thus, the population of Earth-crossing asteroids must be continuously replenished. The origin arguments have largely split along discipline lines. In general, observers believe that the near-Earth asteroids are derived from the main asteroid belt between Mars and Jupiter, because they appear similar spectroscopically to main belt objects. On the other hand, dynamicists have favored a cometary source because of the lack of known dynamical mechanisms to move sufficient numbers of asteroids from the main belt to near-Earth orbits. However, in the past decade two dynamical mechanisms for delivery of main belt asteroids have been recognized which

could account for at least half of the estimated population of near-Earth asteroids (Wetherill 1988), while observational evidence has continued to mount of anomalous Earth-crossing asteroids whose characteristics and behavior might be explained if the objects were indeed extinct cometary nuclei. The recognition that ~70% of the surface of comet Halley was covered by an apparently inert crust was important in lending credibility to the idea that comets could evolve into dormant, asteroidal-appearing objects.

Various tests of cometary versus asteroidal origin have been put forward, usually based on statistical differences between the size, shape, rotation periods, spectral properties, meteor stream associations, and orbital dynamics of their parent populations. However, none of these tests are definitive, and many near-Earth objects display contradictory combinations of physical attributes. The one agreed-upon test of a cometary origin is the ability to generate an appreciable cometary coma around the nucleus. There are no confirmed observations of coma activity in any of the near-Earth asteroids, though there are a number of independent observations of what may well be sporadic activity.

The possible evolution of SP comets to inert-appearing, near-Earth asteroids has been reviewed extensively by Öpik (1963), Wetherill (1971), Degewij and Tedesco (1982), Kresak (1985), Fernandez (1988) and Weissman et al. (1989). The reader is referred to those papers, and other references therein, for a more complete discussion of the problem. The reviews by Degewij and Tedesco and by Weissman et al. are particularly recommended as they arrive at rather opposite conclusions as to the likelihood of extinct or dormant comets among the near-Earth asteroids, and thus provide good examples of the two sides in the debate.

A list of comet-like asteroids and a detailed discussion of the characteristics that make each object a member of the list is given in Weissman et al. (1989). *Probable* cometary candidates identified among the numbered asteroids include: 944 Hidalgo, 2060 Chiron, 2101 Adonis, 2201 Oljato, 2212 Hephaistos, 3200 Phaethon and 3552 Don Quixote; while *possible* candidates include: 1566 Icarus, 1580 Betulia, 1620 Geographos, 1685 Toro, 1862 Apollo, 1866 Sisyphus, 1917 Cuyo, 1981 Midas and 2062 Aten. The discussion here is limited to two objects: 1566 Icarus and 2060 Chiron, because relevant information on them has become available since the publication of that review.

A. 1566 Icarus

The spherical shape, F-type spectrum, and rapid rotation of this object were cited in Weissman et al. (1989) as suggestive of an asteroidal nature. However, recent work of Yeomans (1991) shows that 1566 Icarus displays comet-like nongravitational effects in its orbital motion. Yeomans searched for nongravitational effects in a total of 12 near-Earth asteroids identified as possible extinct comets; only 1566 Icarus showed clear evidence for them. The rapid rotation is not necessarily surprising if its small size, ~1 km, is taken into account. Further observations of this object are clearly warranted.

B. 2060 Chiron

This unique object was initially classified as an asteroid, albeit an unusual one because of its orbit: semimajor axis, 13.7 AU; eccentricity, 0.38 (Kowal et al. 1979). The unexpected brightening of Chiron first noticed by Tholen et al. (1988) and its subsequent irregular photometric behavior (French et al. 1989, and references therein) suggested comet-like activity. A faint coma was first detected in April 1989 (Meech and Belton 1989) and has been observed since by several groups (West 1990; Luu and Jewitt 1990; Meech and Belton 1990). The detection of CN emission around Chiron (Bus et al. 1991*a*) confirmed the cometary nature of this object. Bus et al. (1991*b*) performed photographic photometry on pre-discovery plates of Chiron and found that it was as bright or brighter in the early 1970s at 18.9 AU, near aphelion, than during its recent outburst. Stern (1989) discussed several mechanisms to account for the activity in Chiron and pointed out that if certain gaseous species are detected, strong constraints can be placed on the residence time of this object in the planetary region of the solar system.

Chiron may be the largest known cometary nucleus (Table I). Spectrally, Chiron is neutral in the visible and near infrared, with no evidence for strong H_2O absorption bands (Lebofsky et al. 1984; Hartmann et al. 1989). Dynamically, this object is in a chaotic orbit, evolving at present primarily under the influence of Saturn. Hahn and Bailey (1990) integrated 82 trial orbits for Chiron backward and forward in time 10^5 yr and found that a substantial fraction were perturbed to Earth-crossing SP comet orbits at some time during that period. This raises the interesting possibility that extremely large SP comets may approach the Earth and may occasionally be available for resource exploitation. The catastrophic consequences of such large SP comets in near-Earth space has been speculated on by Clube and Napier (1986).

VIII. FUTURE COMETARY MISSIONS

The 1986 spacecraft encounters with comet Halley provided a quantum leap forward in our knowledge and understanding of comets, just as other planetary exploration spacecraft have done for eight of the nine planets and their satellite systems. However, the Halley flybys were at ultra high speed, and left many questions unanswered. Thus, the need clearly exists for additional missions to comets which will further add to our knowledge of early solar system cosmochemistry, and which can address the question of the heterogeneity of the overall cometary population.

The closest mission to fruition is the Giotto Extended Mission. The Giotto spacecraft is still operational and eight of its ten scientific instruments are functional, though two are significantly degraded. The non-operating instruments are the Halley multicolor camera and the neutral mass spectrometer. There has been some damage to some spacecraft engineering subsystems such as the star scanner and the thermal insulation, the latter leading to degraded spacecraft performance beyond about 1.1 AU. Using an Earth gravity assist

maneuver in July, 1990, the spacecraft was retargeted toward an encounter with periodic comet Grigg-Skjellerup in 1992 (Farquhar et al. 1987). The spacecraft will fly by the comet on July 10, 1992, 12 days before Grigg-Skjellerup passes through perihelion. The encounter is at heliocentric and geocentric distances of 1.01 AU and 1.43 AU, respectively, and at a velocity of 14.0 km s^{-1}. Targeting is expected to be within about 1000 to 2000 km of the nucleus.

Given the reduced instrument complement, the Giotto encounter with Grigg-Skjellerup will emphasize fields and particles measurements of the solar wind interaction with the cometary coma. Observations will be made of the cometary bow shock (or its analog), cometosheath and ionopause. As Grigg-Skjellerup has a much lower gas production rate than Halley, the encounter should provide a considerably different picture of cometary plasma physics.

The next logical step in cometary exploration is the Comet Rendezvous Asteroid Flyby (CRAF) mission (Weissman and Neugebauer 1992) which was approved for a New Start by NASA in 1989. The current mission design calls for a launch in April, 1997 and arrival at comet Kopff in May, 2006. CRAF will match orbits with the comet near aphelion at 5.3 AU from the Sun and then follow it through perihelion and outward again. The planned mission phases include: initial slow flybys at velocities <50 m s^{-1} and distances as close as 50 km; nucleus mapping orbits at typical altitudes of 50 to 80 km; close orbits at less than 5 nucleus radii (<20 km) altitude for gravity field and mass determination; a series of slow fly-throughs of the cometary coma near perihelion at ranges between 50 and 500 km (with some forays to 5000 km); a 50,000 km excursion down the comet tail shortly after perihelion in 2009; and finally, return to a low nucleus orbit as cometary activity wanes post-perihelion. Enroute to the comet CRAF will fly by two large main-belt asteroids: 88 Thisbe and 19 Fortuna which are 232 km and 200 km in diameter, respectively, and which both have a C-type, carbonaceous surface composition. A third, smaller asteroid, 1084 Tamariwa, is also being considered as a possible flyby target. The CRAF trajectory, which uses Mars and Earth gravity assists, is shown in Fig. 8. CRAF is a joint project of the space agencies of the United States, Germany and Italy.

The advantages of the rendezvous mission over the fast Halley flybys are fairly obvious. The rendezvous mission allows for long duration observations of the nucleus, coma and tail, over a complete cometary orbit, noting the changes in phenomena with time and with solar distance. Samples of cometary dust and gas, and possibly ice, can be collected at low relative velocities which do not result in vaporization of volatiles or solids. Thus, it is possible to identify and study cometary molecules, hydrocarbons, minerals and ices, rather than inferring their existence from elemental identifications only. By orbiting the nucleus the spacecraft can map 100% of the nucleus surface, determine its mass and bulk density, and observe changes in surface features as cometary activity builds and declines. The spacecraft may be able to

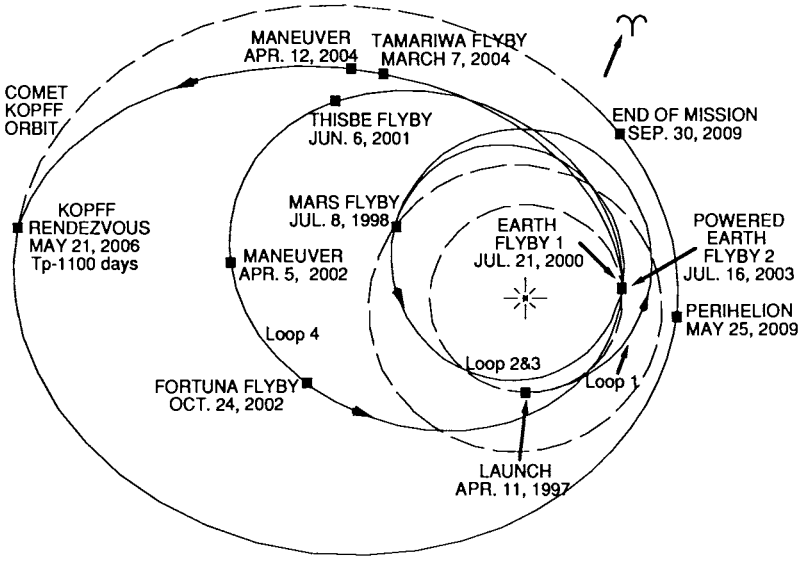


Figure 8. Interplanetary trajectory for the 1997 launch of the Comet Rendezvous Asteroid Flyby mission. The spacecraft will use a gravity assist flyby of Mars and two of Earth to enable it to match orbits with periodic comet Kopff in May, 2006. Two or three asteroid flybys will be conducted enroute to the comet.

maneuver through several individual cometary jets to search for compositional differences. The spacecraft can explore the coma/solar-wind interaction in three-dimensional space around the comet and separate spatial and temporal events. Most importantly, the long residence time at the comet allows design of adaptive science sequences that can build on the knowledge gained early in the mission to define more diagnostic observations.

The CRAF payload is shown in Table V. The narrow angle CCD camera is expected to yield a surface resolution of better than 1 meter/line-pair at an orbital altitude of 63 km. At the same distance, the resolution elements for VIMS and TIREX are 32 and 95 meters, respectively. A secondary ion mass spectrometer will analyze cometary dust particles with a resolution of $\geq$3000 at 13 amu and $\geq$13,000 at 350 amu, allowing detailed elemental and isotopic analyses. Additional compositional analysis of dust and ice will come from CIDEX which includes a gas chromatograph with three columns and an X-ray fluorescence spectrometer. A neutral gas/ion mass spectrometer will determine the composition of gases in the coma and a dust counter will determine the number, mass, velocity and charge of cometary dust particles. Finally, a full complement of magnetospheric instruments will observe the solar-wind interaction with the cometary coma and tail.

By far the most ambitious cometary mission now under consideration is Rosetta, the Comet Nucleus Sample Return mission, which is currently the subject of a joint study by the European Space Agency (ESA) and NASA

TABLE V
CRAF Payload

Instrument	Capabilities
Imaging	2000 mm and 250 mm cameras, $\lambda = 0.4$ to 1.1 μm, 6 and 48 mrad fields of view, 1024×1024 CCDs
Visual/infrared mapping spectrometer	$\lambda = 0.35$ to 5.14 μm, CCD and InSb detectors, 0.5×0.5 mrad pixel with 32×32 mrad field of view
Thermal infrared radiometer	$\lambda = 5$ to 50 μm plus visual channel, 1.5×1.5 mrad field of view
Comet dust environment monitor	Dust detector: momentum $\geq 10^{-11}$ N s^{-1}, charge $> 10^{-15}$ coulomb
Cometary matter analyzer	Mass spectrometer, 1–350 amu, 1–3000 amu, m/dm $\geq$3,000 at 13 amu; $\geq$13,000 at 350 amu
Cometary ice and dust experiment	X-ray fluorescence spectrometer and gas chromatograph
Neutral gas and ion mass spectrometer	Mass spectrometer, 1–300 amu. Dynamic range: 10 to 10^{12} molecules cm^{-3}.
Magnetometer	Field strengths $>$0.03 nT
Cometary retarding ion mass spectrometer	0.01 to 150 eV/q, ion temperatures $\geq$25 K, densities: 10^{-1} to 10^{6} ions cm^{-3}
Suprathermal plasma investigation of cometary environments	Energy range: 1 eV to 80 KeV. Mass: 1 to 65 amu.
Coordinated radio, electrons, and waves experiment	Electrons: 0.5 eV to 6 keV, plasma waves: 0.1 Hz to 6.6 MHz.
Radio science	Nucleus mass to better than 0.1%; higher gravity field harmonics. Mascons. Asteroid mass to better than 1%

(Schwehm and Langevin 1991). ESA has designated Rosetta one of the four "cornerstone" missions of its Horizon 2000 Program. The goal of the mission is to rendezvous with and land on a cometary nucleus, obtain a 10 kg core sample 1 to 3 meters in length, and return it intact (and cold) to terrestrial laboratories for detailed analyses. Additionally, the Rosetta mission would try to return a 1 to 5 kg sample of nonvolatile cometary crust material, and a 10 to 100 g sample of volatile-rich material, perhaps from the bottom of the core sample hole. Rosetta would also carry a camera for imaging the sample collection site and probes to monitor temperatures during sample collection. Additional instruments include an imaging subsystem, an infrared spectrometer, an infrared thermal mapper, a neutral and ion mass spectrometer, a radar altimeter and sounder, and a dust counter.

The Rosetta mission is envisioned to be launched sometime in the next decade, using an expendable Titan 4/Centaur or Ariane 5 launch vehicle. The mission may use the current CRAF Mariner Mark II spacecraft for its baseline design. Study of possible mission options, sampling strategies and techniques, and advanced propulsion systems is continuing.

A simpler proposed option for a cometary sample return mission is a high speed coma fly-through with sample collecting cells mounted on the spacecraft. Cometary dust particles would vaporize as they penetrated the thin foil covering the capture cells and then recondense on the inside of the cells. Or, dust particles might be captured intact by slowly decelerating them in an underdense material, such as styrofoam. It may even be possible to capture samples of coma gases. The spacecraft would be targeted for an Earth return trajectory, and a small, heat shield equipped entry capsule would carry the samples through the atmosphere and descend on a parachute for recovery and eventual analysis. This type mission has been studied by NASA, ESA and Japan, and extensive laboratory work has been conducted towards showing that the proposed capture techniques do work (Tsou et al. 1984).

IX. DISCUSSION

The successful development of a human presence in near-Earth space will depend critically on the ability to exploit large, easily processed sources of volatiles. In particular, water will be among the more difficult resources to find in space, or to process from lunar materials or most near-Earth asteroids, which are primarily S-type, stony-iron objects. In contrast, the SP comets represent a rich source of water, other volatiles, and hydrocarbons, obtainable with minimal processing effort. A typical cometary nucleus is approximately 25% water ice, 5% other volatile ices, 35% hydrocarbons and 35% silicates.

The difficulty, of course, lies in the high encounter velocities between most SP comets and the Earth. Typical hyperbolic encounter velocities are on the order of 25 km s^{-1} or more. However, the velocity distribution is quite broad and some Earth-crossing or Earth-approaching SP comets have encounter velocities below 20 km s^{-1}. For example, the active SP comet

Grigg-Skjellerup, the target of the second Giotto encounter, has a hyperbolic encounter velocity of only 15.3 km s^{-1}, comparable with many Earth-crossing asteroids.

The best hope for easily obtaining volatiles from cometary bodies lies with the possibility that some fraction of the near-Earth asteroids are actually extinct or dormant cometary nuclei. These objects have already been perturbed into orbits with relatively low Earth encounter velocities, and so the difficulty of matching orbits with them is significantly reduced. Moreover, technologies have been proposed for changing the orbits of near-Earth asteroids and comets using mass-drivers, bringing them to orbits and/or sites where they can be more readily utilized. Although the mass-wastage fraction might be quite high in some cases, the total estimated cost is usually still far less than raising water and other volatiles out of deep gravitational potential wells.

Continued physical studies of the near-Earth asteroids are required to ascertain which might be extinct nuclei, coupled with spacecraft missions that can provide a more complete reconnaissance of individual objects, and perhaps even sample return missions. A variety of missions have been proposed to main-belt and near-Earth asteroids (Veverka et al. 1989); those to the Earth-crossing asteroids are the easiest to perform energetically, and have the shortest flight times.

Spacecraft missions to comets will also play a vital role in further evaluating and understanding the most primitive bodies in the solar system. Rendezvous missions with a number of comets are desired to examine the details of this diverse population of objects. Coordinated observations by *in-situ* spacecraft and groundbased (or Earth-orbiting) remote sensing instruments can greatly increase the knowledge in our cometary databases, and address the many questions with regard to cometary homogeneity and heterogeneity. The ultimate near-term desire is a comet sample return mission as proposed by ESA, in which cometary samples can be analyzed slowly and carefully with all the resources of advanced terrestrial laboratories.

Note Added in Proof

Several important developments have occurred since the completion of this chapter. The first likely Kuiper belt object, 1992 QB_1, was discovered in August 1992 by Jewitt and Luu (IAU Circular 5611, 1992) at a heliocentric distance of 41 AU. The 23rd magnitude object is estimated to be 200 km in diameter. A preliminary orbit for the object gives it a semimajor axis of 44.38 AU, an eccentricity of 0.1069, an inclination of 2.22 degrees, and an orbital period of 296 yr (Marsden, IAU Circular 5684, 1992). The perihelion distance of 39.64 AU is well beyond the orbit of Neptune.

Minor planet 4015, 1979 VA was identified by E. Bowell on pre-discovery plates taken in 1949 as showing cometary activity (IAU Circular 5585, 1992). B. Marsden identified this object as comet Wilson-Harrington, 1949 III, one of several "long-period" comets with poorly determined orbits that had been suspected of being short-period (see Sec. II). P/Wilson-Harrington has an

orbital period of 4.3 yr. This discovery has given added credence to the expectation that short-period comets do evolve to dormant, asteroidal objects.

A new detection of Chiron's thermal flux near 20 μm by Campins, Jewitt, and Telesco (IAU Circular 5457, 1992) is in excellent agreement with the value obtained by Lebofsky et al. (1984). If the ILM model is assumed and the faintest V(1,0) (Lebofsky et al. 1984) is taken to be from the "bare" nucleus, the estimated diameter and albedo are 284 (+11,−23) km and 0.04 (−0.01,+0.02), respectively. This implies that Chiron's coma contributed 1/3 of the total brightness in the *R* bandpass at the time of these observations (November, 1991).

The detection of nongravitational forces in the motion of 1566 Icarus (see Sec. VII) has turned out to be erroneous (Yeomans, *Astron. J.* 104:1266, 1992). The problem resulted from neglect of relativistic terms which are important for this small perihelion asteroid.

The Giotto spacecraft successfully encountered periodic comet Grigg-Skjellerup on 10 July 1992. It is estimated that the spacecraft passed within 200 km of the nucleus, most likely on the dark side. Analysis of scientific measurements are in progress.

The Comet Rendezvous Asteroid Flyby mission was cancelled in early 1992 by the Bush Administration. The loss of CRAF is a serious setback for cometary science. Current NASA studies of comet missions are focusing on the low cost Discovery Program with primary attention given to multi-comet flyby missions and simplified orbiters, launched at the end of this decade. In addition, the ESA Rosetta mission will likely be rescoped into a comet nucleus orbiter without a sample return.

Acknowledgment We thank B. Marsden and W. Huebner for their valuable reviews, and D. Yeomans and M. Mumma for useful comments on an earlier draft of this paper. We also thank G. Schwehm for providing the latest information on planning for the Giotto Extended Mission. This research was supported by grants from the National Science Foundation, the NASA Planetary Astronomy Program, and the NASA Planetary Geology and Geophysics Program. The work was performed, in part, at the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration.

REFERENCES

A'Hearn, M. F. 1984. Chemistry of comets. *Chem. Eng. News* 62:32–49.

A'Hearn, M. F. 1988. Observations of cometary nuclei. *Ann. Rev. Earth Planet. Sci.* 16:273–293.

A'Hearn, M. F., and Feldman, P. D. 1984. S_2: A clue to the origin of cometary ice? In *Ices in the Solar System*, eds. J. Klinger, D. Benest, A. Dollfus and R.

Smoluchowski (Dordrecht: D. Reidel), pp. 463–471.
A'Hearn, M. F., and Festou, M. C. 1990. The neutral coma. In *Physics and Chemistry of Comets*, ed. W. F. Huebner (Berlin: Springer Verlag), pp. 69–112.
A'Hearn, M. F., Feldman, P. D., and Schleicher, D. G. 1983. The discovery of S_2 in comet IRAS-Araki-Alcock 1983d. *Astrophys. J. Lett.* 274:99–103.
A'Hearn, M. F., Campins, H., and Schleicher, D. G. 1988. Periodic comet Tempel 2 (1987g). *IAU Circ.* 4619.
A'Hearn, M. F., Campins, H., Schleicher, D. G., and Millis, R. L. 1989. The nucleus of comet P/Tempel 2. *Astrophys. J.* 347:1155–1166.
A'Hearn, M. F., Millis, R. L., Schleicher, D. G., Osip, D. J., and Birch, P. V. 1991. Evolutionary effects in chemical abundances. In *Asteroids, Comets and Meteors 1991*, LPI Contrib. No. 765, p. 1 (abstract).
Allen, D. A., and Wickramasinghe, D. T. 1987. Discovery of organic grains in comet Wilson. *Nature* 329:615–616.
Allen, M., Delitsky, M., Huntress, W., Yung, Y., Ip, W.-H., Schwenn, R., Rosenbauer, H., Shelley, E., Balsiger, H., and Geiss, J. 1987. Evidence for methane and ammonia in the coma of comet Halley. *Astron. Astrophys.* 187:505–512.
Alvarez, W., Alvarez, L. W., Asaro, F., and Michel, H. V. 1980. Extra-terrestrial cause for the Cretaceous-Tertiary extinction: Experimental results and theoretical interpretation. *Science* 208:1095–1108.
Anders, E., and Grevesse, E. 1989. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* 53:197–214.
Anderson, J. D., and Standish, E. M., Jr. 1986. Dynamical evidence for Planet X. In *The Galaxy and the Solar System*, eds. R. Smoluchowski, J. N. Bahcall and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 286–296.
Bailey, M. E., and Stagg, C. R. 1988. Cratering constraints on the inner Oort cloud: Steady state models. *Mon. Not. Roy. Astron. Soc.* 235:1–35.
Bailey, M. E., and Stagg, C. R. 1990. The origin of short-period comets. *Icarus* 86:2–8.
Belton, M. J. S., Spinrad, H., Wehinger, P. A., Wyckoff, S., and Yeomans, D. K. 1987. The spectral behavior of P/Halley at large heliocentric distance in light of the Giotto/Vega results. *Astron. Astrophys.* 187:569–574.
Bockelée-Morvan, D., and Crovisier, J. 1992. Formaldehyde in comets II: Excitation of the rotational lines. *Astron. Astrophys.* 264:282–291.
Bockelée-Morvan, D., Crovisier, J., Colom, P., Despois, D., and Paubert, G. 1990. Observations of parent molecules in comets P/Brorsen-Metcalf, Austin (1989c1) and Levy (1990c) at millimetre wavelengths: HCN, H_2S, H_2CO, and CH_3OH. In *Proc. 24th ESLAB Symposium: Formation of Stars and Planets*, ed. B. Battrick, ESA SP-315, pp. 143–148.
Bockelée-Morvan, D., Colom, P., Crovisier, J., Despois, D., Paubert, G. 1991. Microwave detection of hydrogen sulphide and methanol in comet Austin (1989c1). *Nature* 350:318–320.
Boice, D. C., Huebner, W. F., Sablik, M. J., and Konno, I. 1990. Distributed coma sources and the CH_4/CO ratio in comet Halley. *Geophys. Res. Lett.* 16:1813–1816.
Bradley, J. P. 1988. Analysis of chondritic interplanetary dust thin-sections. *Geochim. Cosmochim. Acta* 52:889–900.
Bregman, J. D., Campins, H., Witteborn, F. C., Wooden, D. H., Rank, D. M., Allamandola, L. J., Cohen, M., and Tielens, A. G. G. M. 1987. Airborne and groundbased spectrophotometry of comet P/Halley from 5–13 micrometers. *Astron. Astrophys.* 187:616–620.
Brin, G. D., and Mendis, D. A. 1979. Dust release and mantle development in comets. *Astrophys. J.* 229:402–408.

Brooke, T. Y., and Knacke, R. F. 1986. The nucleus of comet P/Arend-Rigaux. *Icarus* 67:80–87.

Brooke, T. Y., Knacke, R. F., Owen, T. C., and Tokunaga, A. T. 1989. Spectroscopy of emission features near 3 microns in comet Wilson (1986l). *Astrophys. J.* 336:971–978.

Brooke, T. Y., Tokunaga, A. T., Weaver, H. A., Chin, G., and Geballe, T. R. 1991*a*. A sensitive upper limit on the methane abundance in comet Levy (1990c). *Astrophys. J. Lett.* 372:113–116.

Brooke, T. Y., Tokunaga, A. T., and Knacke, R. F. 1991*b*. Detection of the 3.4 μm emission feature in comets P/Brorsen-Metcalf and Okazaki-Levy-Rudenko (1989r) and an observational summary. *Astron. J.* 101:268–278.

Brownlee, D. E. 1985. Cosmic dust: collection and research. *Ann. Rev. Earth Planet. Sci.* 13:147–173.

Brownlee, D. E., Wheelock, M. M., Temple, S., Bradley, J. P., and Kissel, J. 1987. A quantitative comparison of comet Halley and carbonaceous chondrites at the submicron level. *Lunar Planet. Sci. Conf.* XVIII:133–134 (abstract).

Bus, S. J., Bowell, E., Harris, A. W., and Hewitt, A. V. 1989. 2060 Chiron: CCD and electronographic photometry. *Icarus* 77:223–238.

Bus, S. J., A'Hearn, M. F., Schleicher, D. G. and Bowell, E. 1991*a*. Detection of CN emission from (2060) Chiron. *Science* 251:774–777.

Bus, S. J., Stern, S. A., and A'Hearn, M. F. 1991*b*. Chiron: Evidence for historic cometary activity. In *Asteroids, Comets and Meteors 1991*, LPI Contrib. No. 765, p. 34 (abstract).

Campbell, D. B., Harmon, J. K., and Shapiro, I. I. 1989. Radar observations of comet Halley. *Astrophys. J.*338:1094–1105.

Campins, H. 1988. The anomalous dust production in periodic comet Encke. *Icarus* 73:508–515.

Campins, H., and Ryan, E. V. 1989. The identification of crystalline olivine in cometary silicates. *Astrophys. J.* 341:1059–1066.

Campins, H., Rieke, G. H., and Lebofsky, M. J. 1983. Ice in comet Bowell. *Nature* 301:405–406.

Campins, H., Rieke, G. H., and Lebofsky, M. J. 1985. Absolute calibration of photometry at 1 through 5 microns. *Astron. J.* 90:896–899.

Campins, H., Bregman, J. D., Witteborn, F. C., Wooden, D. H., Rank, D. M., Allamandola, L. J., Cohen, M., and Tielens, A. G. G. M. 1986. Airborne spectrophotometry of comet Halley from 5 to 9 microns. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 121–124.

Campins, H., A'Hearn, M. F., and McFadden, L.-A. 1987. The bare nucleus of comet Neujmin 1. *Astrophys. J.* 316:847–857.

Carusi, A., Kresak, L., Perozzi, E., and Valsecchi, G. B. 1985. *Long-Term Evolution of Short-Period Comets* (Bristol: Adam Hilger).

Clube, S. V. M., and Napier, W. M. 1986. Giant comets and the galaxy: Implications of the terrestrial record. In *The Galaxy and the Solar System*, eds. R. Smoluchowski, J. N. Bahcall and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 260–285.

Cochran, A. L. 1987. Another look at abundance correlations among comets. *Astron. J.* 92:231–238.

Cochran, A. L. 1989. Are all comets created equal? In *Asteroids, Comets and Meteors III*, eds. C.-I. Lagerkvist, H. Rickman, B. A. Lindblad and M. Lindgren (Uppsala: Univ. of Uppsala), pp. 281–284.

Cochran, A. L., and Barker, E. S. 1987. Comet Giacobini-Zinner: A normal comet? *Astron. J.* 92:239–243.

Cochran, A. L., Cochran, W. D., and Barker, E. S. 1982. Spectrophotometry of

comet Schwassmann-Wachmann 1. II. Its color and CO^+ emission. *Astrophys. J.* 254:816–822.

Colom, P., Crovisier, J., Bockelée-Morvan, D., Despois, D., and Paubert, G. 1992. Formaldehyde in comets. I: Microwave observations of P/Brorsen-Metcalf (1989X), Austin (1989c1), and Levy (1990c). *Astron. Astrophys.* 264:270–281.

Combes, M., Moroz, V. I., Crifo, J. F., Bibring, J. P., Coron, N., Crovisier, J., Encrenaz, T., Sanko, N., Grigoriev, A., Bockelée-Morvan, D., Gispert, R., Emerich, C., Lamarre, J. M., Rocard, F., Krasnopolsky, V., and Owen, T. 1986. Detection of parent molecules in comet Halley from the IKS-Vega experiment. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 1, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 353–358.

Combes, M., Moroz, V. I., Crovisier, J., Encrenaz, T., Bibring, J. P., Grigoriev, A. V., Sanko, N. F., Coron, N., Crifo, J. F., Gispert, R., Bockelée-Morvan, D., Nikolsky, Yu. V., Krasnopolsky, V. A., Owen, T., Emerich, C., Lamarre, J. M., and Rocard, F. 1988. The 2.5-12 μm spectrum of comet Halley from the IKS-VEGA experiment. *Icaurs* 76:404–436.

Crovisier, J. 1991. Radio spectroscopy of comets. In *Asteroids, Comets and Meteors 1991*, LPI Contrib. No. 765, p. 44 (abstract).

Crovisier, J., and Schloerb, F. P. 1991. The study of comets at radio wavelengths. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 149–173.

Crovisier, J., Despois, D., Bockelée-Morvan, D., Colom, P., and Paubert, G. 1991. Microwave observations of hydrogen sulfide and searches for other sulphur compounds in comets Austin (1989c1) and Levy (1990c). *Icarus* 93:246–258.

Cruikshank, D. P., and Brown, R. H. 1983. The nucleus of comet P/Schwassmann-Wachmann 1. *Icarus* 56:377–380.

Cruikshank, D. P., Hartmann, W. K., and Tholen, D. J. 1985. Colour, albedo and nucleus size of Halley's comet. *Nature* 315:122–124. Corrigendum: 315:690.

Degewij, J., and Tedesco, E. F. 1982. Do comets evolve into asteroids? Evidence from physical studies. In *Comets*, ed. L. L. Wilkening (Tucson: Univ. of Arizona Press), pp. 665–695.

Delamere, W. A., Reitsema, H. J., Huebner, W. F., Schmidt, H. U., Keller, H. U., Schmidt, W. K. H., Wilhelm, K., and Whipple, F. L. 1986. Radiometric observations of the nucleus of comet Halley. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 355–357.

Delsemme, A. H. 1973. Origin of the short-period comets. *Astron. Astrophys.* 29:377–381.

Delsemme, A. H. 1982. Chemical composition of cometary nuclei. In *Comets*, ed. L. L. Wilkening (Tucson: Univ. of Arizona Press), pp. 85–130.

Delsemme, A. H. 1991. Nature and history of the organic compounds in comets: An astrophysical view. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 377–428.

DiSanti, M., Mumma, M., Hoban, S., Reuter, D., Espenak, F., Lacy, J., and Parmar, R. 1990. A search for CO emission in comet Austin (1990c1). *Bull. Amer. Astron. Soc.* 22:1094 (abstract).

Donn, B., Daniels, P. A., and Hughes, D. W. 1985. On the structure of the cometary nucleus. *Bull. Amer. Astron. Soc.* 17:520 (abstract).

Duncan, M., Quinn, T., and Tremaine, S. 1988. The origin of short-period comets. *Astrophys. J. Lett.* 328:69–73.

Eberhardt, P., Dolder, U., Schulte, W., Krankowsky, D., Lämmerzahl, P., Hoffman, J. H., Hodges, R. R., Berthelier, J. J., and Illiano, J. M. 1986. The D/H ratio in water from Halley. In *20th ESLAB Symposium on the Exploration of Halley's*

Comet, vol. 1, eds. B. Battrick, E. J. Rolfe, and R. Reinhard, ESA SP-250, pp. 539–541.

Eberhardt, P., Krankowsky, D., Schulte, W., Dolder, U., Lämmerzahle, P., Berthelier, J. J., Woweries, J., Stubbeman, U., Hodges, R. R., Hoffman, J. H., and Illian, J. M. 1987. The CO and N_2 abundance in comet P/Halley. *Astron. Astrophys.* 187:481–484.

Eberhardt, P., Meier, R., Krankowsky, D., and Hodges, R. R. 1991. Methanol abundances in comet P/Halley from in-situ measurements. *Bull. Amer. Astron. Soc.* 23:1161 (abstract).

Encrenaz, T., and Knacke, R. 1991. Carbonaceous compounds in comets: Infrared observations. In *Comets in the Post-Halley Era,* eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 107–137.

Everhart, E. 1967. Intrinsic distributions of cometary perihelia and magnitudes. *Astron. J.* 72:1002–1011.

Everhart, E. 1972. The origin of short-period comets. *Astrophys. Lett.* 10:131–135.

Everhart, E. 1978. The evolution of comet orbits as perturbed by Uranus and Neptune. In *Comets, Asteroids, and Meteorites: Interrelations, Evolution and Origins,* ed. A. H. Delsemme (Toledo: Univ. of Toledo Press), pp. 99–104.

Fanale, F. P., and Salvail, J. R. 1984. An idealized short-period comet model: Surface insolation, H_2O flux, dust flux, and mantle evolution. *Icarus* 60:476–511.

Farquhar, R., Dunham, D. W., and Hsu, S. C. 1987. Alternative cometary targets for the Giotto extended mission. In *Diversity and Similarity of Comets*, eds. E. J. Rolfe and B. Battrick, ESA SP-278, pp. 727–731.

Feldman, P. D., Festou, M. C., A'Hearn, M. F., Arpigny, C., Butterworth, P. S., Cosmovici, C. B., Danks, A. C., Gilmozzi, R., Jackson, W. M., McFadden, L. A., Patriarchi, P., Schleicher, D. G., Tozzi, G. P., Wallis, M. K., Weaver, H. A., and Woods, T. N. 1987. IUE observations of comet P/Halley: Evolution of the ultraviolet spectrum between September 1985 and July 1986. *Astron. Astrophys.* 187:325–328.

Fernandez, J. A. 1980. On the existence of a comet belt beyond Neptune. *Mon. Not. Roy. Astron. Soc.* 192:481–491.

Fernandez, J. A. 1988. End-states of short-period comets and their role in maintaining the zodiacal dust cloud. *Moon and Planets* 41:155–161.

Fink, U. 1991. Comet Yanaka (1988r): A new class of carbon poor comet. *Bull. Amer. Astron. Soc.* 23:1160 (abstract).

French, L. M., Vilas, F., Hartmann, W. K., and Tholen, D. J. 1989. Distant asteroids and Chiron. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 880–920.

Geiss, J., Altwegg, K., Anders, E., Balsiger, H., Ip, W.-H., Meier, A., Neugebauer, M., Rosenbauer, H., and Shelley, E. G. 1991. Interpretation of the ion mass spectra in the mass per charge range 25–35 amu/e^- obtained in the inner coma of Halley's comet by the HIS-sensor of the Giotto IMS experiment. *Astron. Astrophys.* 247:226–234.

Glaccum, W., Moseley, S. H., Campins, H., and Loewenstein, R. F. 1987. Airborne spectrophotometry of P/Halley from 20 to 65 microns. *Astron. Astrophys.* 187:635–638.

Goldstein, R. M., Jurgens, R. F., and Sekanina, Z. 1984. A radar study of comet IRAS-Araki-Alcock 1983d. *Icarus* 89:1745–1754.

Gombosi, T. I., and Houpis, H. L. 1986. The icy-glue model of the cometary nucleus. *Nature* 324:43–44.

Greenberg, J. M. 1982. What are comets made of? A model based on interstellar dust. In *Comets,* ed. L. L. Wilkening (Tucson: Univ. of Arizona Press), pp. 163–189.

Greenberg, J. M., and Hage, J. I. 1990. From interstellar dust to comets: A unification

of observational constraints. *Astrophys. J.* 361:260–274.
Greenberg, J. M., Grim, R., and van Izendoorn, L. 1986. Interstellar S_2 in comets. In *Asteroids, Comets and Meteors II,* eds. C.-I. Lagerkvist, B. A. Lindblad, H. Lundstedt and H. Rickman (Uppsala: Uppsala Univ.), pp. 225–227.
Grün, E., and Jessberger, E. 1990. Dust. In *Physics and Chemistry of Comets,* ed. W. F. Huebner (Berlin: Springer Verlag), pp. 113–176.
Grün, E., Bar-Nun, A., Benkhoff, J., Bischoff, A., Düren, H., Hellermann, H., Hesselbarth, P., Hsiung, P., Keller, H. U., Klinger, J., Knölker, J., Kochan, H., Kohl, H., Kölzer, G., Krankowsky, D., Lämmerzahl, P., Maurersberger, K., Neukum, G., Oehler, A., Ratke, L., Roessler, K., Schwehm, G., Spohn, T., Stöffler, D., and Thiel, K. 1991. Laboratory simulation of cometary processes: Results from the first KOSI experiments. In *Comets in the Post-Halley Era,* eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 277–297.
Hahn, G., and Bailey, M. E. 1990. Rapid dynamical evolution of giant comet Chiron. *Nature* 348:132–136.
Halley, E. 1705. *A Synopsis of the Astronomy of Comets* (London).
Hamid, S. E., Marsden, B. G., and Whipple, F. L. 1968. Influence of a comet belt beyond Neptune on the motions of periodic comets. *Astron. J.* 73:727–729.
Hanner, M. S. 1984. Comet Cernis: Icy grains at last? *Astrophys. J. Lett.* 277:75–78.
Hanner, M. S., and Tokunaga, A. T. 1991. Infrared techniques for comet observations. In *Comets in the Post-Halley Era,* eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 67–91.
Hanner, M. S., Aitken, D. K., Knacke, R., McCorkle, S., Roche, P. F., and Tokunaga, A. T. 1985. Infrared spectrophotometry of comet IRAS-Araki-Alcock (1983d): A bare nucleus revealed? *Icarus* 62:97–109.
Hanner, M. S., Tokunaga, A. T., Golisch, W. F., Griep, D. M., and Kaminski, C. D. 1987. Infrared emission from P/Halley's dust coma during March 1986. *Astron. Astrophys.* 187:653–660.
Hanner, M. S., Newburn, R. L., Gehrz, R. D., Harrison, T., Ney, E. P., and Heyward, T. L. 1990. The infrared spectrum of comet Bradfield (1987s) and the silicate emission feature. *Astrophys. J.* 348:312–321.
Harmon, J. K., Campbell, D. B., Hine, A. A., Shapiro, I. I., and Marsden, B. G. 1989. Radar observations of comet IRAS-Araki-Alcock 1983d. *Astrophys. J.* 338:1071–1093.
Hartmann, W. K., Cruikshank, D. P., and Degewij, J. 1982. Remote comets and related bodies: VJHK colorimetry and surface materials. *Icarus* 52:377–408.
Hartmann, W. K., Tholen, D. J., and Cruikshank, D. P. 1987. The relationship of active comets, "extinct" comets, and dark asteroids. *Icarus* 69:33–50.
Hartmann, W. K., Tholen, D. J., Meech, K. J., and Cruikshank, D. P. 1989. 2060 Chiron: Colorimetry and possible cometary behavior. *Icarus* 83:1–15.
Herter, T., Campins, H., and Gull, G. E. 1987. Airborne spectrophotometry of P/Halley from 16 to 30 microns. *Astron. Astrophys.* 187:629–631.
Hoban, S., Mumma, M., Reuter, D. C., DiSanti, M., Joyce, R. R., and Storrs, A. 1991. A tentative identification of methanol as the progenitor of the 3.52 μm emission feature in several comets. *Icarus* 93:122–134.
Horanyi, M., Gombosi, T. I., Cravens, T. E., Korosmezey, A., Kecskeméty, K., Nagy, A., and Szegö, K. 1984. The friable sponge model of a cometary nucleus. *Astrophys. J.* 278:449–455.
Huebner, W. F., Snyder, L. E., and Buhl, D. 1974. HCN radio emission from comet Kohoutek (1973f). *Icarus* 23:580–584.
Huebner, W. F., Boice, D. C., Schmidt, H. U., and Wegmann, R. 1991. Structure of the coma: Chemistry and solar wind interaction. In *Comets in the Post-Halley Era,* eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp.

907–936.

Ip, W.-H., Rosenbauer, H., Schwenn, R., Balsiger, H., Geiss, J., Meier, A., Goldstein, B. E., Lazarus, A. J., Shelley, E., and Kettmann, G. 1990. Giotto IMS measurements of the production rate of hydrogen cyanide in the coma of comet Halley. *Ann. Geophys.* 8:319–325.

Jessberger, E. K., and Kissel, J. 1991. Chemical properties of cometary dust and a note on carbon isotopes. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 1075–1092.

Jessberger, E. K., Christoforidis, A., and Kissel, J. 1988. Aspects of the major element composition of Halley's dust. *Nature* 332:691–695.

Jewitt, D. C., and Luu, J. X. 1989. A CCD portrait of comet P/Tempel 2. *Astron. J.* 97:1766–1790.

Jewitt, D. C., and Luu, J. X. 1990. The persistent coma of comet P/Schwassmann-Wachmann 1. *Astrophys. J.* 351:277–286.

Jewitt, D. C., and Meech, K. J. 1988. Optical properties of cometary nuclei and a preliminary comparison with asteroids. *Astrophys. J.* 328:974–986.

Johnson, R. E., Cooper, J. F., Lanzerotti, L. J., and Strazzula, G. 1987. Radiation formation of a non-volatile comet crust. *Astron. Astrophys.* 187:889–892.

Joss, P. C. 1973. On the origin of short-period comets. *Astron. Astrophys.* 25:271–273.

Kamoun, P. G., Campbell, D. B., Ostro, S. J., Pettengill, G. H., and Shapiro, I. I. 1982. Comet Encke: Radar detection of nucleus. *Science* 216:293–295.

Kawara, K., Gregory, B., Yamamoto, T., and Shibai, H. 1988. Infrared spectroscopic observation of methane in comet P/Halley. *Astron. Astrophys.* 207:174–181.

Kazimirchak-Polonskaya, E. I. 1972. The major planets as powerful transformers of cometary orbits. In *The Motion, Evolution of Orbits, and Origin of Comets*, eds. G. A. Chebotarev and E. I. Kazimirchak-Polonskaya (Dordrecht: D. Reidel), pp. 373–397.

Keller, H. U. 1990. The nucleus. In *Physics and Chemistry of Comets*, ed. W. F. Huebner (Berlin: Springer Verlag), pp. 13–68.

Kim, S. J., and A'Hearn, M. F. 1991. Upper limits of SO and SO_2 in comets *Icarus* 90:79–95.

Kim, S. J., A'Hearn, M. F., and Larson S. M. 1990. Multi-cycle fluorescence: Application to S_2 in comet IRAS-Araki-Alcock 1983d. *Icarus* 87:440–451.

Kissel, J., Sagdeev, R. Z., Bertaux, J. L., Angarov, V. N., Audouze, J., Blamont, J. E., Büchler, K., Evlanov, E. N., Fechtig, H., Fomenkova, M. N., von Hoerner, H., Inogamov, N. A., Khromov, V. N., Knabe, W., Krueger, F. R., Langevin, Y., Leonas, V. B., Levasseur-Regourd, A. C., Managadze, G. G., Podkolzin, S. N., Shapiro, V. D., Tabaldyev, S. R., and Zubkov, B. V. 1986*a*. Composition of comet Halley dust particles from Vega observations. *Nature* 321:280–282.

Kissel, J., Brownlee, D. E., Büchler, K., Clark, B. C., Fechtig, H., Grün, E., Hornung, K., Igenbergs, E. B., Jessberger, E. K., Krueger, F. R., Kuczera, H., McDonnell, J. A. M., Morfill, G. M., Rahe, J., Schwehm, G. H., Sekanina, Z., Utterback, N. G., Völk, H. J., and Zook, H. A. 1986*b*. Composition of comet Halley dust particles from Giotto observations. *Nature* 321:336–338.

Kleine, M., Wyckoff, S., Wehinger, P. A., and Peterson, B. A. 1991. The carbon isotope abundance ratios in comets. *Bull. Amer. Astron. Soc.* 23:1166 (abstract).

Kowal, C. T., Liller, W., and Marsden, B. G. 1979. The discovery and orbit of (2060) Chiron. In *Dynamics of the Solar System*, ed. R. L. Duncombe (Dordrecht: D. Reidel), pp. 245–250.

Krankowsky, D. 1991. The composition of comets. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 855–877.

Krankowsky, D., Lämmerzahl, P., Herrwerth, I., Woweries, J., Eberhardt, P., Dolder,

U., Herrmann, U., Schulte, W., Berthelier, J. J., Illiano, J. M., Hodges, R. R., and Hoffman, J. H. 1986. In situ gas and ion composition measurements at comet Halley. *Nature* 321:326–329.

Krasnopolsky, V. A., and Tkachuk, A. Yu. 1991. TKS-Vega experiment: NH and NH_2 bands in comet Halley. *Astron. J.* 101:1915–1919.

Kresak, L. 1985. The aging and lifetimes of comets. In *Dynamics of Comets: Their Origin and Evolution,* eds. A. Carusi and G. B. Valsecchi (Dordrecht: D. Reidel), pp. 279–302.

Krueger, F. R., and Kissel, J. 1987. The chemical composition of the dust in comet P/Halley as measured by "PUMA" on board Vega-1. *Naturwissenchaften* 74:312–316.

Krueger, F. R., Korth, A., and Kissel, J. 1990. The organic matter of comet P/Halley as inferred by joint gas phase and solid phase analyses. *Space Sci. Rev.* 56:167–175.

Kuiper, G. P. 1951. On the origin of the solar system. In *Astrophysics,* ed. J. A. Hynek (New York: McGraw Hill), pp. 357–424.

Larson, H. P., Weaver, H. A., Mumma, M. J., and Drapatz, S. 1989. Airborne infrared spectroscopy of comet Wilson (1986l) and comparisons with comet Halley. *Astrophys. J.* 338:1106–1114.

Lebofsky, L. A., and Spencer, J. R. 1989. Radiometry thermal modeling of asteroids. In *Asteroids II,* eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 128–147.

Lebofsky, L. A., Tholen, D. J., Rieke, G. H., and Lebofsky, M. J. 1984. 2060 Chiron: Visual and thermal infrared observations. *Icarus* 60:532–537.

Levison, H. F. 1991. The long-term dynamical behavior of small bodies in the Kuiper belt. *Astron. J.* 102:787–794..

Levison, H. F., and Duncan, M. J. 1990. A search for proto-comets in the outer regions of the solar system. *Astron. J.* 100:1669–1675.

Lewis, J. S. 1974. The temperature gradient in the solar nebula. *Science* 186:440–443.

Lowery, B. E. 1973. The effect of multiple encounters on short-period comet orbits. *Astron. J.* 78:428–437.

Luu, J., and Jewitt, D. 1990. The nucleus of comet P/Encke. *Icarus* 86:69–81.

Maas, R. W., Ney, E. P., and Woolf, N. F. 1970. The 10-micron emission peak of comet Bennett 1969i. *Astrophys. J. Lett.* 160:101–104.

Magee-Sauer, K., Scherb, F., Roesler, F. L., Harlander, J., and Lutz, B. L. 1989. Fabry-Perot observations of the NH_2 emission from comet Halley. *Icarus* 82:50–60.

Marconi, M. L., Mendis, D. A., Korth, A., Pin, R. P., Mitchell, D. L., and Reme, H. 1990. The identification of H_3S^+ with the ion of mass per charge (m/q) 35 observed in the coma of comet Halley. *Astrophys. J. Lett.* 352:17–120.

Marsden, B. G. 1970. On the relationship between comets and minor planets. *Astron. J.* 75:206–217.

Marsden, B. G. 1992. *Catalogue of Cometary Orbits,* 7th ed. (Cambridge, Mass.: Smithsonian Astrophysical Obs.).

Marsden, B. G., Sekanina, Z., and Yeomans, D. K. 1973. Comets and nongravitational forces. V. *Astron. J.* 78:211–225.

McDonnell, J. A. M., Lamy, P. L., and Pankiewicz, G. S. 1991. Physical properties of cometary dust. In *Comets in the Post-Halley Era,* eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 1043–1073.

McFadden, L. A., Ostro, S. J., Barker, E. S., Cochran, A. L., Cruikshank, D. P., Hartmann, W. K., Soifer, B. T., and Veeder, G. J. 1984. 2201 Oljato: An asteroid, a comet, or both? *Bull. Amer. Astron. Soc.* 16:691 (abstract).

McFadden, L. A., Tholen, D. J., and Veeder, G. J. 1989. Physical properties of Aten, Apollo and Amor asteroids. In *Asteroids II,* eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 442–467.

Meech, K. J., and Belton, M. J. S. 1989. 2060 Chiron. *IAU Circ.* 4770.

Meech, K. J., and Belton, M. J. S. 1990. The atmosphere of 2060 Chiron. *Astron. J.* 100:1323–1338.

Mendis, D. A., and Brin, G. D. 1977. The monochromatic brightness variations of comets—II. The core-mantle model. *Moon and Planets* 17:359–372.

Millis, R. L., A'Hearn, M. F., and Campins, H. C. 1988. The investigation of the nucleus and coma of comet P/Arend-Rigaux. *Astrophys. J.* 324:1194–1209.

Millman, P. M. 1977. The chemical composition of cometary meteoroids. In *Comets, Asteroids, Meteorites: Interrelations, Evolution and Origins*, ed. A. H. Delsemme (Toledo: Univ. of Toledo Press), pp. 127–132.

Möhlmann, D., Börner, H., Danz, M., Elter, G., Mangoldt, T., Rubbert, B., and Weidlich, U. 1986. Physical properties of P/Halley derived from Vega images. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 339–340.

Mumma, M. J., and Reuter, D. C. 1989. On the identification of formaldehyde in Halley's comet. *Astrophys. J.* 344:940–948.

Mumma, M. J., Weaver, H. A., Larson, H. P., Davis, D. S., and Williams, M. 1986. Detection of water vapor in Halley's comet. *Science* 232:1523–1528.

Mumma, M. J., Reuter, D., and Magee-Sauer, K. 1990. Heterogeneity of the nucleus of comet Halley. *Bull. Amer. Astron. Soc.* 22:1088 (abstract).

Mumma, M. J., Stern, S. A., and Weissman, P. R. 1993. Comets and the origin of the solar system: Reading the Rosetta stone. In *Protostars and Planets III*, eds. E. H. Levy and J. I. Lunine (Tucson: Univ. of Arizona Press), pp. 1177–1252.

Newburn, R. L., and Spinrad, H. 1989. Spectrophotometry of 25 comets: Post-Halley updates for 17 comets plus new observations for eight additional comets. *Astron. J.* 97:552–569.

Newton, H. A. 1893. On the capture of comets by planets, especially their capture by Jupiter. *Mem. Natl. Acad. Sci.* 6:7–23.

Ney, E. P. 1982. Optical and infrared observations of comets in the range 0.5 μm to 20 μm. In *Comets*, ed. L. L. Wilkening (Tucson: Univ. of Arizona Press), pp. 323–340.

Oort, J. H. 1950. The structure of the cloud of comets surrounding the solar system and a hypothesis concerning its origin. *Bull. Astron. Inst. Neth.* 11:91–110.

Öpik, E. J. 1963. The stray bodies in the solar system. Part I. Survival of cometary nuclei and the asteroids. *Astron. Astrophys.* 2:219–262.

Osip, D. J., Schleicher, D. G., Millis, R. L., A'Hearn, M. F., and Birch, P. V. 1991. 15 yr of comet photometry: A comparative analysis of 80 comets. In *Asteroids, Comets, and Meteors 1991*, LPI Contrib. No. 765, p. 160 (abstract).

Peale, S. J. 1989. On the density of Halley's comet. *Icarus* 82:36–49.

Prialnik, D., and Bar-Nun, A. 1988. The formation of a permanent dust mantle and its effect on cometary activity. *Icarus* 74:272–283.

Prinn, R. G., and Fegley, B., Jr. 1989. Solar nebula chemistry: Origin of planetary, satellite, and cometary volatiles. In *Origin and Evolution of Planetary and Satellite Atmospheres*, eds. S. K. Atreya, J. B. Pollack and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 78–136.

Quinn, T., Tremaine, S., and Duncan, M. 1991. Planetary perturbations and the origin of short-period comets. *Astrophys. J.* 355:667–679.

Reuter, D. C. 1992. The contribution of methanol to the 3.4 μm emission feature in comets. *Astrophys. J.* 386:330–335.

Rickman, H. 1986. Masses and densities of comets Halley and Kopff. In *Comet Nucleus Sample Return Mission*, ed. O. Melita, ESA SP-249, pp. 195–205.

Rickman, H. 1989. The nucleus of comet Halley: Surface structure, mean density, gas and dust production. *Adv. Space Res.* 9(3):59–71.

Rickman, H. 1991. The thermal history and structure of cometary nuclei. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 733–760.

Rickman, H., and Fernandez, J. A. 1986. Formation and blowoff of a cometary dust mantle. In *The Comet Nucleus Sample Return Mission*, ed. O. Melita, ESA SP-249, pp. 185–194.

Ryan, E. V., and Campins, H. 1991. Comet Halley: Spatial and temporal variability of the silicate emission feature. *Astron. J.* 101:695–705.

Sagdeev, R. Z., Avanesov, G. A., Shamis, V. A., Szegö, K., Merenyi, E., Smith, B. A., Ziman, Ya. L., Krasikov, V. A., Tarnopolsky, V. A., and Kuzim, A. A. 1986*a*. TV experiment in Vega mission: Image processing technique and some results. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 295–305.

Sagdeev, R. Z., Avanesov, G. A., Ziman, Ya. L., Smith, B., Toth, I., Moroz, V. I., Tarnopolsky, V. I., Zhukov, B. S., and Shamis, V. A. 1986*b*. TV experiment of the Vega mission: photometry of the nucleus and inner coma. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 317–326.

Sagdeev, R. Z., Elyasberg, P. E., and Moroz, V. I. 1987. Is the nucleus of comet Halley a low density body? *Nature* 331:240–242.

Schleicher, D. G., Millis, R. L., and Birch, P. V. 1987. Photometric observations of comet P/Giacobini-Zinner. *Astron. Astrophys.* 187:531–538.

Schleicher, D. G., Bus, S. J., and Osip, D. J. 1991. The anomalous molecular abundances of comet P/Wolf-Harrington. In *Asteroids, Comets and Meteors 1991*, LPI Contrib. No. 765, p. 185 (abstract).

Schwehm, G. H., and Langevin, Y. 1991. *Rosetta: A Comet Nucleus Sample Return Mission*, ESA SP-1125.

Sekanina, Z. 1990. Rotation vector of Halley's comet. In *Comet Halley: Investigations, Results, Interpretations*, vol. 2, ed. J. W. Mason (Chichester: Ellis Horwood), pp. 189–201.

Sekanina, Z. 1991. Comprehensive model for the nucleus of periodic comet Tempel 2 and its activity. *Astron. J.* 102:350–388.

Shapiro, I. I., Marsden, B. G., Whipple, F. L., Campbell, D. B., Harmon, J. K., and Hine, A. A. 1983. Radar observations of comets IRAS-Araki-Alcock and Sugano-Saigusa-Fujikawa. *Bull. Amer. Astron. Soc.* 15:800 (abstract).

Smoluchowski, R. 1981. Amorphous ice and the behavior of cometary nuclei. *Astrophys. J. Lett.* 244:31–34.

Snyder, L. E., Palmer, P., and de Pater, I. 1990. Observations of formaldehyde in comet Machholz (1988j). *Icarus* 86:289–298.

Spinrad, H., Stauffer, J., and Newburn, R. L., Jr. 1979. Optical spectrophotometry of comet Tempel 2 far from the Sun. *Publ. Astron. Soc. Pacific* 91:707–711.

Stagg, C. R., and Bailey, M. E. 1989. Stochastic capture of short-period comets. *Mon. Not. Roy. Astron. Soc.* 241:507–541.

Stern, S. A. 1986. The effects of mechanical interaction between the interstellar medium and comets. *Icarus* 68:276–283.

Stern, S. A. 1989. Implications of volatile release from object 2060 Chiron. *Publ. Astron. Soc. Pacific* 101:126–132.

Stern, S. A. 1990. ISM induced erosion and gas dynamical drag in the Oort cloud. *Icarus* 84:447–466.

Stern, S. A., and Shull, J. M. 1988. The thermal evolution of comets in the Oort cloud by passing stars and stochastic supernovae. *Nature* 332:407–411.

Sykes, M. V., and Walker, R. G. 1991. Constraints on the diameter and albedo of 2060 Chiron. *Science* 251:777–780.

Sykes, M. V., and Walker, R. G. 1992. Cometary dust trails, I: Survey. *Icarus* 95:180–210.

Sykes, M. V., Lebofsky, L. A., Hunten, D. M., and Low, F. 1986. The discovery of dust trails in the orbits of periodic comets. *Science* 232:1115–1117.

Tholen, D. J., Hartmann, W. K., and Cruikshank, D. P. 1988. (2060) Chiron. *IAU Circ.* 4554.

Thomas, N., and Keller, H. U. 1989. The colour of comet Halley's nucleus and dust. *Astron. Astrophys.* 213:487–494.

Tokunaga, A. T., and Hanner, M. S. 1985. Does comet P/Arend-Rigaux have a large dark nucleus? *Astrophys. J.* 296:L13–L16.

Torbett, M. V. 1989. Chaotic motion in a comet disk beyond Neptune: The delivery of short-period comets. *Astron. J.* 98:1477–1481.

Torbett, M. V., and Smoluchowski, R. 1990. Chaotic motion in a primordial comet disk beyond Neptune and comet influx. *Nature* 345:49–51.

Tsou, P. D., Brownlee, D., and Albee, A. 1984. Experiments on intact capture of hypervelocity particles. *Lunar Planet. Sci.* XV:866–867 (abstract).

Ulich, B. L., and Conklin, E. J. 1974. Detection of methyl cyanide in comet Kohoutek. *Nature* 248:121–122.

Vanysek, V. 1991. Isotopic ratios in comets. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer, J. Rahe (Dordrecht: Kluwer), pp. 879–895.

Veeder, G., Hanner, M. S., and Tholen, D. J. 1987. The nucleus of comet P/Arend-Rigaux. *Astron. J.* 94:169–173.

Veverka, J., Langevin, Y., Farquhar, R., and Fulchignoni, M. 1989. Spacecraft exploration of asteroids: The 1988 perspective. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 970–993.

Weaver, H. A., Mumma, M. J., Larson, H. P., and Davis, D. S. 1986. Post-perihelion observations of water in comet Halley. *Nature* 324:441–444.

Weissman, P. R. 1979. Physical and dynamical evolution of long-period comets. In *Dynamics of the Solar System*, ed. R. L. Duncombe (Dordrecht: D. Reidel), pp. 277–282.

Weissman, P. R. 1982. Terrestrial impact rates for long and short-period comets. In *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, eds. L. T. Silver and P. H. Schultz, Geological Soc. of America SP-190 (Boulder: Geological Soc. of America), pp. 15–24.

Weissman, P. R. 1986*a*. Are cometary nuclei primordial rubble piles? *Nature* 320:242–244.

Weissman, P. R. 1986*b*. How pristine are cometary nuclei? In *The Comet Nucleus Sample Return Mission*, ed. O. Melita, ESA SP-249, pp. 15–25.

Weissman, P. R. 1987*a*. How typical is Halley's comet? In *Diversity and Similarity of Comets*, eds. E. J. Rolfe and B. Battrick, ESA SP-278, pp. 31–36.

Weissman, P. R. 1987*b*. Post-perihelion brightening of Halley's comet: Spring time for Halley. *Astron. Astrophys.* 187:873–878.

Weissman, P. R. 1990. The Oort cloud. *Nature* 344:825–830.

Weissman, P. R. 1991*a*. The cometary impactor flux at the Earth. In *Global Catastrophes in Earth History*, eds. V. Sharpton and P. Ward, Geological Soc. of America SP-247 (Boulder: Geological Soc. of America), pp. 171–180.

Weissman, P. R. 1991*b*. Why did Halley hiccup? *Nature* 353:793–794.

Weissman, P. R., and Kieffer, H. H. 1981. Thermal modeling of cometary nuclei. *Icarus* 47:302–311.

Weissman, P., and Neugebauer, M. 1992. The comet rendezvous asteroid flyby mission: A status report. In *Asteroids, Comets and Meteors 1991*, eds. A. Harris and E. Bowell (Houston: Lunar and Planetary Inst.), pp. 629–632,

Weissman, P. R., and Stern, S. A. 1993. Physical processing of cometary nuclei. In

Workshop on Analysis of Returned Comet Nucleus Samples, ed. S. Chang, NASA, in press.

Weissman, P. R., A'Hearn, M. F., McFadden, L. A., and Rickman, H. 1989. Evolution of comets into asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 880–920.

West, R. M. 1990. 2060 Chiron. *IAU Circ.* 4970.

Wetherill, G. W. 1971. Cometary versus asteroidal origin of chondritic meteorites. In *Physical Studies of Minor Planets*, ed. T. Gehrels, NASA SP-267, pp. 447–460.

Wetherill, G. W. 1975. Late heavy bombardment of the moon and terrestrial planets. In *Proc. Lunar Sci. Conf.* 6:1539–1561.

Wetherill, G. W. 1988. Where do the Apollo objects come from? *Icarus* 76:1–18.

Wetherill, G. W. 1991. End products of cometary evolution: Cometary origin of Earth-crossing bodies of asteroidal appearance. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 537–556.

Whipple, F. L. 1950. A comet model I: The acceleration of comet Encke. *Astrophys. J.* 111:375–394.

Whipple, F. L. 1964. Evidence for a comet belt beyond Neptune. *Proc. Natl. Acad. Sci. U. S.* 51:711–718.

Wilhelm, K., Cosmovici, C. B., Delamere, W. A., Huebner, W. F., Keller, H. U., Reitsema, H., Schmidt, H. U., and Whipple, F. L. 1986. A three-dimensional model of the nucleus of comet Halley. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 367–369.

Wisniewski, W. Z. 1988. Periodic comet Tempel 2. *IAU Circ.* 4603.

Wisniewski, W. Z., Fay, T., and Gehrels, T. 1986. Light variations in comets. In *Asteroids, Comets, Meteors II*, eds. C.-I. Lagerkvist, B. Lindblad, H. Lundstedt and H. Rickman (Uppsala: Univ. of Uppsala), pp. 337–339.

Woods, T. N., Feldman, P. D., Dymond, K. F., and Sahnow, D. J. 1986. Rocket ultraviolet spectroscopy of comet Halley and abundance of carbon monoxide and carbon. *Nature* 324:436–438.

Woods, T. N., Feldman, P. D., and Dymond, K. F. 1987. The atomic carbon distribution in the coma of comet P/Halley. *Astron. Astrophys.* 187:380–384.

Wyckoff, S., and Theobald, J. 1989. Molecular ions in comets. *Adv. Space Res.* 9(3):157–161.

Wyckoff, S., Lindholm, E., Wehinger, P. A., Peterson, B. A., Zucconi, J.-M., and Festou, M. C. 1989. The $^{12}C/^{13}C$ abundance ratio in comet Halley. *Astrophys. J.* 339:488–500.

Wyckoff, S., Tegler, S. C., and Engel, L. 1991. Ammonia abundances in four comets. *Astrophys. J.* 368:279–286.

Yamamoto, T. 1991. Chemical theories on the origin of comets. In *Comets in the Post-Halley Era*, eds. R. L. Newburn, M. Neugebauer and J. Rahe (Dordrecht: Kluwer), pp. 361–376.

Yeomans, D. K. 1986. Physical interpretations from the motions of comets Halley and Giacobini-Zinner. In *20th ESLAB Symposium on the Exploration of Halley's Comet*, vol. 2, eds. B. Battrick, E. J. Rolfe and R. Reinhard, ESA SP-250, pp. 419-425.

Yeomans, D. K. 1991. A comet among the near-Earth asteroids? *Astron. J.* 101:1920–1928.

Yeomans, D. K., and Chodas, P. W. 1989. An asymmetric outgassing model for cometary nongravitational accelerations. *Astron. J.* 98:1083–1093.

ROLE OF NEAR-EARTH ASTEROIDS IN THE SPACE EXPLORATION INITIATIVE

DONALD R. DAVIS
Planetary Science Institute

ALAN L. FRIEDLANDER
Science Applications International Corporation

and

THOMAS D. JONES
NASA Johnson Space Center

There exist frequent, low ΔV mission opportunities to many bodies in the complete near-Earth asteroid (NEA) population: about 250 of these asteroids larger than 1 km diameter are easier to reach than the Moon's surface. Opportunities for fast missions of less than one year duration that require less ΔV than a round-trip lunar mission occur about 9 to 10 times each year on average to a near-Earth asteroid larger than 0.5 km in diameter. However, less than 10% of the approximately 1,700 NEAs larger than 1 km have been discovered through mid-1991. Hence, a significant increase in the discovery rate of NEAs is needed to realize the mission potential in the NEA population. This point is illustrated by the fact that the best candidate (by far) for a fast-trip mission is the recently discovered asteroid 1991JW which was unknown at the time of the Tucson Conference on "Resources of Near-Earth Space," in January 1991. Given that there are many opportunities to visit these asteroids and that a large investment will be made in a Mars landing mission as part of the Space Exploration Initiative (SEI), a program of robotic and astronautic missions to NEAs would provide significant additional return on that investment by: (1) providing a precursor to a Mars mission with flight testing of hardware and mission operations (other than those needed for Mars surface operations); (2) providing a highly visible program milestone, sustaining momentum for the SEI in the interval between lunar base establishment and the Mars landing mission; (3) providing a large science return through *in situ* observations and returned samples; and (4) carrying out an assessment of the resources available in the NEA population that could be used for any future large-scale space development.

I. INTRODUCTION

A broad new initiative for the expansion of human beings into space has recently been mandated by the President of the United States. This endeavor is the first top-down directive to the U. S. space program since the ringing 1961 challenge of John F. Kennedy to "place a man on the Moon and return

him safely to the Earth." The focus of this new initiative has been on the Moon and Mars; indeed, the initial response of NASA as set forth in its 90-day study report (NASA 1989*b*) is a series of options regarding the sizing and timing of a program to establish an outpost on the Moon and to send manned missions to Mars. Near-Earth asteroids were not considered in this initial NASA response to the new directive, perhaps due to their relative obscurity compared with the Moon and Mars, or perhaps due to a misperception that they are difficult to get to on any regular basis.

However, as we develop strategies and design options for the Space Exploration Initiative (SEI), all possible nearby targets must be considered. To this end a NASA-commissioned study was carried out to assess the role that the population of near-Earth asteroids (NEAs) could play in the SEI (Davis et al. 1990). It was argued in this study that the NEAs are a fascinating population of bodies that are interesting from a variety of perspectives: solar system science, space resources, and collisional hazards to the Earth. The authors also claimed that these asteroids provide exciting targets for SEI missions that naturally fit into strategies now being considered for the Moon and Mars by providing meaningful targets for precursors to a Mars landing mission—the "Apollo 8" mission of SEI. The results and recommendations of the Davis et al. report were presented to the Synthesis Group, a committee commissioned by the National Space Council as part of an outreach program designed to acquire new ideas and concepts to be incorporated into SEI. Apparently missions to NEAs struck a responsive chord with this committee and were put forth as an option in Architecture II, which has a science emphasis (see Stafford et al. 1991 for a description of all SEI architectures). Just as appropriate, though, is a role for NEAs in Architecture IV, emphasizing space resources in all scenarios that culminate in Mars landing, since NEA missions provide an excellent and realistic precursor before a Mars landing mission. In this chapter we develop the case why NEA missions should be an integral part of the SEI, show that excellent mission opportunities exist to the *known* NEA population, and note that additional, even better, opportunities will almost certainly be discovered in the near future given the increasing discovery rate of NEAs.

What are near-Earth asteroids? NEAs are a population of small bodies, ranging from nearly 40 km to sub-kilometer in diameter, moving on orbits approaching (or crossing) that of the Earth. A formal definition that we adopt here is that used by Helin and Shoemaker (1979): the object's perihelion must be <1.3 AU for it to be classified as a near-Earth asteroid. The first NEA, 433 Eros, was discovered in 1898 by G. Witt of Berlin. Discoveries continued at a slow rate for the next few decades—only 14 were known by 1950, and just 28 by 1970. However, the discovery rate has increased significantly in the last decade due to the dedicated efforts of search programs led by E. Helin of the Jet Propulsion Laboratory, by E. M. Shoemaker of the United States Geological Survey, and, more recently, by T. Gehrels' Spacewatch program at the University of Arizona. Currently (July, 1991), there are 178 known

NEAs, with 20 having been discovered in 1991 as of July 1, and 25 in 1990, which was a banner year for their discovery. Of these 178 asteroids, only a small fraction have been studied in sufficient detail to derive properties other than their orbits and very approximate sizes.

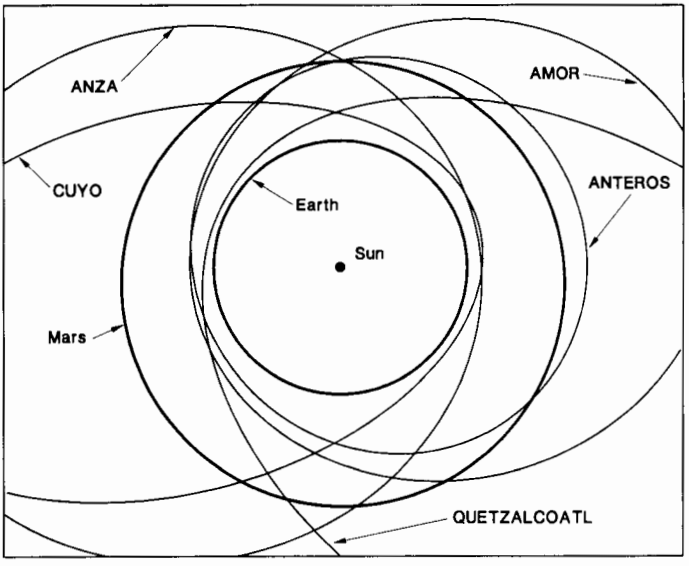


Figure 1. Comparative sizes and shapes of the orbits of Earth and Mars with those of five near-Earth asteroids.

The orbits of near-Earth asteroids are quite diverse, ranging from those with low inclination i and modest eccentricities e to those with high inclinations and large eccentricities (Fig. 1). Most NEAs have aphelia that lie within the region of the main asteroid belt, between Mars and Jupiter. Individual NEAs are transient bodies on the geologic (and astronomical) time scale: because their orbits bring them close to Earth, Mars and Venus, they will occasionally have their orbits significantly perturbed by the gravity of Earth (or Mars or Venus) or may even hit one of these planets. Calculations have shown that on the time scale of a several tens of Myr, the population would be significantly reduced by such planetary encounters. This time scale is short compared with the 4.6 Gyr age of the solar system, so, clearly, the NEAs are not a remnant population from the formation of the solar system; rather they are continually being resupplied from several sources. Their orbits point to the main asteroid belt as the leading candidate for the principal resupply source. Another source for NEAs is the population of short-period comets: current evidence suggests that these comets become inactive, i.e., lose their ability to blow off dust and gas, after about 1000 or so orbits around the Sun. What remains is probably an extinct core, a volatile-depleted (at least

in its outer layers), very dark body with an orbit similar to those observed for some NEAs. The fraction of NEAs that comes from the asteroid belt and from extinct comets is a topic of current research (Hartmann 1987). Almost certainly, both sources contribute to the observed NEA population; however, it is widely accepted that a majority of the NEAs are derived from the main asteroid region as fragments of collisionally broken rock and/or metal parent asteroids. Bodies from this source are more likely to have low-e, low-i orbits, and hence be more accessible to spacecraft.

NEAs are fascinating scientifically because many are samples of main-belt asteroids which are, in turn, most likely the debris left over when a planet failed to form in the region between Mars and Jupiter. Groundbased telescopic studies of asteroid spectra have identified three or four main compositional classes, and half a dozen or more minor ones. The NEAs are a "grab-bag" mixture of different asteroid classes from the main belt plus perhaps occasional new objects not identified so far in the main belt. The most common class in the outer belt is the black C class, thought to be similar to the carbonaceous chondrite meteorites. The most common class in the inner belt is the S class, apparently composed of common silicate rock-forming minerals, plus metal. One of the most fundamental controversies in asteroid studies is the nature of S class asteroids. Are they the source bodies of the ordinary chondrite meteorites: primitive, never-melted stones that are the most common type of meteorite? Are they collisionally exposed metal-rich cores of bodies that partially melted in the past? Or do both types exist within the S population? The answer to the S-class question could be provided by missions to S-class NEAs, and would provide a major breakthrough in understanding the small bodies of the solar system. Additional scientific rationale for missions to NEAs is given in the recent study by the Discovery Science Working Group (Discovery 1991) on the Near-Earth Asteroid Rendezvous Mission (NEAR Science Working Group 1986).

Potential resources offer another compelling incentive to study the NEAs. Water is very likely to be present in abundance on a significant fraction of these bodies, especially the C-class objects and extinct comet nuclei. Indeed, these objects are likely to be the only significant sources of water between the oceans of Earth and the frozen deposits on Mars (polar caps, permafrost, and water of hydration), and will be a water source that is not in a deep gravity well. This is especially relevant, given the most recent evidence from spectroscopic studies by Bell et al. (1989) and the Soviet Phobos-2 mission that the surface soils of the Martian satellites Phobos and Deimos are completely dehydrated. The water in NEAs may be present as hydrated minerals for asteroids derived from the main asteroid region; Jones et al. (1990) find that about two-thirds of the main-belt C asteroids show spectral evidence for water of hydration. Or the water may be in the form of ice in the interior of asteroids or inactive comets expected to be found in the NEA population. Another potential resource from NEAs is to provide bulk mass far from deep planetary gravity wells, needed for shielding large space structures in Earth orbit. Further aspects of

the resource potential inherent in NEAs are found in the recent book by Lewis and Lewis (1987), *Space Resources: Breaking the Bonds of Earth.*

The impact hazard from by NEAs is perhaps the most "down-to-Earth" reason for wanting to learn more about this population of objects. There has been a rising awareness of the threat of widespread destruction on Earth if it is hit by even a small NEA. That such impacts occur is without dispute: Meteor Crater in Arizona was formed by a very small NEA hitting the Earth about 40,000 yr ago. The Tunguska explosion in Siberia in 1908 and a grazing encounter by a house-sized body in 1972, which was recorded by a wide variety of instruments ranging from cameras on the ground to satellites in Earth orbit as the meteoroid skimmed through the upper atmosphere, confirm the frequency of close encounters. Studies of the iridium anomaly in the Cretaceous/Tertiary boundary layer indicate that a large NEA hit the Earth at that time, and perhaps was instrumental in eradicating dinosaurs and most other species from the Earth. Other studies suggest that the environmental stress resulting from NEA impacts has been a fundamental factor in biological evolution. Obviously, we should clarify the precise nature of the NEA threat and define plausible actions needed to eliminate, or at least mitigate, the threat of an approaching NEA by altering its orbit. To do that, however, we must understand more about the range of physical attributes that NEAs may be expected to have. However, as documented in the recent book by Chapman and Morrison (1989), we must first define the details of the threat by inventorying the population of NEAs down to a small size and determining their orbits precisely.

A final compelling reason for studying NEAs, and the one that is the focus of the rest of this Chapter, is the role that they could play in demonstrating mission hardware and techniques for the SEI. We shall make the case that manned missions to near-Earth asteroids would be excellent precursors to sending humans to Mars. Earlier studies by Science Application International Corporation (SAIC) (cf. Sec. IV) have shown that there are indeed asteroids that are more accessible, particularly for round trips, than the lunar surface. The frequency of opportunities to go to these objects depends on their numbers. A necessary step in understanding how NEAs might fit into the SEI is to explore the ΔV requirements and mission frequency for both minimum ΔV rendezvous missions and for short trip time missions (less than one year in duration desired for astronaut missions) to both the *known* NEA population and the *complete* population. To accomplish this step, in Sec. II we estimate the size and orbit distributions of the complete NEA population based on the sample of these bodies currently known. In Sec. III we examine both minimum energy and "fast" missions, considering both the known population and the complete, but as yet largely undiscovered, population of NEAs. In Sec. IV, we describe in detail potential missions to near-Earth asteroids, and in Sec. V we present reasons why NEA missions should be in the SEI. In the final section, we summarize our results and put forth recommendations for the next steps to be taken in order to realize the exciting potential of near-Earth

asteroid missions.

II. ESTIMATING THE TRUE POPULATION OF NEAR-EARTH ASTEROIDS AND THEIR ORBITS

The known population of NEAs is only a sample of the much larger true population. In this section, we estimate the characteristics of the true population: how many are there? What orbits do they move in? What are they likely made of? We start with the number of NEAs as a function of their size. Shoemaker et al. (1990) have estimated the size distribution of the true population of Earth-crossing asteroids based on the number discovered through 1989 and the fraction of the volume of space that the population occupies that search programs cover. Figure 2 compares the number of asteroids known today with the estimated true population as a function of absolute magnitude, based on size distribution data for asteroids and lunar craters. (Note that this figure is for the Earth-crossing asteroids, a subset of the more general class of near-Earth asteroids that we study in this chapter.) The Earth-crossing asteroids include objects that will evolve into Earth crossers at some future time but currently have a perihelion distance between 1.02 and ~1.08 AU. About 60% of the known population of NEAs are Earth-crossing asteroids. Assuming this ratio to be characteristic of the total NEA population, we estimate there to be 1700±800 NEAs larger than ~1 km diameter. The number of NEAs increases dramatically at smaller sizes, a result of well understood power-law size relations among asteroid fragments. Just reducing the diameter limit to 0.5 km increases the number larger than this size to ~5700±2600 objects. This figure clearly shows that there are a large number of asteroids yet to be discovered: less than 10% of the estimated 1700 bodies >1 km diameter have been discovered. These conclusions are generally supported by statistics of rediscovery rates. The accidental rediscovery rate of NEAs is strongly dependent on the brightness of the asteroid and the fraction of asteroids that has been accidentally rediscovered closely matches the completeness shown in Fig. 2 (Shoemaker et al. 1990).

The orbits of the near-Earth asteroids determine the ΔV required to visit them, and knowing the distribution of orbital elements for the true population of NEAs is a prerequisite to finding mission ΔV's. Our estimate of the orbit element distributions for the true NEA population is given in Fig. 3, along with the distributions for the known population. To derive these distributions, we first plot the orbit elements of the known asteroids and represent them with simple mathematical expressions. A χ^2 test for the quality of our fits shows that the semimajor axis a, the eccentricity e, and the inclination i are adequately represented by a normal distribution; these elements are most important for mission ΔV calculation. The argument of periapse ω is nearly uniformly distributed, while the node longitude Ω can be approximated either by a uniform or a normal distribution, but neither representation is very good.

The observed NEA orbits, however, are a somewhat biased sample of the

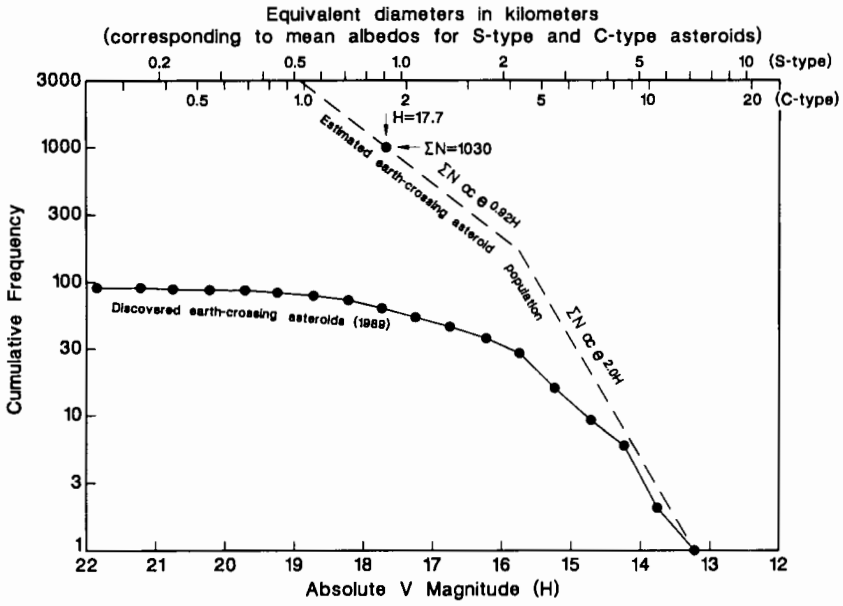


Figure 2. Magnitude-frequency distribution of Earth-crossing asteroids (figure from Shoemaker et al. 1990).

true population's orbits in the sense that those asteroids whose orbits bring them close to Earth or have low encounter speeds with the Earth are the most likely to be discovered. We apply the bias correction factors described by Shoemaker et al. (1990) to obtain our estimate of the unbiased orbit element distributions of the true NEA population (Davis et al. 1990) shown in Fig. 3.

The NEA population contains evidence of a variety of types of compositions. The S-class asteroid is the most common type found in the NEA population. However, as shown by Jewitt and Luu (1989), the true population probably contains about equal numbers of dark asteroids (C, D, F, and T class) and bright ones (S, E, M, V, and A class) down to ~1 km diameter objects. This is because the dark asteroids of a given size are much fainter (due to different photometric phase functions and because the mean albedo of the C-class is 0.04, while that of the S-class is 0.15) and therefore less likely to be discovered than comparably sized bright objects.

Finally, NEAs have a variety of rotation periods and shapes. Spin periods range from just over 2 hr to nearly 24 hr among the 28 asteroids measured to date. The variations in the brightness of an asteroid as it rotates, the lightcurve amplitude, is a measure of the degree of elongation of the asteroid; spherical bodies with a uniform albedo surface would have no amplitude to their lightcurves. Many NEAs show large lightcurve amplitudes, indicating elongated and perhaps very irregularly shaped bodies. Also, radar observations

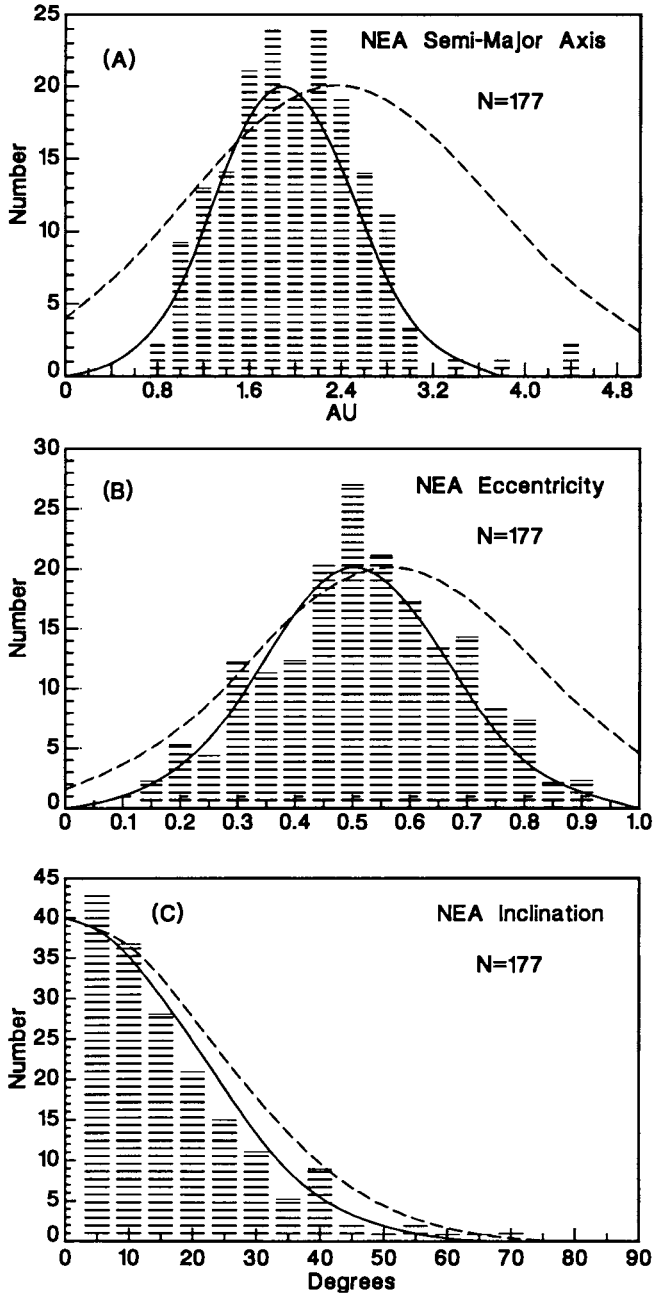


Figure 3. Histograms for the semimajor axis (A), eccentricity (B), and inclination (C) distributions for 177 near-Earth asteroids (1983SN is lost). The solid lines are a Gaussian approximation (or half-Gaussian for the inclination) while the dotted lines represent our estimate of the bias corrected orbit element distributions.

of NEAs indicate very irregular shapes for many of the bodies including a strongly bifurcated or "contact binary" dumbbell shape for 1989PB and possibly for others (Ostro et al. 1990). 433 Eros is also known to have a very elongated shape.

III. MINIMUM ΔV REQUIREMENTS FOR NEAR-EARTH ASTEROID MISSIONS

In this section we investigate the lower bound on trajectory energy requirements to rendezvous with NEAs for two types of missions as distinguished by trip time: (1) minimum-energy transfers generally characterized by lower ΔV and longer transit time, and (2) "fast" transfers having higher ΔV requirements and of interest primarily for shorter round trips appropriate for astronaut missions. We examine here both the known (discovered) population and the true (including the as yet undiscovered bodies) population of NEAs.

The parameter space of all possible transfers between Earth and a target asteroid is rather complex. It involves a number of variables and performance measures such as departure/arrival dates, ΔV impulses corresponding to these dates, and, in the case of round trips, stay time at the asteroid, total trip time, and velocity at Earth return. A quantitative description of this parameter space depends on the asteroid's orbital elements as well as the relative time phasing of the Earth and the asteroid in their respective orbits. Practical interest is generally focused in the region of optimal solutions (e.g., minimum total ΔV) because transfer energy requirements become excessive outside these regions—but there usually exist several locally optimal solution regions. In evaluating the mission performance potential of an entire population of NEAs, it would be very time consuming to calculate actual time-dependent trajectories throughout this parameter space. A simplified and efficient method described below has been used instead for this preliminary evaluation. Refined solutions based on real trajectory data will be presented in Sec. IV.

A. Analytical Figure-of-Merit

The lower bound on transfer energy requirements is obtained under the assumption of ideal longitudinal phasing between Earth and a target asteroid leading to Hohmann transfers. Earth's orbit is taken to be circular at 1 AU distance. The requisite plane change is split appropriately between the launch and arrival ΔV impulses assuming that the target orbit's apse line and nodal lines are collinear. Simple analytical expressions have been derived by Shoemaker in defining a performance figure-of-merit representing the total ΔV of the rendezvous transfer trajectory (Shoemaker and Helin 1978). Mission initiation begins in low Earth orbit (LEO); total ΔV is thus the sum of the launch or injection impulse to depart LEO and the impulse at asteroid rendezvous. Shoemaker's figure-of-merit depends only on the asteroid's orbital elements (a, e, i) and includes an adjustment factor to reflect the non-ideal phasing geometry of real mission opportunities.

We will not list the analytical expressions in this chapter for the sake of brevity, noting instead several of the key characteristics of the calculation method. In the case of minimum-energy transfers, rendezvous occurs at aphelion of the asteroid's orbit for Amors and Apollos and at perihelion for Aten-type asteroids; the transfer trajectory is tangent to both Earth's orbit and the asteroid's rendezvous apse point. Fast transfers are achieved by rendezvous at perihelion for Amors and Apollos ($a > 1.0$ AU) and at aphelion for Atens ($a < 1.0$ AU); the semimajor axis of the transfer trajectory is set to 1 AU in the case of fast transfers to either Apollos or Atens. Total ΔV requirements for rendezvous missions are, on average, about 50% higher for fast transfers compared to minimum-energy transfers. However, this penalty depends strongly on the orbital elements of specific targets, and typically varies from 25 to 65% for Amors and Apollos and 10 to 80% for Atens.

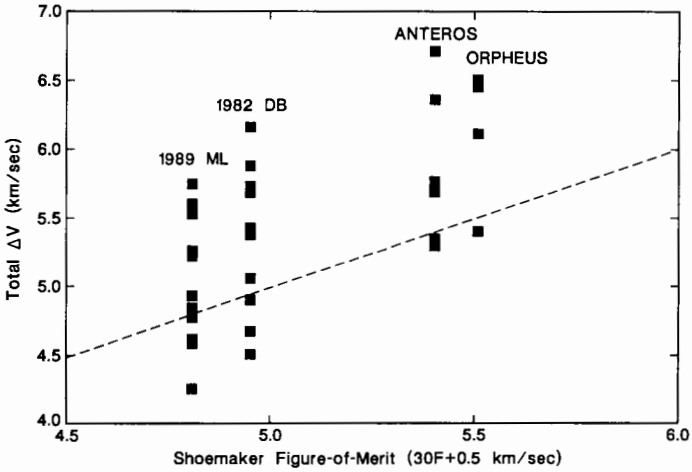


Figure 4. Launch opportunity variations in ΔV requirements for rendezvous with 2 Apollos and 2 Amors.

Figure 4 compares Shoemaker's figure-of-merit with the total ΔV requirement obtained from actual trajectory optimizations over several launch opportunities for two Apollo and two Amor targets. The agreement is within 11% on the low side and 24% on the high side for these cases. This test provided a degree of confidence that we could apply the analytical method in a preliminary global-type study of NEA mission requirements—at least to derive a consistent relative measure of performance characteristics.

Before applying the technique to specific targets in the NEA population, we first sought to obtain some general insight regarding performance trades in orbital element space. Figure 5(a) shows total ΔV contours for minimum-energy rendezvous in (a, e) space assuming a target orbit inclination of 5°, a somewhat typical value of "good" targets. As would be expected because

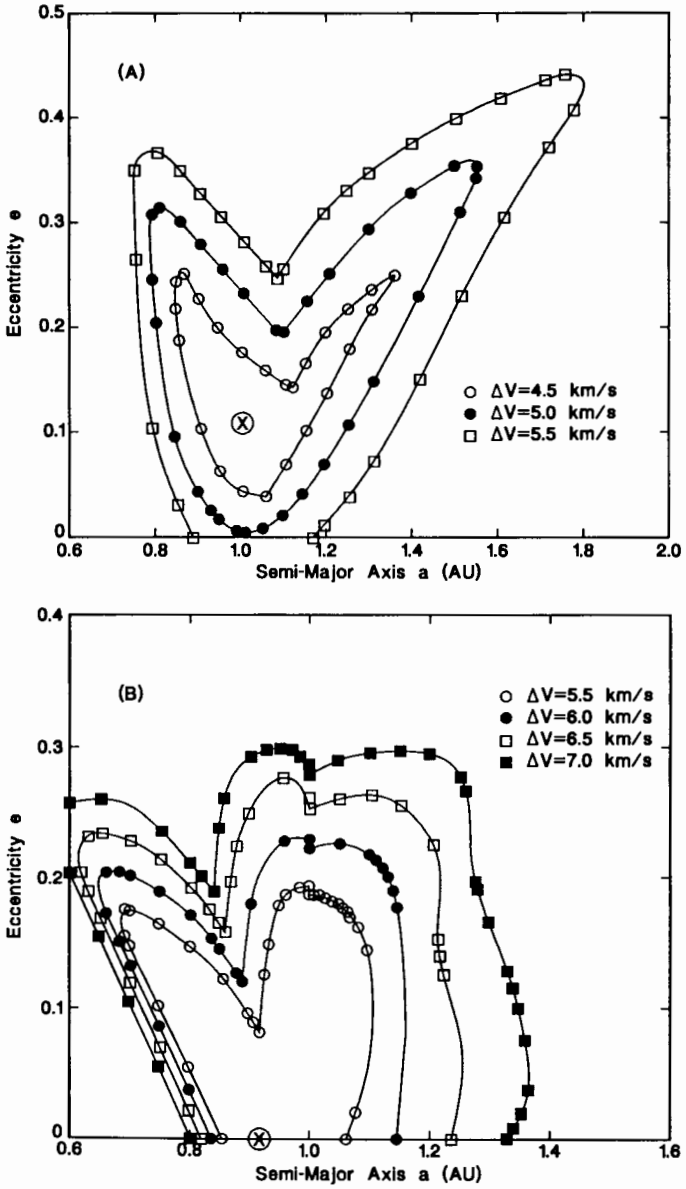


Figure 5. ΔV contours in a–e space for $i = 5$ deg. Part (A) is for minimum-energy rendezvous, while part (B) is for fast trip rendezvous.

orbital velocity is lower at aphelion distances, a constant ΔV contour at a given value of eccentricity is skewed to greater distance from Earth (1 AU) when the target's semimajor axis is >1 AU, i.e., for Apollos and Amors. For example, with a limit of 5 km s^{-1}, an asteroid's semimajor axis and

eccentricity can extend to 1.5 AU and 0.35, respectively, whereas for Atens the bounding a and e are about 0.79 AU and 0.31. The minimum value of the figure-of-merit is 3.89 km s^{-1} corresponding to $a = 1.0$ AU and $e = 0.11$ (indicated by the x in Fig. 5(a)).

Figure 5(b) shows several total ΔV contours for fast trip rendezvous to a target of orbit inclination 5°. Note that the lowest ΔV contour here is the same ΔV as for the highest ΔV contour in Fig. 5(a). Note also that the contours in this case tend to be skewed toward $a < 1.0$ AU. The apparent small discontinuity at $a = 1$ AU and higher eccentricity is an artifact of a switching function in the solution procedure. With a ΔV limit of 6.0 km s^{-1}, the bounding semimajor axis values are 1.16 AU at an eccentricity of about 0.10 (an Amor object), and 0.65 AU at an eccentricity of 0.2 (an Aten object). Note also that performance is highly sensitive to the orbital elements in the region of low values of a and e. The minimum value of the figure-of-merit is 4.15 km s^{-1} corresponding to $a = 0.91$ AU and $e = 0$ (indicated by the x in Fig. 5(b)).

B. Results for Known Population of NEAs

The analytical method described above was applied to the population of 178 NEAs discovered through mid-1991. Table I shows the performance accessibility of the 50 best targets for minimum-energy rendezvous missions. Results are listed in ascending order of total ΔV and also include the corresponding values of launch energy C3 and rendezvous impulse which compose the total ΔV. The two most accessible targets are 1989ML and 1982DB with total ΔV's under 5 km s^{-1}. There are 22 accessible targets having rendezvous mission ΔV less than 6 km s^{-1}, while 50 asteroids are reachable with ΔV less than 6.7 km s^{-1}. Figure 6(a) graphs the cumulative distribution of this data. As a point of comparison, the total ΔV needed to go from LEO to the lunar surface is approximately 6 km s^{-1}; thus, it is energetically easier to rendezvous with about 22 known NEAs than it is to place a payload on the lunar surface. It is exciting to see a more than 50% increase in the past year for the number of asteroids that are energetically more accessible than the Moon (22 in 1991 vs 14 in 1990). This dramatic increase in the number of excellent mission opportunities was predicted by Davis et al. (1990) and it is reassuring to see it occurring even more rapidly than anticipated.

Table II and Fig. 6(b) present the ordered listing and cumulative distribution graph for fast trip rendezvous with the known NEA population. The three most accessible targets in this case are 1991JW, 1989UQ and 1989ML, with total ΔVs between 5.9 and 6.5 km s^{-1}. The asteroid 1991JW, newly discovered in May, 1991 (Helin and Lawrence 1991), is a superb candidate for an astronaut mission: it requires less ΔV to reach than the Moon, yet a round-trip mission can be done in one year (see Sec. IV). We may estimate the total round-trip ΔV requirements by assuming a short stay time, a departure impulse comparable in magnitude with that required for rendezvous, and an aerocapture return to LEO. Then the total round-trip mission ΔV is 7.9 km s^{-1}

for 1991JW, 8.3 km s^{-1} for 1989UQ, 9.2 km s^{-1} for 1989FC and 9.5 km s^{-1} for both 1989ML and 3361 Orpheus. For comparison, a round-trip mission to the Moon that returns to LEO via aerocapture requires a total ΔV of 9.4 km s^{-1}. It is quite interesting to note that four of the seven best targets in the fast round-trip mission class were discovered only in the past two years. This again underscores one of the main points of this chapter: there are many good mission opportunities to targets awaiting discovery. To test the validity of this hypothesis, we have undertaken a Monte Carlo assessment of the distribution of mission ΔV performance for the true population of NEAs.

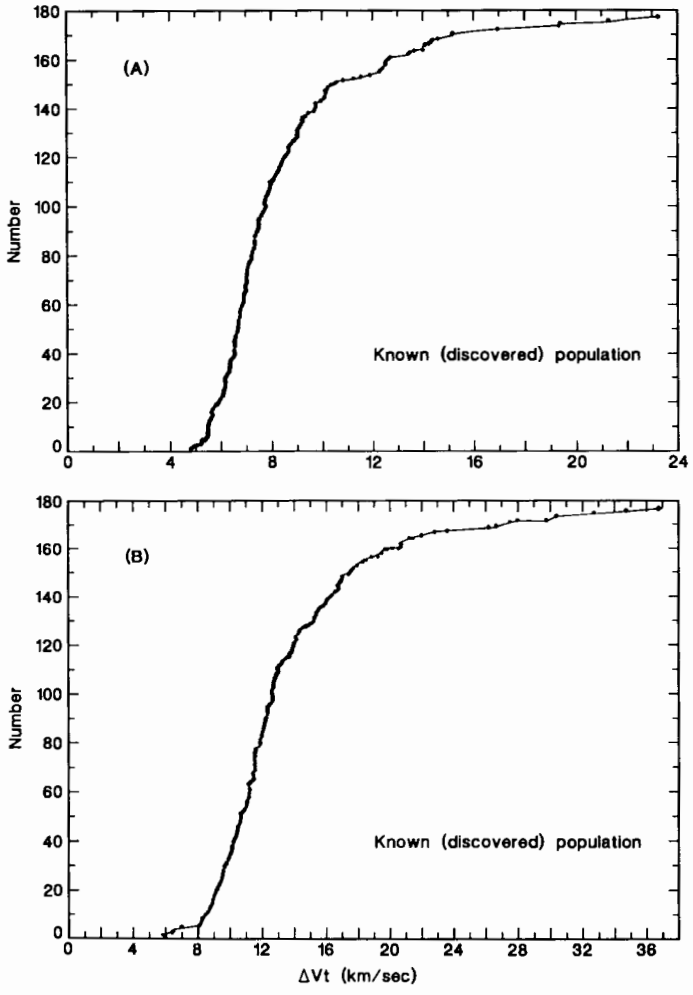


Figure 6. Cumulative distribution of total ΔV (from LEO) for: (A) minimum-energy rendezvous with near-Earth asteroids; (B) fast trip rendezvous.

TABLE I

Figure-of-Merit[a] for Minimum Energy Rendezvous Missions to Near-Earth Asteroids: The Fifty Lowest Total ΔV Asteroids

Rank	Asteroid	Total ΔV ΔV_t (km s^{-1})	Launch Energy C3 (km s^{-1})^2	Rendezvous ΔV ΔV_r (km s^{-1})
1	1989 ML	4.811	13.25	1.021
2	4660 1982 DB	4.953	28.07	0.536
3	1954 XA	5.247	35.33	0.534
4	1989 UQ	5.252	15.40	1.369
5	1943 ANTEROS	5.403	27.62	1.004
6	1991 JR	5.461	28.95	1.008
7	1989 VB	5.463	45.69	0.341
8	1990 MF	5.475	41.39	0.521
9	3361 ORPHEUS	5.507	16.93	1.557
10	3757 1982 XB	5.512	44.99	0.417
11	1990 VA	5.520	25.25	1.219
12	1991 BN	5.534	28.62	1.095
13	1991 JW	5.593	12.16	1.850
14	1989 UP	5.622	46.99	0.450
15	1990 BA	5.622	33.19	0.996
16	1990 OS	5.624	38.71	0.777
17	3908 1980 PA	5.671	47.59	0.476
18	1980 AA	5.745	46.92	0.575
19	3988 1986 LA	5.824	37.39	1.029
20	1977 VA	5.913	43.17	0.889
21	1991 FA	5.969	49.12	0.714
22	1991 DG	5.994	35.59	1.271
23	433 EROS	6.071	30.47	1.555
24	1991 DB	6.120	48.63	0.884
25	1988 TA	6.139	35.09	1.436
26	1990 UQ	6.145	35.94	1.409
27	1988 SM	6.161	43.07	1.142
28	1987 SF3	6.209	61.04	0.503
29	3352 MCAULIFFE	6.216	43.67	1.173
30	3288 SELEUCUS	6.236	53.50	0.814
31	2061 ANZA	6.250	61.73	0.518
32	1627 IVAR	6.276	48.39	1.049
33	1989 DA	6.289	61.47	0.567
34	1972 RB	6.344	57.51	0.770
35	1988 XB	6.368	32.63	1.765
36	1990 UN	6.384	43.80	1.335
37	1991 JX	6.407	70.12	0.367
38	4688 1980 WF	6.485	62.18	0.737
39	4015 1979 VA	6.562	73.87	0.386
40	3551 1983 RD	6.571	61.20	0.859
41	1990 UA	6.577	44.20	1.513

TABLE I (cont.)

Rank	Asteroid	Total ΔV ΔV_t (km s^{-1})	Launch Energy C3 (km s^{-1})^2	Rendezvous ΔV ΔV_r (km s^{-1})
42	1987 WC	6.578	37.86	1.765
43	2340 HATHOR	6.583	45.00	1.488
44	1989 AZ	6.590	49.34	1.326
45	1991 FE	6.612	56.41	1.079
46	1685 TORO	6.626	33.20	1.999
47	1221 AMOR	6.654	57.78	1.070
48	6743 P-L	6.658	46.67	1.498
49	4179 TOUTATIS	6.700	70.83	0.634
50	1620 GEOGRAPHOS	6.702	31.70	2.136

[a] Shoemaker and Helin 1978.

C. Results for True Population of NEAs

A statistical Monte Carlo evaluation of Shoemaker's figure-of-merit was conducted using the orbital element distributions for a, e and i as defined in Sec. II. In sampling these element distributions, we included the apparent correlation coefficient between a and e of $R = 0.5$, and placed bounds on perihelion distance ($q < 1.3$ AU) and aphelion distance ($Q > 0.7$ AU). A sample size of 5000 was selected for this experiment—about equivalent to the estimated mean population of NEAs with diameters >0.5 km. We assume that results obtained could be scaled in an approximate sense to a smaller population, as appropriate to quantify the frequency of mission opportunities for larger asteroids or specific taxonomic types.

Figure 7(a) shows the distribution of total ΔV for minimum-energy rendezvous with emphasis only on the lower ΔV tail of this distribution, i.e., <9 km s^{-1}. Table III(a) lists the 25 lowest ΔV cases from the sample size of 5000. Similar results are shown in Fig. 7(b) and Table III(b) for fast trip rendezvous trajectories. In the case of minimum-energy rendezvous, there are 47 asteroids (0.9%) that could be accessible for a $\Delta V < 5$ km s^{-1}, 257 asteroids (5.1%) for <6 km s^{-1}, and 709 asteroids (14.2%) for <7 km s^{-1}. An estimate of the number of accessible C-class asteroids of diameter 1.0 km or larger, assuming that one-third of the total true population of about 1700 objects are C asteroids with no orbit element bias, would then be 5, 29 and 80, respectively, for ΔV limits of 5, 6 and 7 km s^{-1}.

A comparison of Figs. 7(a) and 7(b) shows about an order of magnitude fewer targets for fast trip rendezvous at the same value of ΔV accessibility than for minimum energy rendezvous. We find that about 24 asteroids (0.5%) could be reached for a $\Delta V < 6$ km s^{-1}, 65 asteroids (1.3%) for 7 km s^{-1}, and 145 asteroids (2.9%) for up to 8 km s^{-1}. Monte Carlo calculations were also made for fast round-trip missions assuming approximately equal values for the asteroid arrival and departure impulses, and aerocapture to LEO at Earth return. The 5000 sample distribution results for total round-trip ΔV

TABLE II

Figure-of-Merit[a] for Fast-Trip Rendezvous Missions to Near-Earth Asteroids: The Fifty Lowest Total ΔV Asteroids

Rank	Asteroid	Total ΔV ΔV_t (km s^{-1})	Launch Energy C3 (km s^{-1})^2	Rendezvous ΔV ΔV_r (km s^{-1})
1	1991 JW	5.946	17.29	1.981
2	1989 UQ	6.383	28.78	1.937
3	1989 ML	6.530	7.25	3.003
4	3361 ORPHEUS	7.144	35.92	2.408
5	1990 VA	8.093	81.12	1.660
6	4660 1982 DB	8.175	7.68	4.629
7	1991 BN	8.303	22.02	4.138
8	4581 1989 FC	8.415	116.52	0.782
9	4544 1989 FB	8.515	62.81	2.743
10	1943 ANTEROS	8.584	11.06	4.889
11	2062 ATEN	8.586	48.55	3.353
12	1954 XA	8.789	8.47	5.208
13	2063 BACCHUS	8.790	94.08	1.906
14	1991 JR	8.805	12.75	5.036
15	1990 BA	8.868	6.78	5.362
16	1989 QF	8.998	104.08	1.775
17	433 EROS	9.050	14.72	5.196
18	1989 UR	9.118	98.57	2.081
19	1990 OS	9.130	14.10	5.303
20	1620 GEOGRAPHOS	9.187	44.85	4.098
21	1991 DG	9.194	21.55	5.048
22	1988 TA	9.313	41.59	4.351
23	1990 UQ	9.354	39.76	4.465
24	1990 MF	9.400	7.95	5.842
25	1988 XB	9.494	58.76	3.874
26	2340 HATHOR	9.543	53.73	4.112
27	3757 1982 XB	9.605	6.80	6.098
28	1685 TORO	9.621	59.47	3.973
29	3988 1986 LA	9.634	13.80	5.819
30	1977 VA	9.758	6.98	6.243
31	1990 SP	9.838	44.23	4.773
32	3352 MCAULIFFE	9.841	8.74	6.248
33	1989 VB	9.866	5.87	6.401
34	1988 TJ1	9.967	23.89	5.724
35	1989 UP	10.009	6.84	6.501
36	1990 UN	10.026	40.50	5.108
37	1980 AA	10.100	6.89	6.589
38	1987 WC	10.101	23.09	5.891
39	1988 SM	10.114	14.38	6.275
40	3908 1980 PA	10.115	5.95	6.646
41	1991 FA	10.232	6.65	6.731

TABLE II (cont.)

Rank	Asteroid	Total ΔV ΔV_t (km s^{-1})	Launch Energy C3 (km s^{-1})^2	Rendezvous ΔV ΔVr (km s^{-1})
42	3554 AMUN	10.325	97.72	3.317
43	1990 UA	10.356	54.28	4.904
44	1627 IVAR	10.369	11.34	6.662
45	6743 P-L	10.432	45.72	5.308
46	1989 AZ	10.433	29.02	5.978
47	1988 EG	10.546	130.64	2.460
48	2100 RA-SHALOM	10.575	57.12	5.016
49	2368 BELTROVATA	10.576	9.93	6.930
50	1991 DB	10.588	14.68	6.735

[a] Shoemaker and Helin 1978.

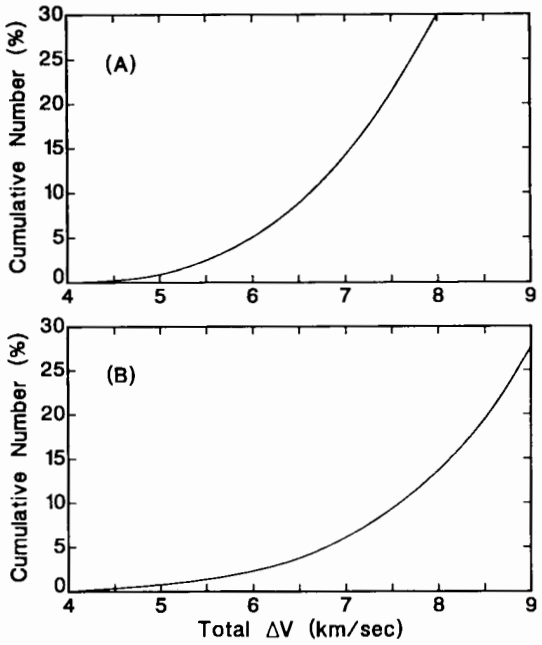


Figure 7. ΔV distribution of Monte Carlo data/Shoemaker figure-of-merit for: (A) minimum-energy rendezvous and (B) fast trip rendezvous.

are: 5 asteroids (0.1%) at <6 km s^{-1}, 35 asteroids (0.7%) at <8 km s^{-1}, and 85 asteroids (1.7%) at less than 10 km s^{-1}. For comparison purposes, we estimate that 67 (1.3%) potential NEA targets are accessible for a $\Delta V <$ the ~9.4 km s^{-1} needed for a lunar surface mission with aerocapture return. If all of these targets were discovered, there should be an estimated 9 to 10 NEAs available each year, assuming an average interval of 7 yr between the most

TABLE III(a)
25 Lowest ΔV Cases from Monte Carlo Sample of 5000[a]
Rendezvous Mission for Minimum ΔV

	a (AU)	e	i (deg)	ΔV (km s^{-1})
1	0.979	0.004	0.18	3.83
2	0.993	0.126	5.95	3.98
3	1.218	0.175	3.28	4.12
4	0.929	0.170	5.23	4.12
5	1.271	0.200	1.90	4.19
6	0.908	0.196	5.34	4.26
7	1.279	0.234	0.24	4.29
8	1.164	0.103	2.85	4.29
9	1.270	0.229	1.64	4.30
10	0.881	0.217	5.06	4.35
11	1.260	0.186	3.73	4.39
12	0.939	0.213	6.54	4.40
13	0.991	0.109	0.26	4.45
14	0.946	0.146	7.45	4.47
15	1.136	0.135	5.72	4.48
16	1.023	0.022	4.01	4.50
17	0.912	0.241	7.94	4.52
18	1.347	0.221	0.15	4.53
19	1.291	0.169	1.48	4.55
20	1.327	0.245	5.90	4.59
21	1.282	0.184	4.51	4.60
22	1.191	0.218	3.50	4.61
23	1.319	0.198	3.17	4.62
24	0.862	0.242	7.00	4.62
25	0.821	0.153	2.88	4.66

[a] (Shoemaker and Helin 1978).

favorable opportunities to a given target body.

IV. SAMPLE MISSIONS TO NEAR-EARTH ASTEROIDS

A considerable amount of NEA trajectory and mission analysis work has been accomplished during the past decade and a half. Most of this work has been performed at SAIC and at JPL with both groups often collaborating on related studies. Rendezvous and sample return missions in robotic scientific exploration have been investigated and, more recently, manned round-trip applications have been examined as well. Among the key publications comprising the data base of trajectory and mission performance analysis are the following: Niehoff (1977); Friedlander et al. (1974); Stancati and Soldner (1981); Lau and Hulkower (1983); NEAR Science Working Group (1986);

TABLE III(b)
25 Lowest ΔV Cases from Monte Carlo Sample of 5000 for Fast-Trip Rendezvous Missions

	a (AU)	e	i (deg)	ΔV (km s^{-1})
1	0.979	0.004	0.18	3.88
2	0.991	0.109	0.26	4.24
3	0.885	0.019	4.50	4.39
4	0.807	0.095	5.37	4.66
5	1.015	0.159	2.20	4.83
6	1.023	0.022	4.01	4.85
7	0.993	0.126	5.95	4.97
8	0.932	0.021	7.02	5.10
9	1.003	0.081	7.24	5.26
10	0.741	0.140	6.76	5.35
11	0.784	0.161	2.10	5.36
12	1.034	0.202	2.34	5.38
13	0.962	0.047	6.19	5.45
14	1.021	0.181	5.02	5.46
15	1.032	0.118	7.18	5.51
16	0.821	0.153	2.88	5.59
17	0.929	0.170	5.23	5.64
18	1.046	0.193	5.27	5.68
19	1.164	0.103	2.85	5.69
20	1.043	0.215	3.94	5.73
21	0.946	0.146	7.45	5.74
22	0.957	0.022	8.24	5.77
23	0.650	0.187	4.82	5.91
24	1.146	0.221	2.59	5.93
25	0.908	0.196	5.34	6.04

TABLE IV
Ballistic Rendezvous with Near-Earth Asteroids[a]

	Minimum	**Mean**	**Maximum**
Flight time (days)	153	686	1444
Launch C3 (km s^{-1})^2	8.3	35.8	82.0
Spacecraft ΔV (km s^{-1})	0.29	1.49	4.59
Total ΔV[a] (km s^{-1})	4.25 (4.81)[b]	6.13 (6.47)[b]	9.95 (8.60)[b]

[a] Includes injection ΔV from low Earth orbit.
[b] Shoemaker's analytic figure-of-merit ($30F$ + 0.5) (Shoemaker and Helin 1978).

Yen (1989); McAdams (1992); Farquhar et al. (1993); and NASA (1989*a*). In this section, we will summarize the trajectory database information that is available with emphasis on post-1998 mission launch opportunities.

TABLE V

Rendezvous Opportunities to Near-Earth Asteroids (Launch Years 1998–2010)

Asteroid	Earth Launch Date	Launch Energy, C3 $(km\ s^{-1})^2$	Midcourse Maneuver Date	Asteroid Arrival Date	Post-Launch ΔV $(km\ s^{-1})$	Total ΔV $(km\ s^{-1})$	Flight Time (days)
1982DB	01/07/98	30.88	——	01/14/00	1.155	5.671	737
	01/10/00	25.93	——	10/17/01	0.701	5.012	646
	01/22/02	22.64	08/13/02	10/24/03	0.284	4.457	640
	01/26/04	17.50	08/19/04	08/21/05	0.894	4.848	573
	01/29/06	12.04	08/21/06	06/20/07	1.647	5.364	507
	01/26/07	32.28	——	06/17/09	1.286	5.860	873
	02/03/08	6.68	08/08/08	04/18/09	2.605	6.084	440
	01/08/09	28.36	——	12/02/10	0.928	5.340	693
Anteros	05/28/99	24.72	04/28/00	09/03/00	1.842	6.103	463
(1943)	05/24/00	43.24	07/20/01	07/09/02	1.861	6.829	776
	05/26/00	45.31	04/30/01	07/13/02	1.767	6.862	779
	05/26/02	38.49	06/02/03	04/02/04	0.916	5.741	677
	05/26/04	30.52	06/08/05	10/11/05	1.155	5.656	504
	05/29/06	17.84	03/26/07	07/20/07	2.875	6.844	416
	05/25/07	57.27	04/19/08	03/27/09	1.298	6.853	673
	05/26/07	40.40	06/22/08	05/15/09	1.360	6.261	721
	05/24/09	35.71	——	07/25/10	0.595	5.308	427
Orpheus	01/25/98	30.40	——	06/09/98	2.125	6.622	134
(3361)	10/14/00	21.11	05/15/01	02/08/02	2.371	6.479	482
	01/15/02	37.20	——	05/22/02	2.183	6.956	127
	01/04/03	16.80	——	06/06/04	2.452	6.376	519
	10/14/04	21.12	01/15/05	02/06/06	2.371	6.480	480

Eros (433)	01/21/98	42.09	——	12/25/98	1.115	6.083	338
	01/06/06	42.81	——	05/07/06	2.241	7.237	120
	01/04/07	16.61	——	06/02/08	2.428	6.344	516
	10/14/08	21.13	01/15/09	01/30/10	2.366	6.475	473
	12/10/09	40.56	——	11/27/10	1.269	6.176	351
	01/21/00	29.66	10/29/00	07/14/01	2.319	6.785	540
	01/25/01	45.20	01/22/02	04/04/03	2.547	7.637	799
	01/26/03	41.53	12/14/03	01/08/05	1.913	6.859	713
	01/25/05	38.43	——	11/13/05	1.112	5.934	292
	01/21/07	31.17	11/03/07	07/29/08	2.096	6.624	555
	01/26/08	45.89	01/29/09	04/20/10	2.648	7.766	815
	01/26/10	42.17	12/21/10	01/24/12	2.040	7.011	729
Alinda	01/05/98	66.59	04/11/00	08/05/01	1.040	6.943	1308
(887)	01/08/99	52.12	08/15/00	08/06/01	2.316	7.675	942
	01/05/02	65.87	03/27/04	07/13/05	1.101	6.978	1285
	01/08/03	50.90	07/31/04	07/15/05	2.426	7.738	919
	01/05/06	65.14	03/11/08	06/20/09	1.165	7.015	1262
	01/08/07	49.63	07/16/08	06/22/09	2.542	7.805	896
	01/06/10	64.38	02/25/12	05/28/13	1.230	7.058	1239
1989 ML	07/30/98	8.66	——	08/31/00	1.986	5.553	763
	07/06/99	11.49	——	07/31/01	0.514	4.206	757
	07/12/99	12.49	——	06/20/00	0.820	4.556	344
	07/16/02	10.61	02/07/04	10/09/04	1.187	4.841	816
	08/01/02	8.29	——	07/03/03	2.140	5.691	334
	07/08/03	15.59	02/07/04	01/24/05	1.298	5.170	566
	07/05/06	14.54	——	12/05/06	1.650	5.476	153
	07/04/06	12.19	——	10/10/08	1.173	4.896	829
	07/05/06	11.48	——	03/26/07	0.513	4.205	264
	07/11/09	12.56	——	10/07/11	0.790	4.529	818
	07/24/09	9.85	——	07/08/10	1.577	5.197	349

A. Rendezvous Mission Trajectory Performance

Table IV shows a statistical summary of rendezvous opportunities to a sampleof Apollo/Amor asteroids. The sample statistics are based on 81 launch opportunities to 29 NEA targets over the 20-yr period 1989–2009; these targets include Apollos and Amors having orbital inclinations $<13°$. In general, the data are representative of locally optimal total ΔV solutions of the minimum-energy type. Thus, flight time tends to be fairly long with a mean value of 686 days and a very wide variation from 153 to 1444 days. We also note that statistical data for Shoemaker's figure-of-merit for these targets is in reasonably close agreement to actual trajectory results.

Table V lists detailed data for rendezvous trajectories to six NEA targets over a range of launch dates from 1998 to 2010 with Earth departure from LEO. Altogether there are 52 different mission opportunities represented. In a few cases, we give two locally optimal solutions during the same "launch period" when such data is available. The information listed includes Earth launch date, launch energy C3, midcourse maneuver date (if any, to improve performance), asteroid arrival date, post-launch ΔV, total ΔV, and flight time.

As noted earlier, there is a significant variation in mission performance over the launch opportunities to any given asteroid, reflecting the changing orbital position geometry between Earth and the target body. Anteros illustrates this point. Over the nine missions listed, launch energy C3 varies from 17.8 to 57.3 $(\text{km s}^{-1})^2$, post-launchΔV from 0.59 to 2.87 km s^{-1}, total ΔV from 5.33 to 6.88 km s^{-1}, and flight time from 416 to 779 days. Note that optimal launch dates for Anteros (as well as for many other targets in elliptical, inclined orbits) tend to occur in a narrow region of the calendar year. This is due to the preference of a nearly fixed longitude of Earth at launch as lying between the asteroid orbit's apsidal and nodal lines. The wide variation in trajectory performance parameters simply reflects this general condition of optimality and the fact that the asteroid's position in its orbit does not usually repeat between successive launch opportunities. There does exist a close repeat after a number of years as determined by the period of syzygy (lowest order close resonance) with Earth; for Anteros this is about 12 yr.

B. Round-Trip Mission Trajectory Performance

Table VI lists detailed trajectory data for round-trip missions to 6 NEA targets over a range of launch dates from 1998 to 2010 with Earth departure from LEO. Altogether there are 54 different mission opportunities represented. These are locally optimal total ΔV solutions assuming aerocapture at Earth return, i.e., Earth orbit capture ΔV is not included in total ΔV performance. Most of the data that were generated represent low-energy, long flight time solutions (3 to 4 yr trips) applicable mainly to robotic (sample return) missions rather than manned missions.

Three target asteroids were specifically examined for shorter trips of <1.5 yr duration; namely, 1989UQ, 1989ML and Orpheus—the best perfor-

mance targets indicated by Shoemaker's figure-of-merit. As 1989UQ is in a close 8/7 orbit period resonance with Earth, the lowest ΔV launch opportunities recur every 7 years. The launch years 2009 and 2010, for example, are two fairly good contiguous opportunities; while 2009 launches offer lower values of total ΔV for trip times around 500 days, the 2010 opportunity provides lower return speeds for Earth aerocapture with correspondingly lower aerodynamic g-forces and heating rates. Performance results for 1989ML and Orpheus are clearly not as good as 1989UQ—about 1 to 2 km s^{-1} higher total ΔV for equivalent trip times—as predicted by the analytical method.

Constrained round-trip times <500 days are also shown in Table VI for these targets, but total ΔV begins to increase significantly. Taking 1989UQ as an example, a 365-day trip launched in 2009 requires a total ΔV of 9.5 km s^{-1} while a 215-day trip launched in 2010 has a total ΔV of 10.5 km s^{-1}. A 365-day trip to 1989ML launched in 2005 requires a total ΔV of 9.9 km s^{-1}. The Davis et al. (1990) study did not find any other known NEA targets that have short (6–12 months) round trips with total ΔV requirements under 12 km s^{-1}.

To show what can be expected for low ΔV short round trips, an asteroid target was "made up" in Davis et al. (1990) which was named 1994OK (ideal), indicating that a target such as this might be discovered in the near future. For this purpose, they used the semimajor axis and inclination of the real asteroid 1989FC (4581 Asclepius) (a = 1.02264 AU, i = 4.912 deg) but changed its eccentricity, ascending node, argument of perihelion, and date of perihelion to the following values: 0.05, 307 deg, 0 deg, and July 31, 2005. Table VI lists the performance data for a 6-month round trip with a total ΔV requirement of only 5.5 km s^{-1}. The trajectory profile for this nearly ideal mission is illustrated in Fig. 8(a). At no time during this mission is the spacecraft very far away from Earth; at approach to rendezvous the distance to Earth is 0.13 AU and the solar phase angle is 82 deg.

The asteroid 1991JW has many of the desired characteristics of 1994OK: low eccentricity (e = 0.11827), semimajor axis near 1 AU (a = 1.0378), and modest inclination (i = 8.70539). Twelve-month and six-month round-trip missions are illustrated in Figs. 8(b) and 8(c) for comparison with the mission to 1994OK. For 1991JW, the six-month mission is quite feasible, requiring a total ΔV of 10.9 km s^{-1}. A twelve-month round trip-mission to this asteroid needs a total ΔV of only 5.9 km s^{-1}—less than the ΔV needed for minimum energy rendezvous for most NEAs. Future discoveries will undoubtedly bring us nearer to the ideal NEA for astronaut missions.

Round-trip mission requirements to 1991JW show a significant performance variation over the cycle of yearly launch opportunities. This is due to the asteroid's semimajor axis being close to that of Earth's orbit with the resulting slow drift in relative longitudinal positions. The period of syzygy (lowest close order resonance) is 18 yr; the year 2009 is an optimal launch opportunity in this repeat cycle. Table VI shows the ΔV variation for 1-yr class round trips with launches in 2008, 2009 and 2010. Data over the complete

TABLE VI

Round-Trip Opportunities to Near-Earth Asteroids (Launch Years 1998–2010)

Asteroid	Earth Launch Date	Launch Energy C3 $(km\ s^{-1})^2$	Midcourse Maneuver Date	Asteroid Arrival Date	Post-Launch ΔV $(km\ s^{-1})$	Stay Time (days)	Asteroid Departure Date	Departure ΔV $(km\ s^{-1})$	Earth Return Date	Earth Return $V\infty$ $(km\ s^{-1})$	Total ΔV $(km\ s^{-1})$	Flight Time (days)
1982DB	01/07/98	30.88	—	01/14/00	1.155	222	08/23/00	0.083	02/08/02	–NA–	5.754	1492
	01/02/00	25.62	—	09/05/01	0.869	30	10/05/01	0.582	12/21/01	5.544	5.749	720
	01/10/00	25.93	—	10/17/01	0.701	70	12/26/01	0.565	02/07/04	–NA–	5.577	1489
	12/24/00	35.97	—	11/03/01	6.573	30	12/03/01	1.862	12/24/01	5.440	13.158	365
	12/21/01	32.95	—	01/30/02	0.955	30	03/01/02	0.668	02/11/04	6.354	6.224	782
	01/22/02	22.64	08/13/02	10/24/03	0.284	70	01/02/04	1.127	01/09/06	–NA–	5.584	1448
	01/26/04	17.50	08/19/04	08/21/05	0.894	53	10/13/05	1.495	01/06/08	–NA–	6.343	1440
	02/12/04	28.37	—	09/05/05	0.761	30	10/05/05	2.317	10/25/06	5.979	7.491	986
	01/29/06	12.04	08/21/06	06/20/07	1.647	43	08/02/07	1.817	01/04/10	–NA–	7.181	1436
	01/06/07	33.05	—	02/18/09	1.353	180	08/17/09	0.339	02/16/11	–NA–	6.297	1502
	02/03/08	5.96	08/15/08	04/11/09	2.663	60	06/10/09	2.139	01/31/12	–NA–	8.248	1458
	01/03/09	27.98	—	04/16/10	3.456	30	05/16/10	2.033	02/05/11	7.000	9.886	763
	01/08/09	28.36	—	12/02/10	0.928	45	01/16/11	0.336	02/07/13	–NA–	5.676	1490
	06/21/10	15.98	—	12/27/10	5.498	30	01/26/11	5.146	05/28/11	6.283	14.532	342
Anteros	05/28/99	24.72	04/28/00	09/03/00	1.842	233	04/24/01	0.492	05/27/02	–NA–	6.595	1094
(1943)	04/03/00	27.16	—	10/11/01	5.251	30	11/10/01	4.857	04/04/02	5.706	14.471	731
	12/04/01	32.50	—	05/08/02	4.591	14	05/22/02	4.871	12/04/02	5.389	14.045	365
	05/17/02	32.49	09/30/02	06/17/03	2.228	330	05/12/04	2.875	05/30/05	–NA–	9.685	1110

	05/27/02	58.43	—	04/07/03	2.313	30	05/07/03	4.900	05/27/04	6.880	12.812	731
	05/20/04	42.54	—	02/04/06	1.280	30	03/06/06	1.256	05/18/07	6.463	7.522	1093
	05/26/04	30.52	06/08/05	10/11/05	1.155	142	03/02/06	1.076	05/26/07	–NA–	6.732	1096
	05/30/06	17.91	03/14/07	07/14/07	2.873	30	08/13/07	0.406	05/27/09	–NA–	7.251	1093
	05/08/07	28.39	—	08/09/08	3.969	30	09/08/08	4.316	05/08/09	5.873	12.699	731
	05/24/09	35.71	—	07/25/10	0.595	237	03/19/11	1.765	05/25/12	–NA–	7.073	1096
	05/31/09	38.68	—	01/05/10	2.645	30	02/04/10	5.974	06/01/11	6.880	13.451	731
Orpheus	01/25/98	30.40	—	06/09/98	2.125	352	05/27/99	1.444	04/04/01	–NA–	8.066	1164
(3361)	01/13/02	36.60	—	05/11/02	2.250	30	06/10/02	2.827	04.28/03	–NA–	9.826	470
	01/04/03	16.80	—	06/06/04	2.452	135	10/19/04	1.276	11/12/05	–NA–	7.652	1044
	05/20/05	17.64	—	02/03/06	2.898	10	02/13/06	2.733	09/30/06	3.927	9.591	498
	05/20/05	17.66	—	02/02/06	2.925	10	02/12/06	3.205	09/12/06	4.279	10.091	480
	06/02/05	15.51	—	02/07/06	3.144	30	03/09/06	4.085	07/27/06	4.332	11.097	420
	07/08/05	20.06	12/11/05	02/22/06	3.971	10	03/04/06	3.994	08/02/06	4.151	12.029	390
	07/19/05	16.54	—	12/09/05	5.511	10	12/19/05	3.588	03/16/06	5.819	13.012	240
	01/04/07	16.61	—	06/02/08	2.428	275	03/04/09	0.937	11/01/09	–NA–	7.281	1033
1989ML	12/23/05	9.65	—	07/11/06	3.270	10	07/21/06	2.981	01/05/07	2.399	9.862	378
	12/30/05	1.46	—	07/12/06	3.182	10	07/22/06	3.041	12/30/06	2.078	9.914	365
	01/01/06	12.24	—	07/14/06	3.088	10	07/24/06	3.519	11/27/06	2.039	10.332	330
	12/27/05	10.60	—	07/12/06	3.180	10	07/22/06	4.225	10/23/06	3.438	11.058	300
	01/06/06	15.07	—	07/11/06	3.247	10	07/21/06	5.108	10/03/06	4.818	12.204	270
	07/06/06	11.48	08/22/08	03/23/07	1.177	327	02/13/08	1.448	07/06/09	4.833	6.317	1097

TABLE VI (cont.)

Round-Trip Opportunities to Near-Earth Asteroids (Launch Years 1998–2010)

Asteroid	Earth Launch Date	Launch Energy C3 $(km\ s^{-1})^2$	Midcourse Maneuver Date	Asteroid Arrival Date	Post-Launch ΔV $(km\ s^{-1})$	Stay Time (days)	Asteroid Departure Date	Departure ΔV $(km\ s^{-1})$	Earth Return Date	Earth Return $V\infty$ $(km\ s^{-1})$	Total ΔV $(km\ s^{-1})$	Flight Time (days)
1989UQ	03/04/09	2.26	—	12/11/09	3.207	10	12/21/09	0.953	07/30/10	6.362	7.439	513
	04/02/09	6.21	—	12/16/09	3.119	10	12/26/09	0.989	07/26/10	6.458	7.566	480
	04/30/09	10.01	08/29/09	12/30/09	3.186	10	01/09/10	1.049	07/24/10	6.556	7.862	450
	05/28/09	24.09	08/29/09	01/01/10	2.809	10	01/11/10	1.081	07/22/10	6.671	8.124	420
	06/22/09	48.84	—	01/13/10	2.142	10	01/23/10	1.217	07/17/10	6.969	8.591	390
	07/10/09	69.13	—	01/27/10	1.865	10	02/06/10	1.612	07/10/10	7.636	9.473	365
	02/10/10	12.38	—	-9/21/10	2.458	32	10/23/10	2.571	06/05/11	3.252	8.761	480
	02/26/10	13.11	—	94/14/10	2.655	47	11/01/10	2.794	05/22/11	3.362	9.213	450
	02/12/10	12.28	—	09/22/10	2.482	67	11/28/10	3.425	04/08/11	4.313	9.643	420
	03/09/10	13.84	—	09/13/10	2.845	80	12/02/10	3.605	04/03/11	4.522	10.245	390
	03/23/10	16.69	—	09/04/10	3.081	96	12/09/10	4.074	03/23/11	5.106	11.074	365
	03/06/10	15.73	—	08/02/10	3.765	10	08/12/10	2.698	11/06/10	6.088	10.341	345
	03/25/10	20.76	—	08/10/10	3.692	10	08/11/10	2.693	10/26/10	7.232	10.478	215
1994OK (IDEAL)	01/29/05	9.41	—	05/19/05	0.968	10	05/29/05	0.953	08/05/05	3.561	5.522	188
1991JW	05/12/08	43.05	—	12/18/08	1.982	36	01/23/09	1.186	05/25/09	6.969	6.988	378
	05/16/09	21.16	—	08/23/09	1.790	40	10/02/09	0.628	05/15/10	5.208	5.900	363
	11/16/10	30.10	—	03/01/11	1.959	10	03/11/11	1.524	11/30/11	6.439	6.443	379

[a] Trajectory statistics based on 81 launch opportunities (1989–2009) to 29 targets (9 Apollos, 20 Amors), including <13 deg (6.7 deg average) with locally optimal total ΔV solutions.

[b] Includes injection ΔV from low Earth orbit.

[c] Shoemaker's analytic figure-of-merit ($30F + 0.5$) (Shoemaker and Helin 1978).

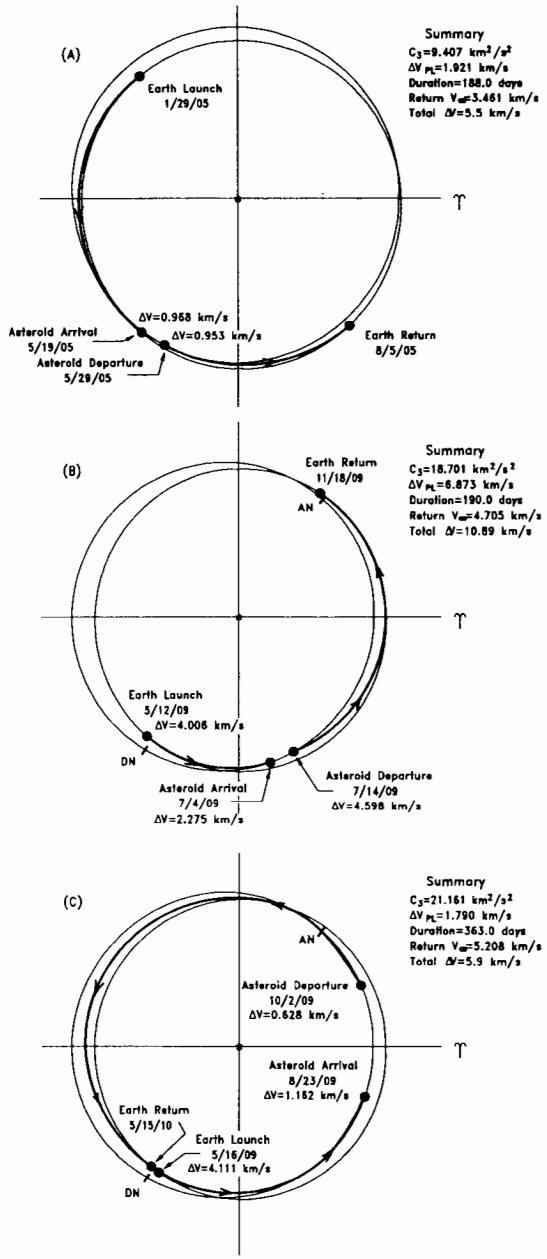


Figure 8. Mission profiles for: (A) 2005 launch for 6-month round trip to 1994 OK (ideal); (B) 2009 launch for 6-month round trip to 1991 JW; and (C) 2009 launch for a one-year round trip to 1991JW.

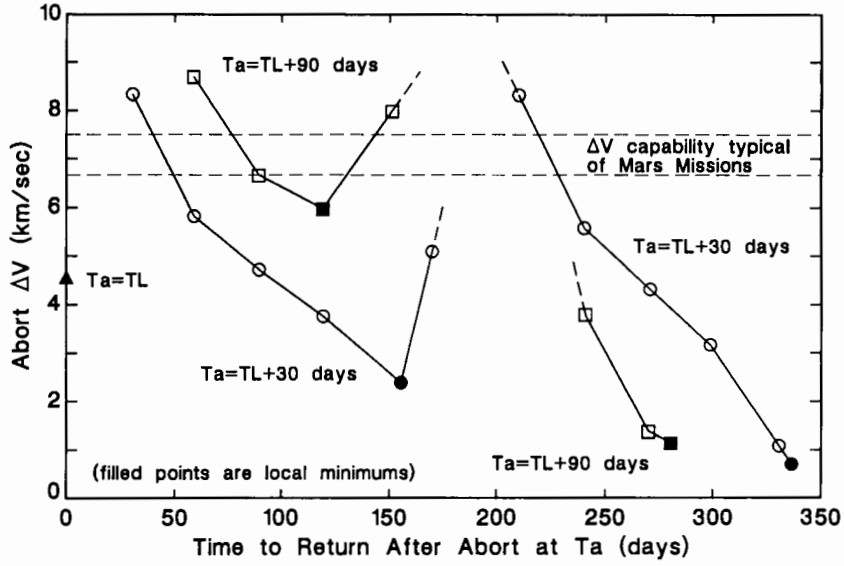


Figure 9. Abort requirements for one-year round trip to 1991JW.

cycle have not yet been generated.

The issue of crew safety is an important one for manned missions to NEAs as it is for lunar and Mars missions. It is desirable to be able to abort the mission if there is a system failure or a medical emergency and return to Earth as quickly as possible. This would be possible because the NEA mission, being a dress rehearsal for a Mars mission, has excess propellant capability by virtue of its system design for Mars application. Figure 10 (Sec. V) shows the abort ΔV requirements at several abort times on the outbound leg to 1991JW for the 1-yr round trip mission launched in 2009. The ΔV cost for an immediate return after launch is 4.6 km s^{-1}. If the abort occurs 30 days after launch, the local minimum ΔV requirement is 2.4 km s^{-1} for return 155 days later, but faster returns are possible with more propellant expenditure. For an abort 90 days after launch (just before rendezvous), the local minimum ΔV requirement is 6.0 km s^{-1} for return 118 days later and 6.9 km will return the spacecraft to Earth in 90 days. If fuel becomes a problem, then at 90 days after launch, just 1.1 km s^{-1} of ΔV will return the spacecraft 279 days later, compared with 2.4 km s^{-1} needed to return to Earth if the nominal mission profile is followed. The calculations illustrate the degree of flexibility available to NEA missions if extra fuel is carried. The dashed lines in Fig. 9 show the ΔV range needed for a round trip Mars orbit mission, and if this same ΔV capability were to be carried on an NEA mission, it would provide significant capability for quick return to Earth.

V. THE ROLE OF NEAR-EARTH ASTEROID MISSIONS IN THE SPACE EXPLORATION INITIATIVE

A. Mission Rationale

What are the reasons for considering missions to near-Earth asteroids as part of the SEI? What are the prerequisites for implementing NEA strategies now being defined, and what are the long-term benefits of such missions? We have shown in Secs. III and IV that good mission opportunities to NEAs with round-trip flight times around one year occur, on average, several times annually. Provided search programs are mounted soon to begin discovering these bodies, we can expect to incorporate these opportunities into SEI planning. Given that mission opportunities are frequent, what benefits do they offer to the SEI? The answer has three aspects: (a) exploration; (b) capability demonstration; and (c) programmatic momentum.

Planetary exploration is the essence of SEI, and the NEAs are the very building blocks of planets—they demand to be explored from both scientific and resource perspectives. Moreover, the near-Earth asteroids are the most accessible unexplored bodies in the solar system, and many of them can be visited much more easily than the Moon, the terrestrial planets, or comets. Given the large investment required to send humans to explore Mars, it makes eminent sense to increase the overall return by exploring representatives of the NEA population. The incremental cost of piloted NEA exploration would be small when compared both with the investment made to land on Mars and the scientific benefits to be realized.

Beyond the compelling exploration opportunities offered by an NEA mission, it would also serve as an excellent test bed for the Mars flights to follow. The NEA mission would demonstrate the *capabilities* of all hardware to be used for Mars orbital flights under actual deep space conditions, exercising the entire Mars spacecraft "stack," the crew, and ground support organizations away from the comfortable familiarity of Earth-Moon space. The entire Mars mission vehicle (except the Mars descent stage) would be used for the NEA rendezvous mission (see Fig. 10). The flight would depart from LEO using the trans-Mars injection stage, just as a real Mars mission would, and use its propulsive Mars capture capability to rendezvous with the asteroid. Should nuclear thermal rockets be used for propulsion, these mission phases would demonstrate for the first time the deep space operation of that system without demanding the performance necessary for a full-up Mars mission.

Initially, the spacecraft would orbit the asteroid at a high altitude for purposes of global reconnaissance, gradually lowering the orbit as more details on the surface and gravity field are gathered. With landing sites selected, the spacecraft would move sufficiently close for astronauts to employ an EVA transporter and space suits for surface exploration. Upon completion of this surface exploration phase, the crew would depart the asteroid using the same propulsion system designed for escape from Mars orbit. Recovery at Earth will use the same propulsive, aerobraking, or direct-entry technique to be used

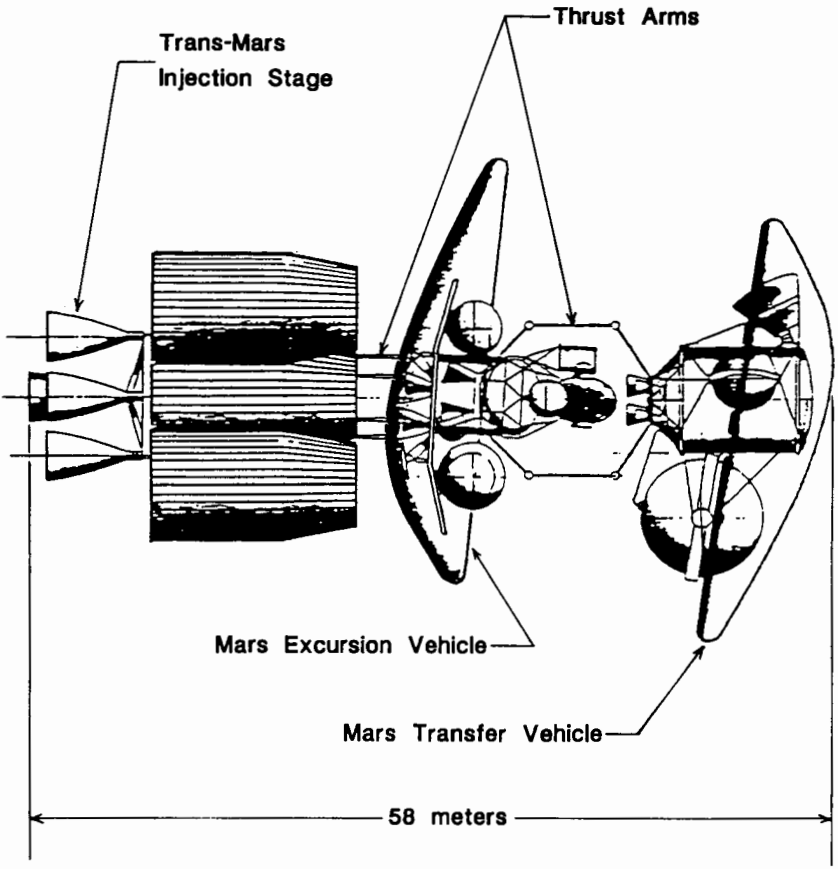


Figure 10. Mars Transportation System (from the 90-day study report).

on return from an actual Mars mission.

With a profile that includes all major mission phases except Mars powered descent/ascent, the NEA mission can serve as the essential shakedown cruise for the Mars transportation system, but at less than half the duration of an actual landing mission. Overall risk is also considerably less, because, should an abort be necessary, it could be accomplished more quickly and using less propellant than from a Mars trajectory as discussed in Sec. IV. As an added safety factor, carrying the full propellant load for a Mars mission might yield an appreciably faster return capability. Operational demonstration of the Mars mission elements via an NEA exploration flight would provide the same confidence to the SEI that the Apollo 8 and 10 missions did to the lunar landing program in 1968–69.

Aside from invaluable flight experience, an NEA mission would provide a major boost to the *momentum* of SEI. Assuming a return to the Moon begins around 2005, and that expeditions to Mars are planned for circa 2015, there

will be a gap of about 10 years separating these two achievements. An NEA mission in the 2010 time frame would prove to be an exciting and highly visible intermediate milestone, demonstrating major progress in accomplishing the SEI's goals. Exploration of these small islands in space, with attendant images of astronauts hovering just above the asteroid surface, the spacecraft and bright blue dot of Earth suspended in the background starfield, would provide dramatic visual proof of the vitality and potential of the SEI, whetting the appetite for the Mars exploration to follow.

B. Human Exploration of NEAs

We have noted the attractive programmatic benefits that human missions to NEA will provide to the SEI's Moon-Mars program. Beyond a flight hardware test, operational experience, and a highly visible program milestone, a piloted asteroid mission would provide a scientific return whose importance is difficult to overestimate. Essentially, a human mission to an NEA using Moon/Mars-derived hardware would provide access to a third "planetary" surface for the price of the initial two. That third surface, however, will differ starkly from a lunar or Martian environment, and so the exploration approach will differ as well. In this section, we discuss some operational factors important for NEA missions, and outline some concepts for human exploration of their surfaces.

An asteroid surface presents difficult challenges to both robotic and human explorers. The complexity of the exploration task is high, considering the spectrum of compositions, geologic histories, and surface morphologies encompassed by the NEA population. Explorers will have to contend with micro-g surface gravities, lengthy round-trip communication times, and complex orbital navigation tasks. Lunar and Mars rovers and their associated traverse strategies are unlikely to be of much use in near-weightlessness on an asteroid surface. Combined with the demanding environment, the complexity of scientific and sampling tasks on a heterogeneous asteroid may make development of a suitable robotic surface explorer impractical from both a technological and a budgetary standpoint. Highly-trained scientist-astronauts, however, should be able to cope with the environment and deliver a level of scientific return more than worth the expense of mounting the mission.

The actual strategies and techniques for NEA surface exploration are still under development, but some operational factors seem clear even at this early date. Dupont and coworkers at NASA Johnson Space Center's Lunar and Mars Exploration Program Office examined asteroid surface exploration techniques when studying Phobos mission options (Dupont et al. 1991). They noted that the asteroid environment is distinctly different from that found on the lunar or Martian surfaces. Explorers will encounter very close horizons, rapid day-night cycles, unique, varying topography, uniform colors due to dust or regolith, and extreme lighting conditions (e.g., harsh shadowing). All of these factors will confront the astronaut with confusing, conflicting visual orientation cues, or perhaps none at all.

Mobility will be the explorer's biggest problem. Some aspects of the

NEA environment will enhance mobility compared to planetary surfaces: the ease of terrain traverse and obstacle avoidance by flight; and the ability to reach nearly the entire surface of the NEA from any point in orbit. But NEA surface conditions also impose penalties: the need for a capable guidance and navigation system for orbital or surface operations; fuel expenditure for transit from the Mars cruise vehicle to the surface; an astronaut transporter with significantly greater propellant requirements than existing EVA systems; and the possibility of thruster-produced surface dust and attendant contamination of equipment, work areas and surface samples.

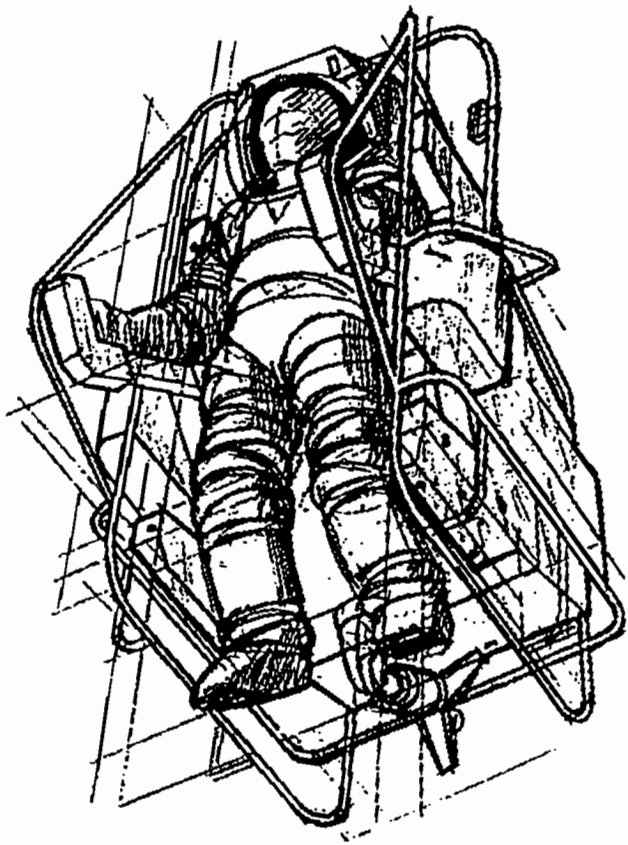


Figure 11. Personnel Maneuvering Unit showing anchor.

For both global exploration and transit between the mother craft and surface, the best solution may be a Personal Maneuvering Unit (PMU) (Fig. 11) derived from the shuttle's cold-gas powered Manned Maneuvering Unit. The PMU will require anchoring devices, flood lights, versatile navigation gear, a gimbaled work platform for surface operations, and about 75 m s^{-1} ΔV capa-

bility, compared with the current MMU's 20 to 25 m s^{-1}. A dual-propellant thruster system may be required, one a conventional bipropellant hot-gas system for efficient orbital maneuvers, the other a cold-gas system for delicate surface operations.

Sustained, day-long EVAs in a near-weightless environment will be the most daunting physical challenge that human explorers face at an NEA. Rather than the familiar Apollo geology traverse, NEA surface operations will be more akin to performing field work strapped to a vertical rock face, where both tools and samples leave the vicinity with the slightest provocation. To cope with the micro-*g* environment and do productive science, the astronaut's modified lunar EVA suit will have to work in concert with a system of anchoring devices and work platforms.

An overall exploration strategy would include the following elements: after an extensive survey of the NEA from the Mars vehicle, the crew would select a number of scientific sites for detailed investigation using the EVA suit/PMU combination. After transit to the surface, the exploration team would emplace an anchoring network and perform intensive field work in an area a few tens of meters in radius. Short hops to adjacent areas would broaden the sample collection and take advantage of new discoveries. Operations would then shift to another scientific site. Suit limitations would impose an approximate 8-hr EVA stay-time on the surface crew, but by using alternating teams on successive days, a comprehensive exploration of the surface of a small NEA could be accomplished in about 2 weeks.

Such a strategy will be well within our experience base. By the time an NEA mission is mounted, astronauts will have some 40 yr of EVA experience under their collective belts, gained from Apollo, Freedom, and SEI lunar operations. During Apollo, humans demonstrated just the sort of field exploration capacity required for NEA exploration. Even in the harsh NEA environment, human presence will provide: (a) efficient, adaptable science exploration; (b) interactive orbit control and controlled flight over the entire asteroid surface; (c) global, frequent surface access using upgrades of existing EVA systems; (d) precise resource assessment and plant siting; and (e) perhaps most important, a large, selective sample return. Based on past human flight experience, the versatility and adaptability of human explorers will enhance significantly the probability of satisfactory scientific return and overall NEA mission success.

In summary, human NEA exploration is within the projected astronaut operational capability contemplated for SEI. There are major, unresolved issues associated with deep space flight: human adaptation to micro-*g*, life support, crew interaction and human factors, and space radiation hazards, to name a few. All must be confronted and solved by any Mars exploration effort, and are not unique to NEA missions. The Synthesis Group has identified technology developments or improvements required for NEA as well as Mars exploration: EVA suit, nuclear propulsion, telerobotics, zero-*g* countermeasures, radiation effects and shielding, automated rendezvous and docking of

large spacecraft, and human factors involved with long-duration space flight. Assuming these problems can be solved, the development of a Mars transportation capability (even a robust lunar capability) will inherently provide the means for safe and practical NEA exploration missions. Considering the size of the potential science return from even a single NEA mission, failure to capitalize on the SEI investment by undertaking asteroid exploration would be a classic case of missed opportunity and lack of vision.

C. Programmatic Prerequisites for NEA Missions in the Space Exploration Initiative

A two-phase program to discover and characterize the near-Earth asteroids is an essential prerequisite to sending human beings to these bodies. First, we must discover them and do the necessary groundbased observing to characterize the asteroids in a preliminary fashion. Second, several asteroids need to be explored by precursor robotic missions. Such a strategy is advocated by the Synthesis Group. In the committee report *America at the Threshold* (Stafford et al. 1991), the most productive near-term precursor for asteroid exploration is "an expanded Earth-based survey to find the most favorable mission targets." Following this enhanced cataloging phase is to be a characterization operation featuring robotic missions to "provide imagery, spectroscopy, structural information and sample returns from exploitable asteroids."

The current discovery rate of near-Earth asteroids, about 30 to 40 objects per year, needs to be increased by a factor of at least 10 if we are to take advantage of the potentially good missions to these bodies. The scientific community is doing an excellent job of increasing the discovery rate recently; despite very limited resources, the discovery rate was 25 in 1990 and is already 20 in the first half of 1991. This significant increase is due largely to the discoveries of the Spacewatch program of Gehrels and coworkers at the University of Arizona and the "down under" AANEAS program in Australia by Steele and McNaught. A 20-yr program to discover more than 90% of the population of the 1000 to 2000 asteroids larger than 1 km in diameter has recently been proposed by a NASA study (Morrison 1992). This search would also discover a significant (~40–50%) sample of those asteroids larger than 0.5 km in diameter and would almost certainly yield several more asteroids that could be reached with a mission ΔV similar to that of the recently discovered 1991JW. Techniques to expand the existing NEA searches have been proposed by Helin and Dunbar (1990), Gehrels et al. (1991), and Morrison (1992). Accompanying the expanded discovery program must be a program to perform the physical observations needed to characterize the objects: photometry, radiometry, spectroscopy, radar, etc. This last technique is expected to be particularly useful for studying NEAs, following the upgrades now under way to the Arecibo and Goldstone radar observatories (Ostro 1989).

For several reasons, robotic exploration of several NEAs is needed prior to sending humans. *In-situ* exploration is necessary to validate and to expand

upon what we can learn from groundbased observations. Are the surfaces of NEAs covered with a powdery regolith, or are they bare rock, metal, or rock-metal mixtures? If there is a regolith, how deep is it? Are the surfaces homogenous in composition or is there geological structure? These and many other questions would be addressed with precursor robotic missions.

Several missions would be needed in order to sample the many different classes of asteroids, and the differing materials that are to be found on the different classes. (Multi-asteroid rendezvous missions for NEAs typically require large ΔV budgets and/or long flight times.) Also, robotic exploration may not include targets for astronaut exploration due to the long interval between favorable mission opportunities to the same asteroid. Thus several asteroids need to be explored by robotic spacecraft in order to understand the range of environments expected to be found within the NEA population.

The first robotic exploration is actually under consideration now as part of the Discovery program at NASA Headquarters. In addition to the science instrumentation (see the Discovery [1991] report), a desirable experiment for missions to suspected extinct comet nuclei would be subsurface exploration for volatiles, perhaps using penetrators or by simply setting off explosive charges on the surface of the body. Besides the science return and resource assessment, robotic missions would provide the information needed to plan astronautic exploration of NEAs, such as how to collect and preserve samples, what type of experiments should astronauts conduct, what type of science package should remain on the asteroid following the mission, etc.

D. Long-Term Benefits from NEA Exploration

The long-term benefits from the NEA exploration will be in the areas of resources and science. The resources available from NEAs are important, diverse, and strategically located for large-scale space operations. Transport of resources from asteroids—water, metals, or just shielding mass, for example—is cheap in terms of ΔV, provided aerocapture is used at Earth (or at Mars or Venus to aid capture into Earth orbit; see Lewis and Lewis 1987). The exploitation of these asteroidal resources is essential to large-scale developments in space, and the assaying of these "resource bodies" will most likely be the dominant long- term payoff from NEA missions.

Astronauts will be essential to developing the infrastructure needed to extract the asteroid resources for use in LEO, GEO, or on the Moon. The exploration of NEAs as part of SEI will provide the knowledge base from which the large-scale development of NEA resources can be planned.

Scientific studies of NEAs will address fundamental questions about the formation and evolution of our solar system. These bodies almost certainly include samples of very primitive cometary and asteroidal bodies as well as fragments of parent bodies which underwent considerable thermal processing. As many NEAs are fragments of broken up parent bodies, they will give us samples of the interior of larger bodies which probably formed as planetesimals in the early solar system. Through close-up studies of several near-Earth

asteroids, we will be better able to unravel the history of our solar system.

Acknowledgments. We thank E. Helin, C. R. Chapman, and S. J. Weidenschilling for critical comment on the manuscript and C. Pilcher and W. Quaide for timely support of this project. The Planetary Science Institute is a division of Science Applications International Corp.

REFERENCES

Bell, J. F., Lucey, P. G., Gradie, J. C., Granahan, J. C., Tholen, D. J., Piscitelli, J. R., and Lebofsky, L. A. 1989. Reflection spectroscopy of Phobos and Deimos. *Bull. Amer. Astron. Soc.* 21:991 (abstract).

Chapman, C. R., and Morrison, D. 1989. *Cosmic Catastrophes* (New York: Plenum Press).

Davis, D. R., Hartmann, W. K., Friedlander, A., Collins, J., Niehoff, J., and Jones, T. 1990. The Role of Near-Earth Asteroids in the Space Exploration Initiative. SAIC Study No. 1-120-232-S28.

Discovery. 1991. Near-Earth Asteroid Rendezvous (NEAR), Report of the Discovery Science Working Group, Executive Summary, October.

Dupont, A., Blackshear, J., Bailey, P., Ewan, P., and Kincade, R. 1991. Treatise on Spaceflight Amid a Two-Body Influence, Especially Related to the Mars-Phobos System and the Exploration of Phobos. Unpublished manuscript.

Farquhar, R., Jen, S.-C., and McAdams, J. V. 1993. Extended Mission Opportunities for a Discovery Class Asteroid Rendezvous Mission. Paper AAS93-127, presented at the AAS/AIAA Spaceflight Mechanics Meeting, Feb. 22–24, Pasadena, Ca.

Friedlander, A. L., Davis, E. R., and Heppenheimer, T. A. 1974. Measurement Error Analysis in Determination of Small-Body Gravity Fields. AIAA Paper No. 74-218.

Gehrels, T. 1991. Scanning with charged-coupled devices. *Space Sci. Rev.* 58:347–375.

Hartmann, W. K. 1987. The relationship of active comets, "extinct" comets, and dark asteroids. *Icarus* 69:33–50.

Helin, E. F., and Dunbar, R. S. 1990. Search techniques for near-Earth asteroids. *Vistas in Astron.* 33:21–37.

Helin, E. F., and Lawrence, K. 1991. *IAU Circ. 5266*, 13 May.

Helin, E. F., and Shoemaker, E. M. 1979. The Palomar planet-crossing asteroid survey, 1973–1978. *Icarus* 40:321–328.

Jewitt, D., and Luu, J. 1989. On the relative numbers of C types and S types among near-Earth asteroids. *Astron. J.* 98:1905–1911.

Jones, T., Lebofsky, L., Lewis, J. S., and Marley, M. S. 1990. The composition and origin of C, P, and D asteroids: Water as a tracer of thermal evolution in the outer belt. *Icarus* 88:172–192.

Lau, C. O., and Hulkower, N. D. 1983. On the Accessibility of Near-Earth Asteroids. AAS Paper No. 85-352.

Lewis, J. S., and Lewis, R. A. 1987. *Space Resources: Breaking the Bonds of Earth* (New York: Columbia Univ. Press).

McAdams, J. V. 1992. Mission options for rendezvous with the most accessible near-Earth asteroid—1989 ML. *J. Astron. Sci.* 40:351–368.

Morrison, D. 1992. *The Spaceguard Survey: Report of the NASA International Near-Earth-Object Detection Workshop*, ed. D. Morrison (Pasadena: Jet Propulsion Laboratory).

NASA. 1989*a*. Report of the 90 Day Study on Human Exploration of the Moon and Mars. NASA Internal Report, Nov. 20.

NASA. 1989*b*. *Exploration Studies Technical Report*. FY 1989 Annual Report.

NEAR (Near-Earth Asteroid Rendezvous) Science Working Group. 1986. Report. JPL No. 86-7.

Niehoff, J. C. 1977. Round-trip mission requirements for asteroids 1976AA and 1973EC. *Icarus* 31:430–438.

Ostro, S. 1989. Radar observations of asteroids. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 192–212.

Ostro, S., Chandler, J. F. Hine, A. A., Rosema, K. D., Shapiro, I. I., and Yeomans, D. K. 1990. Radar images of asteroid 1989PB. *Science* 248:1523–1528.

Shoemaker, E. M., and Helin, E. F. 1978. Earth-approaching asteroids as targets for exploration. In *Asteroids: An Exploration Assessment*, NASA CP-2053, pp. 245–256.

Shoemaker, E. M., Wolfe, R. F., and Shoemaker, C. S. 1990. Asteroid and comet flux in the neighborhood of Earth. In *Global Catastrophes in Earth History*, eds. V. I. Sharpton and P. D. Ward, Geological Soc. of America SP-247 (Boulder: Geological Soc. of America), pp. 155–170.

Stafford, T. P., and the Synthesis Group. 1991. *America at the Threshold: Report of the Synthesis Group on America's Space Exploration Initiative* (Washington, D. C.: U. S. Government Printing Office.

Stancati, M. L., and Soldner, J. K. 1981. Near-Earth Asteroids: A Survey of Ballistic Rendezvous and Sample Return Missions. AAS paper 81-185.

Yen, C. L. 1989. Mission opportunity maps for rendezvous with Earth-crossing asteroids. *J. Astronaut. Sci.* 37:399–415.

PART IV

Mars and Beyond

THE PHYSICAL AND CHEMICAL PROPERTIES AND RESOURCE POTENTIAL OF MARTIAN SURFACE SOILS

C. R. STOKER
NASA Ames Research Center

J. L. GOODING
NASA Johnson Space Center

T. ROUSH
San Francisco State University

A. BANIN
Hebrew University

D. BURT
Arizona State University

B. C. CLARK
Martin Marietta Aerospace

G. FLYNN
State University of New York, Plattsburgh

and

O. GWYNNE
NASA Ames Research Center

The physical and chemical properties of Martian surface soils are reviewed from the perspective of providing resources to support human activities on Mars. The relevant properties can only be inferred from limited analyses performed by the Viking Landers, from information derived from remote sensing, and from analysis of the SNC meteorites thought to be from Mars. Several lines of evidence suggest that the Martian surface is globally covered with a fine-grained soil having nearly uniform physical and compositional properties. Viking experiments characterized the physical properties of the soil well enough to allow preliminary designs for resource extraction equipment. The mineralogy of the soil has not been determined but the bulk elemental compositions determined by the Viking Lander X-ray Spectrometer analyses provide evidence for

clay minerals (possibly smectites) or mineraloids (palagonite) admixed with sulfate and chloride salts. The soil unit contains materials bearing useful amounts of H_2O, Fe, Al, Mg, S and Cl. Spectroscopic observations of Mars provide additional constraints on the composition of the soil and the best match to Martian spectra is obtained with certain terrestrial palagonites which contain poorly crystalline iron oxides. Many lines of evidence suggest that the SNC meteorites are samples of Martian crustal materials and thus provide information about the parent minerals from which Martian soil is derived. A number of trace mineral phases have been identified in the SNCs including carbonates, sulfates, and possible smectites. Martian surface materials can be used directly in a number of ways. Martian soil, with appropriate preconditioning, can probably be used as a plant growth medium, supplying mechanical support, nutrient elements, and water at optimal conditions to the plants. Loose Martian soils could be used to cover structures and provide radiation shielding for surface habitats. Martian soil could be wetted and formed into duricrete bricks with strength comparable to concrete for use in construction. Useful elements and compounds can be extracted from Martian soil. Water, one of the most valuable resources, probably exists in hydrated minerals although no specific mineralogical identifications have yet been achieved. Viking experiments showed that water evolved from Martian soils when they were heated to 500 C suggesting the soils contained at least ~1% loosely bound water. Many minerals which might exist on Mars, including gypsum (21% H_2O), kieserite (13% H_2O), goethite (10% H_2O), and nontronite (5–6% H_2O), contain tightly bound (structural) water which could be released by pyrolyses but the required energy input would vary with the identity and abundance of the mineral. The high sulfur content and the implication that sulfur may occur as a soluble salt suggests that sulfur can be readily extracted from Martian soil. Sulfur and sulfur-bearing compounds are used in the manufacture of a wide variety of useful substances including acids, bases, oxidizing and reducing agents, fertilizers, dyes, catalytic agents, detergents, solvents, explosives, etc. Machinable metals such as iron, magnesium and aluminum could also be extracted from Martian soils. Martian soil contains 18 wt% iron oxide which is likely to be in the form of poorly crystalline or nanocrystalline iron oxide minerals. The most important information needed to design resource extraction procedures is to determine the mineralogy of the Martian soil.

One of the most stunning results of the Viking mission to Mars was the realization that the planet possesses, in some accessible form, many or all of the resources needed to sustain human life (Clark 1984). Compared to the Moon, Mars is a resource garden of eden, with a ready source of volatiles residing in the atmosphere (Meyer and McKay 1984) as well as volatiles, useful compounds and metallic elements residing in the soil (Clark 1984). Due to the long supply lines from Earth and the large cost of transportation between Earth and Mars, the use of Martian resources may prove to be essential for enabling human exploration and settlement of Mars.

The properties of the Martian atmosphere have been sufficiently well characterized by spacecraft missions to permit the design of resource extraction procedures. The soils of Mars are less well understood than the atmosphere, but enough is known to begin to develop concepts for using resources derived from the soils. The soil is an attractive source of resources on Mars because there is strong evidence, reviewed below, that Mars is covered with a globally distributed soil unit having quite uniform properties. Thus, the soil resource can be used at virtually any location on Mars and extensive prospecting will

not be necessary before choosing a landing site with access to resources.

The purpose of this chapter is to review what is known about the physical and chemical properties of Martian soil and to suggest what resources might be extracted from the soils, as well as how the soil itself might serve as a resource for human settlement. We begin (Sec. I) by reviewing the physical properties of the soil as deduced from experiments on the Viking Lander. We next discuss (Sec. II) what is known about the chemical properties of Martian soil from Viking *in-situ* measurements. The elemental composition of Martian soil was derived from the Viking Lander X-ray fluoresence spectrometer experiments, but the Viking Landers did not provide information about the mineralogy beyond what could be inferred from the elemental analysis. Thus, in Sec. III we review the information about the soil mineralogy that can be inferred from spectroscopic observations of Mars in the visible and infrared. Section IV focuses on the implications for soil mineralogy that can be deduced from analyzing the composition of the SNC meteorites which are thought to have originated on Mars and therefore could represent the parent minerals from which Martian soil is derived. A significant component of Martian soil is likely to be derived from the influx of meteorites onto the surface of Mars and thus elements are expected to be present in Martian soil which are present in meteorites and would otherwise be quite rare. The meteoritic component is discussed in Sec. V.

The remaining sections of the chapter discuss strategies for using resources derived from the soils. Here the use of the soil resource falls into two main categories: (1) use of the soils directly (with little or no processing) as discussed in Sec. VI; and (2) resources extracted from Martian soil. Relatively little work has been done along the lines of designing strategies for using the soil or soil-derived resources and this is an area ripe for future work which we hope will be aided by the detailed review of the soil properties given here.

I. PHYSICAL PROPERTIES OF SOIL

A. Material Types and Physical Properties

Most of the information about the physical properties of Martian surface materials has been derived from experiments that were performed by the Viking Landers as described in detail by Moore et al. (1987). The physical and chemical properties of Martian surface materials at the Viking landing sites have also been recently reviewed by Arvidson et al. (1989*a*) and much of the following discussion derives from these two papers.

Figure 1 shows images taken by Lander cameras of the Viking 1 and 2 landing sites. Five types of surface materials were characterized (Moore et al. 1977,1978,1979,1982,1987) including drift material, clods and crusts, blocky material, rocks, and features thought to be outcrops of bedrock. The Viking Lander 1 site has two types of fine-grained sediment deposits: drift and blocky material. The drift material, which has the "consistency of loose kitchen flour" (Arvidson et al. 1989*a*) covers about 14% of the Viking Lander 1 site. Blocky

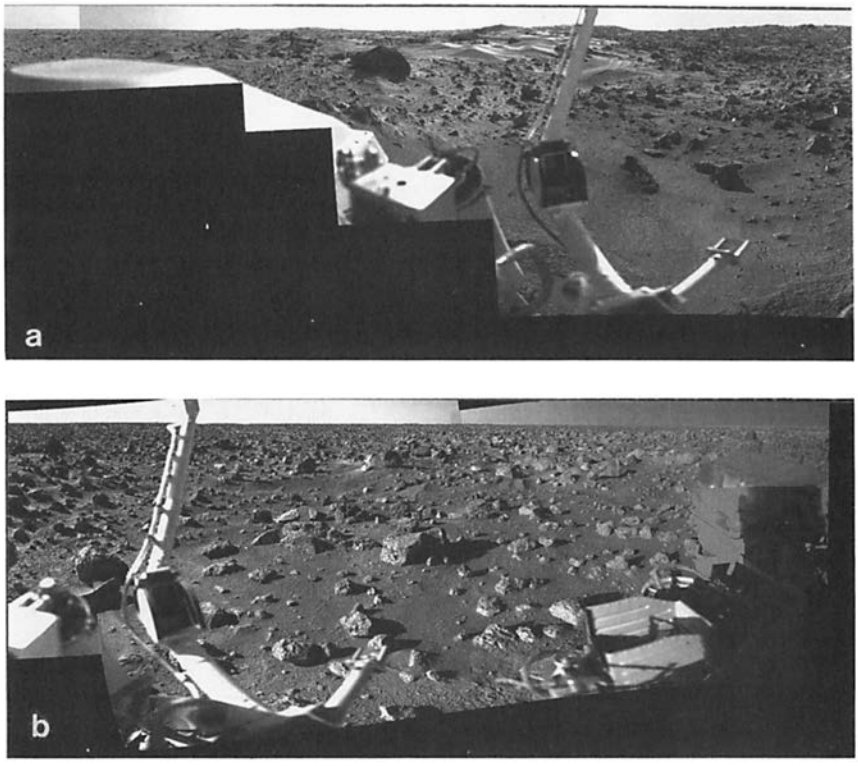


Figure 1. Panoramic views of the Martian surface. Mosaic images generated from VL1 (a) and VL2 (b) camera systems. At the VL1 site, large tracts of dune-like drifts are superposed on a rocky substrate. Rocks range in size from a few cm to 3 m and most rock surfaces are pitted, though some are smooth. The VL2 site has much less drift material which is confined to wind tails among rocks. Rocks are somewhat larger than at VL1 but show the same general morphology. Areas between rocks at both sites appear to be littered with small fragments or clods of cm or smaller size. Images courtesy of H. Moore, U. S. Geological Survey.

material, having the apparent consistency of "dry cloddy garden soil," was also present in the rocky area in front of the Lander where it was usually overlain by drift material. Rocks sat on top of the blocky material and some were set in a matrix of it. Trenches made in the blocky material by the surface sampler experiment left a rubble formed of distinct fragments and clods. The blocky material has been referred to as duricrust in some publications (see, e.g., Mutch et al. 1977) although this term usually refers to a product of deep chemical weathering under conditions of abundant moisture which might not be the genesis of the Martian crusts (Sharp and Malin 1984).

The sediments present at the Viking Lander 2 site were called "crusty to cloddy material" (Moore et al. 1987) because, when trenches were made into the soils, the uppermost layer broke into thin plates of crusts or thicker clods. The strength of these materials was intermediate between the drift and blocky material seen at the Lander 1 site; they were easily disaggregated and did not survive the sieving process used to separate fine- from coarse-grained materials. An attempt was made at both Viking sites to sieve and sample small pebble-sized rock fragments but no rock fragments in this size range were obtained from the soil at either site. (In this chapter, we use the term soil to denote unconsolidated material on the surface of Mars and includes both wind-blown sediments and material that has formed in place by the comminution, decomposition, or alteration of rocks or materials derived from rocks.)

Chemical analysis of the blocky material showed that it had up to 50% higher concentration of sulfur and chlorine, two salt-forming elements, than did drift materials (Clark et al. 1982). Thus, the blocky material may be formed from the fine-grained drift material by a cementation process involving a concentration of salts (Clark and Van Hart 1981; Clark et al. 1982). The mechanical properties of drift, blocky, and crusty material were estimated based on experiments performed with the backhoe soil scoop on the surface sampler experiment, by lander footpad penetration during landing, and by analysis of sample trenches and forces inferred from motor currents (Moore et al. 1987). Table I shows the physical properties inferred for each type of material. Angles of internal friction for the blocky and crusty material are comparable to those of many dry terrestrial soils, which average in the range 26 to 34° (Gibbs et al. 1961). The angle of internal friction of the drift material is 18°, considerably lower than expected for moderate density terrestrial soils and is suggestive of porous compressible soils with low bulk densities. The cohesions of the soils at the landing sites are smaller than those of most dry, fine-grained terrestrial soils (Arvidson et al. 1989*a*).

An attempt was made to sample fragments of rock surfaces by chipping at the rocks with the sampler backhoe. Despite loads of up to 10 MPa, the rocks could not be chipped or scratched and thus do not appear to have a friable weathered rind. The results of the gas exchange experiments (Ballou et al. 1978) suggest that the sizes of individual grains within the drift and clods and crusts materials range from 0.1 to 10 mm in diameter (Table I). The blocky material was not sampled by this experiment and so its grain size was not determined but the similar chemical composition of the blocky, drift, and clods and crusts materials (Toulmin et al. 1977; Clark et al. 1982) imply that they are all composed of the same basic soil unit having clay-to-silt-sized grains which have undergone different degrees of induration.

The physical properties of Martian soils are important for designing equipment which can be used to extract resources from the soils. Viking experiments provide the best available information on the physical characteristics of the Martian soil which can provide guidance in designing vehicles

TABLE I

Summary of Physical Properties of Surface Materials in the Sample Fields at the Viking Landing Sites[a]

	Drift Material	Crusts and Clods	Blocky material	Rocks[b]
Angle of internal friction (degrees)	18.0±2.4	34.5±4.7	30.8±2.4	40–60
Cohesion (kPa)				
Mean	1.6±1.2	1.1±0.8	5.5±2.7	—
Range	0–3.7	0–3.2	2.2–10.6	10^3–10^4
Bulk Density (kg m^{-3}	1150±150	1400±200	1600±400	2600
Sizes				
Grains (mm)	0.1–10	0.1–10	0.1–10 & larger	—
Clods (m)	lumps	0.04	0.04	—
Rock fragments (m)	none	> 0.035	≥0.001–0.035	0.035–1.0
Fractional area covered				
Viking Lander 1	0.14	—	0.78	0.08
Viking Lander 2	—	0.86	—	0.14

[a] Table from Arvidson et al. 1989*a*.

[b] Rock properties were not measured. Typical values for terrestrial crystalline rocks are included for comparison.

for use on Mars. Table II summarizes some of the relevant physical properties of the soils. Unfortunately, the available information does not include some of the most important parameters needed to design vehicles to work on Mars. One of the biggest challenges for engineers designing rovers for Mars is determining how the vehicles will interact with the surface. Rovers will have the same wide range of uses on Mars as they do here on Earth, including everything from moving people and equipment to bulldozing and compacting soil. The historical method for approaching this problem has been derived by Bekker (1969).

TABLE II
Physical Properties of Martian Soil

Bolometric albedo[a]	0.09 to 0.43
Surface thermal inertia[a]	0.67 to 4.6×10^5 J m^{-2} s$^{-1/2}$ K^{-1}
Thermal conductivity[b,c]	0.75 and 1.13×10^3 J m^{-1} s^{-1} K^{-1} (for VL1 and VL2 respectively)
Thermal emissivity[b,c]	~1
Specific heat capacity[b,c]	6.087×10^{-2} (T–220) J kg^{-1} K^{-1}
Angle of internal friction of mixed crusts and fines[d]	28–34
Cohesion of mixed crusts and fines[d]	0.2–2.3 kPa
Bulk density of mixed crusts and fines[d]	1400 ±200 kg m^{-3}

[a] From Kieffer et al. 1977.
[b] Note, these values were not measured and are nominal values determined by Kieffer et al. 1977.
[c] From Kieffer 1976.
[d] From Moore et al. 1987.

The relationship between the load and the sinkage of a tire is given by:

$$p = kz^n \tag{1}$$

where p is the load or vertical average contact pressure, z is the sinkage distance, k is the modulus for soil sinkage, and n is the exponent for sinkage. The two parameters that must be derived in order to determine the vehicle interaction with the soil are the modulus for soil sinkage (k) and the exponent for sinkage (n). The equation for the amount of force needed to overcome the friction, sinkage and soil buildup is the rolling resistance R (in Newtons). It is give by Bekker (1969) as:

$$R = \frac{(3\mathrm{W})^{\frac{2n+2}{2n+1}}}{(3-n)^{\frac{2n+2}{2n+1}}(n+1)(k_c + bk_\phi)^{\frac{1}{2n+1}} D^{\frac{n+1}{2n+1}}} \tag{2}$$

where W is weight of the vehicle, b is the width of the tires, D is the diameter of the tires, and k_c and k_ϕ are soil coefficients, related to k by:

$$k = k_c/b + k_\phi. \tag{3}$$

Here b is the width or radius of the object in contact with the ground, k_c and k_ϕ are coefficients of soil variation.

Using information about the sinkage of the footpads from the two Viking Landers, and the pressure force exerted by each footpad, we can deduce values for the parameters in Eqs. (1) and (3). Figure 2 shows a plot of the sinkage as a function of the applied pressure. By fitting these data to a power law, we derive parameters used in Eq. (1) of $k = 1481.0$ kPa and $n = 1.8$. Using the two different footpad radii, Eq. (3)can be solved to obtain $k_c = 306.8$ kPa and $k_\phi = 20.0$ kPa m^{-1}. Table III compares the soil constants obtained from Viking results with those of some typical Earth soils.

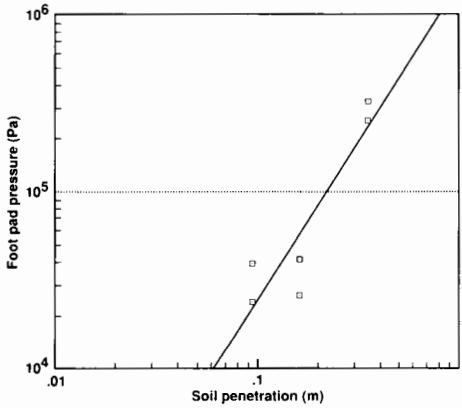


Figure 2. Data for sinkage of Viking Lander 1 and 2 foot pads. The data are fit to a power law (solid line) and used to solve Eq. (1).

We caution that the soil properties for Mars (Table III) are only rough values based on limited data derived from the sinkage of the Lander footpads, an experiment which was not designed to measure such parameters, and should be refined by *in-situ* experimental results. In order to determine the soil constants, a typical procedure is to use round plates and to conduct at least two soil penetration tests using plates of different diameters. Similar tests should be performed on Mars in various locations to determine these parameters.

B. Magnetic Properties

The Martian surface soils were found to have significant magnetic susceptibility. Each of the Viking Landers was equipped with three Sm-Co permanent magnet arrays which were arranged as a central circular magnet surrounded by a concentric ring-shaped magnet. One of these arrays was mounted on the top of the Lander next to a photometric reference test chart, while the other two were mounted on the surface sampler backhoe. These two arrays had "strong" and "weak" magnetic field strengths of 2500 G and 700 G,

TABLE III

Comparison of Soil Parameters for Martian Soils as Measured by Viking with the Range of Values Seen for Typical Earth Soils[a]

Dry Sand[b]	Sandy Loam[b]	Clay[b]	Mars Soil[c]
$k_c = 1\,\text{kPa}$	$k_c = 1 - 52\,\text{kPa}$	$k_c = 2 - 105\,\text{kPa}$	$k_c = 306.8\,\text{kPa}$
$k_\phi = 1527\,\text{kPa/m}$	$k_\phi = 3 - 1500\,\text{kPa/m}$	$k_\phi = 103 - 1723\,\text{kPa/m}$	$k_\phi = 20.0\,\text{kPa/m}$
$n = 1.1$	$n = 0.2 - 0.9$	$n = 0.3 - 0.9$	$n = 1.8$
$c = 1.0\,\text{kPa}$	$c = 1.4 - 17.2\,\text{kPa}$	$c = 1.0 - 70\,\text{kPa}$	$c = 0.2 - 2.3\,\text{kPa}$
$f = 28\,^\circ$	$f = 11 - 38\,^\circ$	$f = 6 - 34\,^\circ$	$f = 28 - 34\,^\circ$

[a] c is the soil cohesion, f is the angle of internal friction.

[b] From Bekker 1969.

[c] Derived from Viking 1 and 2 results.

respectively. The magnetic susceptibility of the surface soils was estimated by imaging the amount of material collected on the magnet arrays with the Lander camera (Hargraves et al. 1977,1979). Magnetic particles collected on both weak and strong magnets mounted on the backhoe during each sample operation. The variation of magnetic properties between samples was difficult to judge because both backhoe arrays became saturated with particles during the first sampling operation. Magnetic particles also collected on the magnet array mounted on the top of the Lander. The source of the material was thought to be dust raised by the firing of retro-rockets on landing and by the surface sampling operations. The amount and grain size of the magnetic material appeared to be similar at both the Viking landing sites. Based on spectrophotometric analysis, the material collected by the magnet arrays was indistinguishable in color and brightness from the normal surface soil material exposed in the trenches.

Based on the maximum amount of material held on the strong magnet at Viking Lander 2 site, and assuming a probable grain size of the material of $<100\,\mu$m (Shorthill et al. 1976), the effective magnetic susceptibility of the surface soils was estimated (Hargraves et al. 1977). The amount of material that can be held on the magnets is a function of the effective magnetic susceptibility of the magnetic material, the grain size, and the presence of composite grains. The data were consistent with the soil grains being composed of 1 to 7% highly magnetic material. Because the whole soil adhered to the magnets, the magnetic component appears to be intimately mixed with the rest of the material. The mineralogy of the magnetic phase is unknown and an X-ray fluorescence spectrometer analysis of materials adhered to the magnets was not technically possible. The highly magnetic material could be discrete grains of magnetite, pyrrhotite, iron with a red iron oxide coating, maghemite (Hargraves et al. 1977,1979] or titanomaghemite (Coey et al. 1990). Laboratory studies with terrestrial palagonititic tuffs and soils have shown that many of them have magnetic susceptibilities comparable to that inferred for Martian soils (Morris et al. 1989,1990). The magnetic mineral in palagonites is thought to be ultrafine-grained (superparamagnetic) hematite which is a common pigment in oxidized, iron-rich soils on Earth. Terrestrial palagonites have also been shown to be close spectral analogs to the soils of Mars as discussed below.

II. CHEMICAL PROPERTIES FROM VIKING MEASUREMENTS

The Viking Lander missions gave the first *in-situ* data relevant to the elemental composition of the Martian soils. Although no comprehensive mineralogical analysis was possible, and attempts to analyze rocks was never a mission objective and likewise was not possible, the compositional data remains the most useful currently available. Elemental analysis was performed using X-ray fluorescence spectrometry (XRFS), but all mineralogical information

must be inferred from the constraints placed by the concentrations of elements and, where possible, other observations.

A. Composition of Soils

The results of the chemical analysis of Martian soils by the Viking XRFS experiment are reported in a number of publications (Baird et al. 1976,1977; Clark et al. 1976,1977; Toulmin et al. 1976,1977; Clark et al. 1982; Arvidson et al. 1989*a*) and are only briefly reviewed here. Several major factors must be considered in interpreting the results of the XRFS experiments. Because of instrument design limitations, the XRFS could not detect elements of atomic number less than 12. Thus, C, N and Na, three important mineral-forming elements, could not be detected. Thus, possible concentrations of carbonates, nitrates and possible Na-bearing compounds are unknown. Samples were introduced to the XRFS via a funnel which included a wire screen with 1.25 × 1.25 cm openings. Table IV shows the elemental concentration of the Martian soils as determined by the XRFS experiment, reported as the oxides of the elements detected.

TABLE IV

Average Weight Percent Composition of Sediments Determined by the Viking Lander X-ray Fluorescence Spectrometers[a]

	Viking Lander 1	Viking Lander 2	Total Absolute Uncertainty
SiO_2	44	43	±6
TiO_2	0.62	0.54	±0.25
Al_2O_3	7.3	~(7)[b]	±4
Fe_2O_3	17.5	17.3	−2 to +5
MgO	6	~(6)[b]	−3 to +5
CaO	5.7	5.7	±2
K_2O	< 0.5	< 0.5	±0.5
SO_3	6.7	7.9	−2 to +6
Cl	0.8	0.4	−0.5 to +1.5
Other[c]	2	2	
Total	91	90	

[a] Table after Clark et al. 1982.

[b] Not measured (as a consequence of instrument noise) but assumed to be the same as Viking Lander 1.

[c] Includes P, Mn and Na, none of which could be unambiguously detected by the XRFS, as well as H (from possible water) and C (from possible carbonates) and N (from possible nitrates).

A salient feature of soil composition, long suspected because of the intensely red color of Mars compared to other inner planetary bodies, is its unusually high concentration of iron, almost 20% Fe_2O_3 (when expressed as ferric oxide). The Martian soils clearly are predominantly silicates, but Al in

the samples is low compared to most basalts on Earth and the Moon, while Mg is not sufficiently high to correspond to familiar ultrabasic igneous rocks on Earth. This element profile alone is sufficient to strongly differentiate Martian soils as distinct from lunar soils or most expectations based upon terrestrial soils. However, the Martian soils do have a grossly basaltic composition and the technical challenges of deriving useful things from this soil will generally parallel those for using terrestrial soils formed from weathered basalts.

All soils at the two Viking sites contain very high amounts of sulfate and chloride salts, as indicated by the relatively high concentrations of S and Cl, and the fact that clods and crusts appeared to be cemented together by these salts. Salts are not found at all on the Moon, and occur on Earth typically as either effluorescences (crusts) or in the form of large, homogeneous bedded deposits. An important difference, of course, is that the Earth's oceans contain enormous quantities of salts which are leached from soil by the work of rainfall and flowing water and delivered to the sea in solution. On Mars, soils which formed during an early warm wet period (see, e.g., McKay and Stoker 1989; Chapter by Jakosky and Zent), would be expected to be similarly leached of their salt burden. This suggests that the salts were emplaced in the soil after the early warm wet period. Indeed, the soils now resident on the Martian surface, including their salts, most probably postdate the early warm wet period.

A model Martian soil composition is given in Table V. The mineralogical composition of the silicate group may be any of the following categories: (1) minimally weathered igneous silicates, mainly pyroxenes, feldspars and magnetite; (2) glass of igneous composition, perhaps altered to palagonite or some other assemblage of mostly noncrystalline components; (3) smectite clays of various compositions, including a major component of nontronite. The soils of Mars may actually include members of all three categories. Based upon the elemental analyses and arguments regarding thermodynamic equilibria, the salts are probably predominantly made up of Mg and Na sulfates, with some degree of hydration, with the chlorides probably dominated by NaCl (halite). Mg, Ca and Fe salts could be present as well, and these three cations are the most likely ones if carbonates are indeed present in the soil. Because the XRFS could not directly measure either C or N elemental concentration in the samples, it cannot be ascertained whether carbonates or nitrates are significant. Planar features interpreted as salt-cemented "duricrust" are evident at both landing sites, and may be ubiquitous below the windblown dust deposits. As discussed above, samples of pebble-sized materials (interpreted as duricrust peds and clods) consistently showed higher concentrations of S and Cl, with no other major changes in elemental profiles, indicating that these samples were of salt-enriched soil (Clark et al. 1982).

B. Globally Uniform Martian Surface Soil

Fine-grained materials at both landing sites gave very similar chemical analyses, in spite of their widely separated geographic distance and differences in photogeologic settings. Presumably, the fine materials sampled are sedi-

TABLE V
Model Martian Soil Composition

Silicate minerals	84–79%
Magnetic mineral(s)	3%
Sulfate salts	12%
Chloride salts	1%
Carbonates	0–4%
Nitrates	0–1%

ments dispersed widely by the winds on Mars. Whereas many, if not most, soils on Earth are heterogeneous mixtures of minerals of different density and average grain size, which separate to at least some extent when acted upon by a sorting agent such as the wind, the near-surface Martian soil appears to behave more uniformly when transported. This could be because the mineral constituents are extremely fine particles, but are cemented together into quasi-homogeneous grains which become wind-borne. Good terrestrial soil analogs might be loess or wind-transported marine clays which, like Martian soils, tend to be globally homogeneous in physical properties and chemical and mineralogical compositions. The spectroscopic observations of Mars (see Sec. III) corroborate the hypothesis that large regions of Mars are covered by a uniform soil. On the other hand, rock and larger-grained soil units beneath this mantling layer of wind-blown sediment could be of distinctly different composition. From a resources standpoint, such a globally distributed soil could be of distinct advantage if a number of processes were optimized for extracting various compounds of interest and if it could be counted upon that this material could be found at most or all landing sites.

C. Water Content of Martian Soils

Although the organics detection experiment on Viking, the GCMS analyzer, detected evolution of water upon heating certain samples (Biemann et al. 1977; Biemann 1979), the quantitative values were highly uncertain, ranging from about 0.1% to 1% by weight of the sample. For several reasons, this is probably a lower limit to the true amount. The samples were heated only for a short period of time (30 s) and before they were placed into the GCMS, they were held for periods ranging from several hours to days in an unsealed sample cavity which was thermally conditioned at temperatures up to +15°C, much higher than Mars' ambient temperatures, which undoubtedly caused release of some adsorbed water. In addition, repeated heating cycles sometimes gave evidence of additional release of water. Finally, the samples were shallow ones and may have been representative only of relatively dehydrated fines because of exposure to the hottest portion of the diurnal temperature cycle.

The actual water content of many or most Martian soils could be much higher than 1%, especially for buried soils. Even if free ice is not present, the soil minerals could contain, in addition to absorbed water, several percent of –OH in their structure if clays, oxyhydroxides, or any of a number of other

minerals are present. The salts could contain up to ~10% water of hydration. Smectite clays could also contain large quantities of interlayer adsorbed water. In the most favorable case, an ice-free Martian soil might still liberate 15% or more of its own mass as water vapor when heated sufficiently.

D. Soil Reactivity and Evidence for Oxidants

The results of the Viking surface chemical experiments have led to the widespread belief that there are oxidants in the Martian soil. The key results upon which this hypothesis is based come from the Viking biology experimental package. The data and their interpretation is of interest from the standpoint of Martian resources and surface operations because oxidants, if present, could represent a resource. Additionally, there have been serious expressions of concern regarding the effect that the soil oxidants may have on materials used in robotic missions and on human health and safety in the event of human missions. For human missions, the concern has centered on the possible irritation of the respiratory system or skin due to dust carried into the Martian habitat through the air locks. Finally, some interpretations of the results of the Viking biology experiments lead to inferences about the mineralogy of Martian soils. Here we briefly review the results of the Viking biology package and discuss some possible interpretations (see detailed reviews by Klein 1978; Mazur et al. 1978; Huguenin 1982).

The significant results of the Viking Lander experiments which imply the presence of oxidants include:

1. The soil released O_2 upon humidification in the gas exchange experiment (Oyama and Berdahl 1977,1979) in amounts ranging from 70 to 770 nmoles cm^{-3} (see Table VI). Heat sterilization of the sample prior to humidification did not eliminate the activity indicating that a biological agent was not involved. No additional O_2 evolved when the samples were wetted with a nutrient solution but there was a slow evolution of CO_2, indicating the organics in the medium were oxidized by a weak oxidizing agent.

TABLE VI

A Comparison of Gas-Exchange O_2 and Labeled-Release ^{14}C Results[a]

Sample	Gas-Exchange O_2 (nanomoles cm^{-3}	Labeled-Release CO_2 (nanomoles cm^{-3}
Viking 1 (surface)	770	~30
Viking 2 (surface)	194	~30
Viking 2 (subrock)	70	~30

[a] Table after Klein 1978.

2. In the labeled-release experiment, $^{14}CO_2$ was rapidly released when Martian surface samples were wetted with an aqueous nutrient medium containing ^{14}C-labeled organics. The kinetics of the initial $^{14}CO_2$ release

were similar to the initial O_2 release in the gas exchange experiment humid mode tests. After the initial release tapered off, ^{14}C continued to evolve slowly until the experiment ended. The total abundance of CO_2 produced in the rapid response was suggestive of a nearly total reaction with only one component (formate) of the nutrient medium. The rapid $^{14}CO_2$ release was completely removed by heating to 160°C for 3 hr, partially destroyed at between 40 and 60°C, and unaffected by storage for short periods at 18°C but lost after two to four months storage at 18°C. One interpretation of the results of the labeled-release experiment was that it indicated microbial activity (Levin and Straat 1981). However, the activity could also be explained by oxidation of labeled formic acid by oxides in the soil, yielding $^{14}CO_2$ (Klein 1979; Oyama and Berdahl 1977; Huguenin 1982; Banin and Margulies 1983; Banin and Rishpon 1979).

3. The GCMS failed to detect organics in surface samples and from samples obtained below the surface. The maximum depth sampled was about 10 cm (Biemann et al. 1977; Biemann 1979). The detection threshold was parts per billion (ppb) for organics containing more than two carbon atoms and parts per million (ppm) for one and two carbon organic compounds. As there are at least two mechanisms that could produce organics on Mars, meteoritic infall (Biemann et al. 1977) and *in-situ* production (Hubbard 1979) the absence of organics suggests that a mechanism for destroying them (such as oxidizing compounds) is present.

The chemical activity and lack of organics have generally been attributed to the action of oxidants in the Martian soil (cf. Klein 1978). Klein (1978,1979) suggested that at least three different oxidants are needed to account for the different results of the Viking biology experiments. The release of O_2 from the soil upon humidification and the persistence of this response after heat treatment suggests the presence of a strong thermally stable oxidant. However, a second strong oxidant is required to explain the decomposition of the organics in the labeled-release experiment because the response was thermally labile. In addition, a weak oxidant is required to explain the slow, continued release of CO_2 in the gas-exchange and the labeled-release experiments.

One explanation for the presence of oxidants in Martian soil is that external energy sources interact with atmospheric gases (Hunten, 1979,1987) and/or the soil (Ponnamperuma et al. 1977) to produce H_2O_2 which is thermally labile, accounting for the labeled-release result. The concentration of H_2O_2 required to explain the labeled-release results is about 1 ppm by mass. In addition to peroxide, thermally stable alkali and alkaline Earth superoxides, which could be produced by a variety of energetic processes (Oyama and Berdahl 1977,1979; Ballou et al. 1978) may account for the gas-exchange experiment O_2 release. If each mole of oxidant released approximately one mole of O_2 and, assuming a soil density of 1.5 g cm^{-3}, then the concentration of oxidant would be about 2 to 25 ppm by mass corresponding to the release

of 70 to 770 nanomoles of O_2 (Table VI). Gooding (1990, Appendix B4) reviewed results from the Viking Lander biology experiments and pointed out that concentrations spanning the range of a few ppb to tens of ppm (by weight) can be inferred for the reactants, depending on which assumptions are made regarding the details of gas production.

A number of experiments have shown that alkali and alkaline earth superoxides including KO_2, ZnO_2 (Ponnamperuma et al. 1977); CaO_2 (Ballou et al. 1978) and MnO_2 (Blackburn et al. 1979) can reproduce the rapid O_2 release in the gas-exchange humid mode experiments. Along these lines, Plumb et al. (1989) have closely duplicated both the labeled-release and the gas-exchange rapid response results using a mixture of KNO_3 and $CaCO_3$ irradiated with ultraviolet light.

It is important to stress that there are alternative explanations for the Viking biology experiments that do not posit the existence of oxidants on the surface. These include the suggestion by Banin and coworkers (Banin and Rishpon 1979; Banin and Margulies 1983) that the gas released in the Viking Biology experiments was due to intrinsically reactive clays. These workers have reproduced the labeled-release response with iron-montmorillonite which catalyzes the oxidation of formate, although the thermal sensitivity of the labeled-release response was not duplicated. However, this model could also explain the puzzling results of the pyrolytic-release experiment (Horowitz et al. 1977) which showed that small quantities of atmospheric carbon were fixed from CO_2 gas, presumably into organic compounds, by Martian soils under conditions of both light and darkness. In a simulation of the pyrolytic-release experiment, Hubbard (1979) reported the synthesis of formate on iron clays provided by Banin. Thus, the reactive clays may be able to account for the results of the pyrolytic-release experiment in addition to the labeled-release experiment because the reduction of formate from CO_2 is the inverse of the catalytic oxidation reaction.

The above discussion serves to illustrate that numerous models for soil properties have been proposed to account for the Viking biology experiments and the presence of oxidants in the soil is by no means certain. Even if oxidants are present, as can be seen by inspection of Table VI, they are not likely to represent a resource at least as a source of O_2 or H_2O_2 because of the large quantities of soil which must be processed to obtain these compounds. To supply 1 kg O_2 by humidification of Martian soil would require processing 40 m^3 or 65 metric tons of soil (Meyer and McKay 1989).

III. MARS MINERALOGY FROM REMOTELY SENSED SPECTROSCOPIC OBSERVATIONS

Remotely sensed spectroscopic measurements from Earth and Mars orbit have provided direct information about the mineralogical nature of the Martian surface materials. Prior to 1986 the most abundant spectral information was obtained from Earth-based telescopes and the Viking Orbiters and Lan-

ders in the wavelength region from 0.35 to 2.55 μm where reflected solar energy is dominant. A smaller body of data existed at longer wavelengths, to 50 μm, again from a combination of Earth-based, airborne and spacecraft observations. At Mars' orbital distance, radiation observed beyond about 5 μm is dominated by thermal emission. Detailed discussions of these observations and their interpretations are provided in Roush et al. (1991*a*), Singer (1985) and Singer et al. (1979). Here we review the observations from the perspective of their implications for the mineralogy of the Martian soil.

A. General Characteristics

To first order, the surface albedo of Mars falls into two main categories throughout the visible and mid-infrared (0.4–4.0 μm) (Blaney 1991; Bell and Crisp 1991): bright and dark regions. This is true on a variety of scales, from the continental dimensions as viewed from Earth to individual bright drifts and dark rocks seen at both Viking Lander sites. Earth-based spectra of typical high- and low-albedo regions are shown in Fig. 3 (McCord and Westphal 1971). The absolute brightness difference observed between the bright and dark regions (on a scale of hundreds of kilometers) is between a factor of 2 and 3 near 1 μm, depending on the homogeneity of the regions and the seasonally variable level of dustiness in the atmosphere and on the surface.

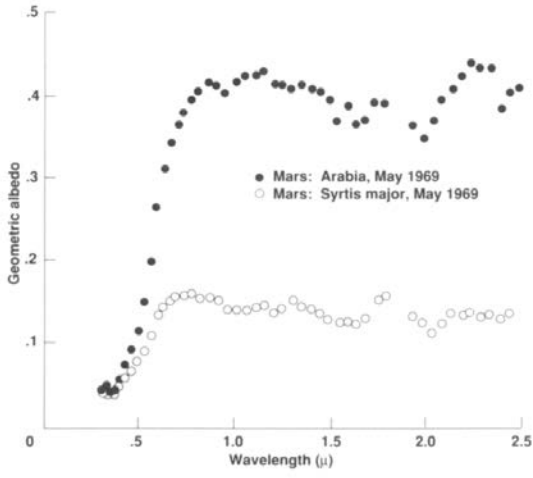


Figure 3. Telescopic spectra of Mars illustrating the absolute brightness differences between typical bright (Arabia) and dark (Syrtis Major) regions (figure from McCord and Westphal 1971).

Telescopic spectra in the visible and near-infrared (0.4 to 2.5 μm) of optically thick dust clouds (McCord et al. 1977) are very similar to spectra of the highest-albedo bright region (McCord et al. 1977; McCord and Westphal

1971). There is good evidence that the bright surface soils on Mars are derived from the same materials as the aerosol dust and as such represent a rather homogeneous, wide-spread unit across the surface of Mars (see, e.g., McCord et al. 1977; Guinness et al. 1979; Arvidson et al. 1989*b*). A comparison of Earth-based bright and dark region spectra of Mars with three spectra of different units observed by the Viking Landers is provided in Fig. 4, which suggests that the soils analyzed by the Viking experiments are related to the classical telescopic bright regions. Thus, interpretation of the bright region spectra is directly applicable to understanding the mineralogical nature of the globally distributed soil unit.

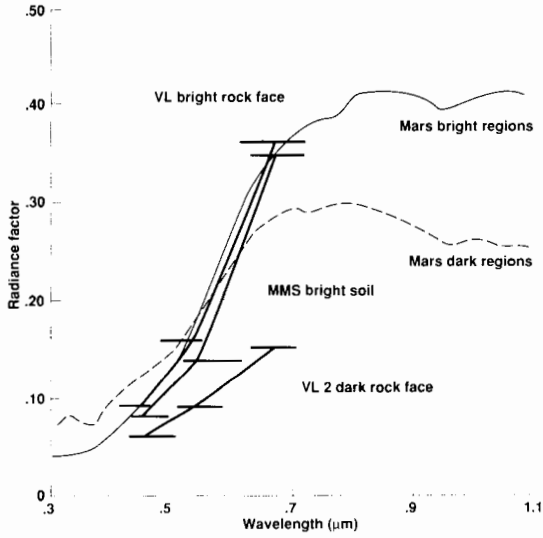


Figure 4. Comparison of spectra obtained by the Viking Lander broad bandpass camera (wide lines and error bars) of various soils and rocks on Mars to telescopic spectra of classical bright and dark regions of Mars (figure from Arvidson et al. 1989*b*). The telescopically observed bright regions are very similar to the bright rocks and soils observed by the Viking Lander.

Figure 5a shows composite spectra of Martian bright and dark regions, relative to a solar analog star and scaled to a value of 1.0 at 1.02 μm to allow easier comparison of spectral features. Both bright- and dark-region spectra are characterized by an intense but relatively featureless reflectance increase from the ultraviolet to a spectral peak near 0.75 μm, although there is a distinct difference in slope between the two regions. Additionally, bright region spectra exhibit subtle slope changes near 0.5 and 0.6 μm, and there is a shallow absorption band centered near 0.86 to 0.87 μm all of which are attributed to Fe^{3+}. Between 0.87 and 1.3 μm, the bright region curve slopes upward slightly, whereas the dark-region spectra exhibit diagnostic Fe^{2+} absorptions in the 1 μm region due to ferromagnesian silicates. Both spectra exhibit weak,

narrow absorptions at 1.45 and 1.62 μm, and a stronger composite absorption from about 1.9 to 2.1 μm all of which are caused by Martian atmospheric CO_2. Throughout the near-infrared, dark region spectra have a distinctive negative slope, which has been suggested to arise from thin coatings (or mixtures) of bright, oxidized material on darker, relatively unoxidized mafic rocks (Singer 1980; Roush 1982; Singer and Roush 1983; Morris and Lauer 1990). In contrast, bright region spectra are relatively flat throughout the near infrared. Spectra of both regions exhibit some weak structure near 2.3 to 2.4 μm which has been suggested to arise from a combination of atmospheric (Encrenaz and Lellouch 1990) and surface sources (Clark et al. 1990).

In the mid-infrared (2.5–5 μm) spectra of Mars exhibit a broad absorption near 3 μm and an increasing level of flux at the longer wavelengths, due to the increased contribution of thermal emission from the surface (Sinton 1967; Moroz 1964; Houck et al. 1973; Roush et al. 1988; Blaney and McCord 1990*a*, *b*). Superimposed upon this flux increase is a subtle absorption feature near 3.8 μm which has been suggested to arise from carbonates (Blaney and McCord 1990*b*). Between 4.2 and 4.4 μm both terrestrial and Martian atmospheric CO_2 prevent observations. An additional feature near 4.5 μm has been interpreted to represent absorptions by sulfates (Blaney and McCord 1990*a*).

In the far infrared (5–50 μm) thermal emission from the surface, and modulation of this energy by atmospheric aerosols dominate the spectra of Mars. Several Martian atmospheric CO_2 features are prominent and have been used to derive atmospheric profiles (see, e.g., Hanel et al. 1972; Conrath et al. 1973). Additionally, two broad featureless absorption bands are present near 10 and 20 μm which are indicative of silicate minerals (Hanel et al. 1972). In recent airborne observations, subtle features are identified near 6, 6.7 and 8.7 μm which have been associated with the presence of hydrates, carbonates and sulfates, respectively (Roush et al. 1989; Pollack et al. 1990).

B. Mineralogical Interpretation of Soils and Dust

Ferric Iron. Because of the visual color of Mars, ferric oxides or other hosts of Fe^{3+} have long been suggested as surface materials, especially for the bright regions. Detailed discussions of ferric oxide spectral reflectance and application to Mars are provided by Singer (1982), Sherman et al. (1982), and Morris et al. (1985,1989,1990). Comparing the Mars spectra with those for candidate Fe^{3+} minerals (Fig. 5 a and b), it appears that the intense slope observed in the visible for Mars demonstrates little of the crystal-field band detail seen for normal crystalline ferric oxides. Even minor amounts ($\sim$1 wt%) of these ferric oxides produce recognizable bands in mixtures diluted with spectroscopically neutral substances (Singer 1981). By contrast, Morris et al. (1989) have demonstrated that spectra of superparamagnetic (nanophase or nanocrystalline) hematite (particle diameters $\leq$10 nanometers) exhibit relatively featureless, intense slopes in the visible similar to that seen for Mars.

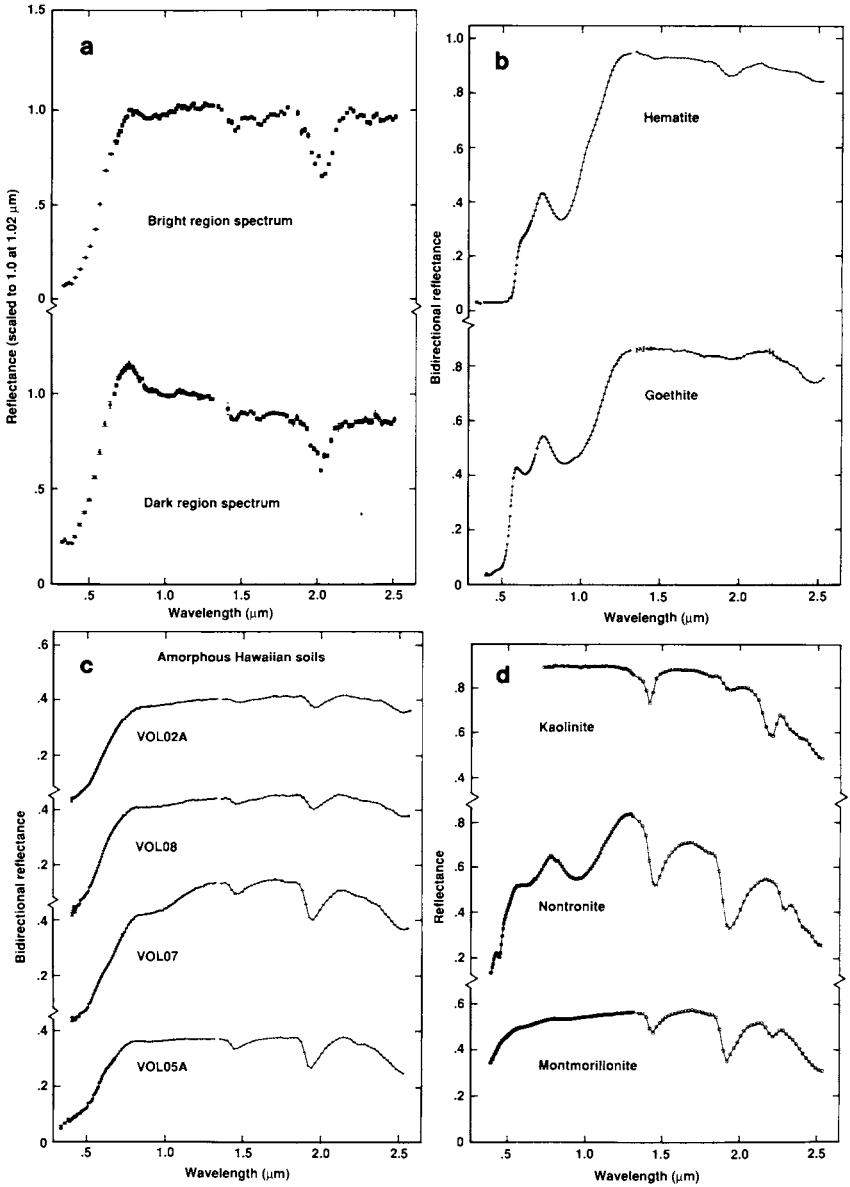


Figure 5. (a) Telescopic relative reflectance spectra of Mars bright and dark regions, scaled to a value of 1.0 at 1.02 mm. (b) Laboratory bidirectional reflectance spectra of the ferric oxides, hematite, αFe_2O_3 (top), and goethite, $\alpha FeOOH$ (bottom). (c) Laboratory bidirectional reflectance spectra of amorphous Hawaiian soils, palagonites. (d) Laboratory reflectance spectra of kaolinite (top), nontronite (middle), and montmorillonite (bottom) (figure from Singer 1985).

Another suggestion, resulting from the Viking Lander chemical measurements, was that the smectite clay nontronite is a major iron-bearing component of Martian soils. However, nontronite spectra (Fig. 5d) tend to have distinctive Fe^{3+} crystal-field absorptions in the visible which are analogous to those seen for the ferric oxides (Singer 1982). Thus, from spectral considerations, nontronite cannot be a major phase on the surface of Mars. Banin et al. (1988) have produced several montmorillonite clays containing Fe^{3+} located in the cation exchange sites. In the visible, their spectra exhibit a relatively smooth absorption edge with little evidence of Fe^{3+} features. Also, as noted above, these samples reproduce many features of the observed behavior of the Viking labeled-release experiment (Banin et al. 1988). However, the near-infrared spectra of these samples exhibit Al-OH absorptions near 2.2 μm which are not consistent with the measured spectra of Mars (Figs. 5a and 5d).

One class of terrestrial materials which contain ferric iron and exhibit a range of crystallinity and composition are palagonites. Palagonite is a broad geologic term, but used here specifically to mean a poorly crystalline hydrated ferric-iron-silica gel, formed by the low-temperature alteration of mafic volcanic glass (ash). There have been numerous suggestions for palagonites or altered glass on Mars (Soderblom and Wenner 1978; Toulmin et al. 1977; Gooding and Keil 1978; Evans and Adams 1979,1980; Allen et al. 1981; Singer 1982). Spectra of four natural palagonitic soils from Hawaii (Singer 1982) are presented in Fig. 5c which show the intense but relatively featureless absorption edges in the visible that are qualitatively and quantitatively very similar to those observed for Mars (Singer 1982; Evans and Adams 1980). Morris et al. (1990) have found that nanophase hematite is an important coloring agent in at least some of these Hawaiian samples.

In the mid-infrared, iron oxide/hydroxide minerals would be expected to exhibit spectral features in the 2.7 to 3.2 μm region, but both Martian and terrestrial atmospheric absorptions prohibit measurements throughout most of this range. In the far-infrared, ferric oxide/hydroxide minerals with well-defined crystalline structure typically exhibit multiple absorptions near 20 μm. Mariner 9 spectra of the globally distributed dust generally exhibit broad single absorption minima near 10 and 20 μm which is inconsistent with the highly crystalline ferric oxides/hydroxides. It is currently unknown whether the spectral properties of nanophase ferric oxides are consistent with the far-infrared Martian data.

From the above discussion, the global dust and majority of high-albedo soil deposits on Mars appear to consist largely of poorly crystalline or nanophase Fe^{3+}-bearing materials. However, this does not preclude the occurrence of some more coarsely crystalline ferric oxides minerals on Mars. Analysis of both Viking Orbiter and Lander multispectral images provide indications of surface regions which are enriched in coarsely crystalline hematite (Soderblom et al. 1978; McCord et al. 1982; Guinness et al. 1987; Arvidson et al. 1989*b*). Morris et al. (1989) found that a minor addition of coarser-grained ($\geq$10 nm) hematite was necessary to simulate the weak 0.86 μm band ob-

served for some regions on Mars. During 1988, spectroscopic measurements of Mars in the visible and near infrared more clearly showed Fe^{3+} spectral features for some regions and were interpreted to indicate a greater abundance of crystalline ferric oxides than previously thought (Bell et al. 1990) or alternatively, variation in one or more ferric-oxide soil components (Singer et al. 1990). While these recent Mars observations indicate a somewhat greater variety of ferric oxide mineralogies than previously known, they remain consistent with an interpretation suggesting an overall relatively low degree of crystallinity and/or short-range crystalline structure in most of the weathered materials.

Silicates. The spectrophotometric interpretation of silicate mineralogy of Mars falls into two broad classes: (1) primary igneous minerals, e.g., pyroxenes and olivines; and (2) minerals which represent significant alteration of the primary igneous materials, e.g., hydrates and hydroxylates.

Spectral evidence suggests that Fe-rich clinopyroxenes are associated with the dark albedo regions on Mars (see Chapter by Singer and McSween). Additionally, vibrational features identified in the far infrared have been interpreted to imply the presence of both olivine (Huguenin 1987) and plagioclase (Aronson and Emslie 1975) associated with the atmospheric dust component. Thus, if all three of these minerals are present on Mars and have a common source region, then this implies mafic rather than ultramafic or silicic sources for Martian pristine crustal materials. Because plagioclase lacks distinctive absorption features in the visible and near-infrared, it is more readily detected in the far infrared. Hence, the planned thermal emission spectrometers to be flown on future missions to Mars (e.g., Mars Observer) should provide additional information regarding the associations of plagioclase and the other igneous silicates.

Bound and Structural Water: Clay Minerals. As common terrestrial rock weathering products, clay minerals have long been considered likely soil constituents on Mars. Fundamental vibrational modes of H_2O and OH^-, associated with clays and other hydrated and/or hydroxylated minerals, occur in the mid to far infrared (see, e.g., Stubican and Roy 1961; Farmer and Russell 1964; Hunt and Salisbury 1970; Farmer 1974; Hunt 1977). Thus, all features seen in the visible and near-infrared region of the spectrum arise from overtone or combination modes of the fundamentals, and as a result are significantly weaker.

Spectra of three representative clay minerals are shown in Fig. 5d. Kaolinite (top), an aluminous clay with little or no interlayer water, has OH spectral bands centered near 1.4 and 2.2 μm. Nontronite (middle), the iron-bearing smectite clay discussed above, displays spectral absorptions near 1.4 and 1.9 μm, due primarily to molecular water occupying the cation exchange sites. The overall negative slope in the near infrared, due to the wing of the fundamentals located near 3 μm, is quite pronounced. The weaker band centered at 2.3 μm is due to structural Fe-OH and is directly analogous to the 2.2 μm features seen in kaolinite and montmorillonite. Mg-bearing clay

minerals have similar bands centered near 2.3–2.35 μm. In montmorillonite (bottom), an Al-bearing smectite clay similar to the nontronite, but with Al^{3+} rather than Fe^{3+} as the cation, the Al-OH combination band is centered just longward of 2.2 μm, nearly the same as for kaolinite. Comparing Figs. 5a and 5d, it is apparent that the Mars data do not exhibit any strong features in the 2.2 to 2.4 μm region and hence are inconsistent with clays as major components on the surface.

The spectra of the iron-doped montmorillonites produced by Banin et al. (1988), discussed above in the context of its possible role in producing the chemical activity seen by the Viking Lander labeled-release experiment, all exhibit a 2.2 μm absorption feature which is inconsistent with the observed spectra of Mars. However, recent studies involving the mixtures of iron-enriched montmorillonites and palagonites indicate that perhaps up to 15 wt% of such clays might remain undetected in the spectra of Mars (Orenberg and Handy 1992).

Earth-based and spacecraft observations of features in the 2.3 μm region were originally interpreted as being produced by clay minerals having Mg-OH bonds while the Al-OH-bond minerals were absent (Singer et al. 1985; Martin and Kieffer 1985; Martin 1985). Alternatively, Clark et al. (1990) have suggested that the 2.36 μm feature is not due to hydroxylated minerals at all but rather to the mineral scapolite. However, scapolites commonly exhibit additional absorption features in the 3.8 to 4.0 μm region (Clark et al. 1990; Swayze and Clark 1990) which are not observed in the Earth-based telescopic spectra of Mars (see, e.g., Roush et al. 1988; Blaney and McCord 1990*a*). Additionally, interpretations at both wavelengths are complicated by spectral interference from Mars atmospheric CO. But, even with this complication, it appears that some solid mineral phase on Mars is contributing to absorptions in the 2.2 to 2.4 μm region. After considering a variety of individual minerals, Clark et al. (1990) conclude that only scapolites, containing bicarbonate and bisulfate anions within their mineral structure, can reproduce the details observed for the absorptions identified in the Mars spectra. They did not, however, consider mixtures of sodium bicarbonate and sodium bisulfate, materials which exhibit a variety of features in this spectral region. Based on the pure mineral accounting for the 2.36 μm absorption feature, Clark et al. (1990) propose that scapolite constitutes 2 to 5 wt% of the soil at a variety of locations. Even if scapolite is a significant component of Martian soil, it does not represent a resource potential in terms of providing a readily available source of volatile compounds.

The presence of bound and/or structural water in Martian soil has been inferred from spectroscopic evidence. Detection of strong absorption in the 3 μm region from Earth (Sinton 1967; Beer et al. 1971; Houck et al. 1973; Singer et al. 1986; Roush et al. 1988; Blaney and McCord 1990*a*, *b*) and early Mariner spacecraft (Herr and Pimentel 1969; Pimentel et al. 1974) indicated water and/or OH^- in some form. Modeling of the 3 μm feature indicated that the water content was 1 to 3 wt% (Houck et al. 1973). The absence of a

reflectance drop-off from ~2.2 to 2.55 μm in these data has been interpreted to suggest that the soils are desiccated, when compared with their terrestrial counterparts (Singer et al. 1985; Roush et al. 1988). Identification of a spectral feature near 6 μm also indicates that molecular water is associated with the surface materials (Pollack et al. 1990).

Carbonates, Sulfates and Nitrates. A number of nonsilicate minerals, most notably carbonates, sulfates and nitrates, have been suggested as likely secondary minerals in Martian soil constituents (Toulmin et al. 1977; Clark et al. 1982). The diagnostic spectral features for these minerals are vibrational in origin with the wavelengths of the fundamental modes occurring in the far infrared. These phases do have weaker overtone and/or combination bands at shorter wavelengths, but observations of the fundamental region provides a more sensitive means for detection of these minerals as minor soil components.

In the mid infrared, the recent Earth-based spectral data have been used to place upper limits of about 1 wt% on the abundance of carbonates (Blaney and McCord 1990*b*) and 3 wt% on the abundance of sulfates (Blaney and McCord 1990*a*) in the near surface environment on Mars. While these data have not been evaluated in these terms, nitrates exhibit a strong spectral signature near 3.6 μm which is not evident in these data. Thus, while we can not currently place upper limits on the abundance of nitrates on the Martian surface, we can state that nitrates do not appear to be a major phase in the near-surface environment on Mars.

In the far infrared, recent airborne spectra exhibit several absorption bands consistent with the interpretation that CO_3^{2-}, HCO_3^{2-}, SO_4^{2-} and HSO_4^{2-}-bearing anionic complexes are located within the dust suspended in the Martian atmosphere (Pollack et al. 1990). Theoretical modeling of the far-infrared data indicate that the abundance of carbonate and sulfate-bearing compounds in the atmospheric dust is roughly 0 to 3% and 10 to 15% by volume, respectively. Thus, this derived amount of carbonate-bearing material, based solely on the far-infrared data, is consistent with abundance estimates based on visible-near- and mid-infrared spectral observations. However, the specifics regarding the chemistry and mineralogy of these materials await interpretation of observations obtained at higher spectral resolution during the 1990 opposition (Roush et al. 1991*b*,1992).

IV. SOIL COMPOSITION FROM SNC METEORITES

Four shergottite, three nakhlite, and one chassignite (SNC) meteorites comprise a clan of igneous rocks for which origin on Mars has been suggested by evidence that has accumulated since the mid-1970s but especially since 1983. The nature and compositions of the SNCs have been reviewed in detail by McSween (1985) and Laul (1986) and will be repeated here only as needed for completeness. There is both direct and indirect evidence that SNCs are from Mars. Only one SNC contains direct evidence with which a measured fingerprint can be unambiguously matched with a measured fingerprint for Mars.

Namely, one of the shergottites (Elephant Moraine, Antarctica, A79001 or EETA79001) contains trapped gases that chemically and isotopically resemble those in the Martian atmosphere (Bogard et al. 1984; Becker and Pepin 1984; Swindle et al. 1984). The other items of evidence are only indirect because, even though they are consistent with Martian origin, the corresponding fingerprints for Mars are either less unique or have not been adequately measured. The indirect evidence includes: (1) young radiometric ages (160–200 Myr for shergottites; 1300 Myr for Nakhla) that distinguish them from all other meteorites (which have ages >4500 Myr); (2) cogenetic grouping on a three-isotope oxygen diagram ($\delta^{17}O$ vs $\delta^{18}O$), showing that they most likely formed on the same planet by geochemical processes resembling those on Earth; (3) traces of carbonate and sulfate minerals, and related phases, that imply origin in an oxidizing environment where water was present.

Because the eight SNCs are genetically related, as inferred with confidence from oxygen isotope compositions, we conclude that all eight are from Mars on the strength of the trapped gas results for EETA79001. If one accepts a Mars origin for SNCs, then their compositions are potentially the most powerful current data regarding resource potentials of igneous Martian materials.

A. Comparison of SNC and Mars Soil Composition

As mentioned above, bulk elemental compositions of shergottites provide reasonable, if not unique, correspondence with the sulfur- and chlorine-free versions of the Viking Lander XRFS analyses. Accordingly, various authors have combined Viking XRFS and shergottite analyses, through interpolation and extrapolation, to build geochemical models for Martian surface materials. From the perspective of Mars resource utilization, scientific details of these models are less important than the more utilitarian use of SNC meteorite compositions as a general guide to Martian rock compositions. Toward that end, Table VII offers a summary of elemental concentrations in the intensively analyzed Shergotty meteorite (the type specimen for shergottites) and in surface sediments at the Viking Lander sites, respectively. Table VII shows that there is currently no evidence in Martian materials for high concentrations of elements that are considered rare or precious on Earth. Unlike lunar rocks, shergottites do not contain high Ti or Al and are distinguished mostly by their high Fe concentrations. Indeed, from the perspective of terrestrial economic geology, shergottite compositions are unimpressive. Although we might expect that the complex geology of Mars could have produced ore-quality materials in some places (Cordell 1985), no information about this can be deduced from SNC meteorite properties.

For resource utilization, of course, one must pay attention not only to the concentration of an element but to its distribution among various minerals. When such distributions are considered, Martian soils are expected to be significantly different "ores" than are rock if, as on Earth, chemical weathering on Mars has consumed primary igneous minerals to form secondary minerals.

TABLE VII

Elemental Compositions of the Shergotty Meteorite and Surface Sediments at Chryse Planitia, Mars (Viking Lander 1) in %, ppm and ppb by Weight

Atomic No.	Shergotty[a]	Mars[b]	Atomic No.	Shergotty[a]
2 He	0.19–0.21 ppb[c]		42 Mo	0.37 ppm
3 Li	3.3–5.6 ppm		47 Ag	6.8–110 ppb
6 C	430–620 ppm		48 Cd	0.014–0.34 ppm
	44–210 ppm[d]		49 In	0.023–0.026 ppm
7 N	132–794 ppb[e]		51 Sb	<5 to 20 ppb
8 O	(40.7–43.6%)[h]		52 Te	3.2–19 ppb
9 F	41–42 ppm		53 I	0.036–0.050 ppm
10 Ne	0.015–0.017 ppb[c]		55 Cs	0.36–0.48 ppm
11 Na	0.95–1.09%		56 Ba	27–40 ppm
12 Mg	5.40–5.7%	3.6%	57 La	1.50–2.44 ppm
13 Al	3.60–4.02%	3.9%	58 Ce	3.51–6.4 ppm
14 Si	23.1–24.0%	21%	59 Pr	0.70–0,88 ppm
15 P	0.24–0.35%		60 Nd	2.60–4.7 ppm
16 S	0.13–0.16%	2.7%	62 Sm	1.01–1.89 ppm
17 Cl	108 ppm	0.8%	63 Eu	0.43–0.65 ppm
18 Ar	3.3–9.9 ppb[c]		64 Gd	1.64–2.8 ppm
19 K	0.12–0.16 %	< 0.4%	65 Tb	0.41–0.52 ppm
20 Ca	6.80–7.15%	4.1%	66 Dy	2.16–4.8 ppm
21 Sc	52–59 ppm		67 Ho	0.56–0.86 ppm
22 Ti	0.4–0.5%	0.37%	69 Tm	0.30–0.38 ppm
23 V	260–265 ppm		70 Yb	1.19–1.80 ppm
24 Cr	0.12–0.16%		71 Lu	0.18–0.26 ppm
25 Mn	0.40–0.42%		72 Hf	1.50–2.23 ppm
26 Fe	15.1–15.6%	12.2%	73 Ta	0.18–0.29 ppm
27 Co	27.2–45 ppm		74 W	0.4–0.5 ppm
28 Ni	56–88 ppm		77 Ir	<5 ppm
29 Cu	26–54 ppm		79 Au	0.81–16 ppb
30 Zn	62–83 ppm		81 Tl	0.15–14.0 ppb
31 Ga	15–17.6 ppm		82 Pb	94 ppb[f]
33 As	0.025 ppm		83 Bi	0.47–2.4 ppb
34 Se	0.29–0.47 ppm		90 Th	0.25–0.39 ppm
35 Br	0.60–0.89 ppm		92 U	0.055–0.17 ppm
37 Rb	4.5–7.27 ppm	<30 ppm[g]		
38 Sr	45–51 ppm	60 ppm[g]		
39 Y	—	70 ppm[g]		
40 Zr	50–67 ppm	<30 ppm[g]		

[a] Laul et al. (1986), except where noted. [b] Average "deep" sample; Clark et al. 1982. [c] All isotopes; Becker and Pepin 1984. [d] Excluding C extracted at <600°C; Wright et al. 1986. [e] Excluding gas extracted at <600°C; Becker and Pepin 1984. [f] Sample 3A; Chen and Wasserburg 1986. [g] Clark et al. 1976. [h] By difference from sum of major elements.

In general, it is likely that processes of differential weathering, soil, and sediment formation on Mars have geochemically fractionated elements from their original sites in igneous minerals and combined them into completely different forms, some of which might offer more encouraging potentials for resource utilization. Clearly, the high sulfur and chlorine concentrations in Martian surface sediments have no counterparts among SNC meteorite compositions. To reconcile such concentrations with shergottite-type rocks, enrichment factors of ~20 to 100 are required for the soil relative to the igneous rocks. The strongly magnetic component of the sediments might represent residual (unweathered) titanomagnetite inherited from SNC-like rocks, representing a "naturally beneficiated" Ti ore; alternatively, it might be an authigenic soil mineral completely unrelated to SNCs and with little or no Ti. Resolving these major questions that affect resource utilization requires that the compositions of soils and sediments be known to at least the same level of detail as the compositions of rocks.

B. Martian Mineralogy Inferred from SNCs

The SNC meteorites consist of minerals expected to crystallize from basic magmas, with some modifications attributable to later shock metamorphism (Table VIII). Dominant minerals are augite, pigeonite and olivine; plagioclase is a minor mineral which, in the shergottites, has been transformed to non-crystalline "maskelynite" by shock pressure, probably during impact cratering on the SNC parent planet. In the eight known SNC specimens, there is no evidence for lunar-like concentrations of ilmenite—the mineral that has been advocated as a possible oxygen source on the Moon. Likewise, magmatic sulfides in SNCs are rare and contain mostly Fe, as opposed to Ni, Cu, Zn or Ag. In fact, by economic geology standards on Earth, these rocks are "barren." Therefore, based on the SNC model, the resource value of Martian rocks should be sought in areas other than recovery of precious metals.

Although SNC meteorites are igneous rocks, several of them contain traces of volatile-bearing minerals that might represent a link to Martian soils or sediments. These minerals occur as rare, small grains (most $<20\,\mu$m size) that were first discovered in 1986 and which are still being identified; the current known inventory is summarized in Table IX and arguments for their pre-terrestrial origins have been given by Gooding et al. (1988,1990,1991). The occurrence of sulfur- and chlorine-rich minerals is consistent with elemental compositions of Martian soils (Table VII), as is the occurrence of possible Fe-rich smectites which have been inferred from Viking XRFS and biology results. One or more ferric oxides or hydroxides co-exist with Fe-rich smectite clay in Nakhla. One should keep in mind, however, that the shergottites, in particular, have been strongly modified by shock metamorphism that is believed to have differentially eradicated (through melting and glass formation) most of the secondary silicates and salts. Therefore, the SNCs might actually under-represent the true variety of Martian clays and salts.

Several of the salt minerals in the SNCs appear to be hydrated (or were

TABLE VIII

Volume Percent Abundances of Igneous Minerals in SNC Meteorites[a]

	Shergotty	Zagami	EETA79001 Lith. A	EETA79001 Lith. B	ALHA[b] 77005	Nakhla	Chassigny
Pigeonite (Mg,Fe,Ca) $(Mg,Fe)Si_2O_6$	36.3	36.5	59.3	39.5	26	nr	nr
Augite (Ca,Na) (Mg,Fe,Al,Ti) $(Si,Al)_2Si_2O_6$	33.5	36.5	6.1	20.0	11	78.6	3.8
Orthopyroxene $(Mg,Fe)_2Si_2O_6$			5.4				4.0
Olivine $(Mg,Fe)_2SiO_4$			8.9		52	15.5	88.5
Plagioclase or Maskelynite (Ca,Na)Al $(Al,Si)Si_2O_8$	23.3	21.7	17.1	29.1	10	3.7	2.6

Oxides	2.3	2.1	3.0	3.5	1	1.9	1.2
Titanomagnetite $(Fe,Ti)_3O_4$	x	x	x	x		x	
Ilmenite $FeTiO_3$	x	x	x	x	x		x
Chromite $FeCr_2O_4$			x		x		x
Sulfides	tr	tr	tr	tr	tr	tr	tr
Troilite FeS					x		
Pyrrhotite Fe_{1-x^s}	x	x	x	x			

TABLE VIII (cont.)

Volume Percent Abundances of Igneous Minerals in SNC Meteorites[a]

	Shergotty	Zagami	EETA79001 Lith. A	EETA79001 Lith. B	ALHA[b] 77005	Nakhla	Chassigny
Pentlandite $(Fe,Ni)_9S_8$							x
Marcasite FeS_2							x
Phosphates	tr	tr	0.2	0.4	tr	tr	tr
Whitlockite $Ca_9(Mg,Fe)H(PO_4)_7$	x	x	x	x	x		
Chlorapatite $Ca_5(PO_4)_3Cl$	x	x	x	x		x	x
Mesostasis (incl. glass; fine-grained Fe-rich olivine; tridymite, SiO_2; baddelyite, ZrO_2)	4.0	2.1	0.1	0.7	tr	nr	nr

[a] Table modified from McSween 1985; nr = not reported; tr = trace (<1% but not measured); x = present but not distinguished as a separate sub-total.

[b] Two other SNCs, Lafayette and Governor Valadares, closely resemble Nakhla but not abundance data are available.

TABLE IX
Secondary Minerals (Possible Weathering Products or Soil/Sediment Minerals) Found at Trace Abundances in SNC Meteorites

Mineral[a]	Shergottite EETA79001	Nakhla	Chassigny
$CaCO_3$	X	X	X
Mg-bearing $CaCO_3$	X		
$MgCO_3$			X
$(Fe,Mn)CO_3$		X	
$CaSO_4 \cdot nH_2O$	X	X	X
$(Mg_x(PO_4)_y \cdot nH_2O$	X		
$(Mg_x(SO_4)_x \cdot nH_2O$		X	
(Na,K)Cl		X	
"Illite" $(K,Na,Ca_{0.5},H_3O)(Al,Mg,Fe)(Si,Al)_4O_{10}[(OH)_2,H_2O]$	X		
S,Cl-bearing "micabole"	X		
Smectite $(Na,Ca_{0.5})_{0.3}(Al,Mg,Fe)_2(Si,Al)_4O_{10}(OH)_2 \cdot nH_2O$		X	
$Fe_2O_3 \cdot nH_2O$		X	

[a] x, y and n are generalized coefficients where stoichiometry is variable or uncertain; X = present but not distinguished as a separate sub-total.

originally hydrated on the SNC parent planet), sustaining the belief that hydrated minerals in Martian soils and sediments might represent a significant and readily accessible source of water. In particular, Ca-sulfate represents a possibly important common denominator in all SNC subgroups, suggesting that gypsum might be one of the more important sinks for water that should be considered.

V. METEORITIC COMPONENT OF SOILS

Meteoritic material continuously rains onto the surface of Mars, providing a globally distributed resource. On the Moon the meteoritic material constitutes about 2% of the mare soils (Anders et al. 1973) and up to 4% in the Apollo 14 soils (Wasson et al. 1975), and similar concentrations might be expected on Mars.

Flynn and McKay (1988,1990) have assessed the meteoritic infall at Mars using measurements of the meteoritic flux at Earth and estimates of the Mars/Earth flux ratio. They arrived at estimates of the meteoritic infall rate of between 2700 and 59,000 ton yr^{-1} onto the surface of Mars. This corresponds to the accretion of a layer of meteoritic material 2 cm to 40 cm deep every Gyr.

The meteoritic material in the lunar soils was determined to be similar in composition to that of the carbonaceous chondrite meteorites (Anders et al. 1973). Thus the meteoritic component in the soils of Mars should enrich the soil in elements rich in the carbonaceous chondrite meteorites but depleted in basalts. This should result in soil enrichments in the siderophile elements, including Ni, Cr and Ge, and volatile elements, including C, Zn and Se, over the concentrations expected in soils derived exclusively from basalts. Nickel, for example, is present in chondritic material at 10 to 50 times its typical basaltic concentration.

The expected meteoritic accretion rates for several elements are given in Table X along with the amount of each meteoritic element accreted per square meter of surface area per billion years. Flynn and McKay (1990) have calculated that, because of the expected size and velocity distributions, more than half of the meteoritic mass arriving at Mars will survive atmospheric entry without melting. This material will consist mostly of particles in the size range from 60 to 1200 μm in diameter. The rate at which indigenous soil is produced on Mars by erosion and impact cratering is estimated to be of order meters per billion years (Flynn and McKay 1990). For an indigenous soil production rate of 1 m Gyr^{-1}, the meteoritic infall would result in meteoritic concentrations ranging from 2 to 29% in the Martian soils. The elemental concentrations produced under these conditions are given in the last column of Table X.

Elemental abundance data for the Martian soils from the Viking and Phobos spacecraft are not inconsistent with a large meteoritic concentration, but no element has yet been measured which unambiguously determines that meteoritic concentration. Limits on the Ni concentration, an element present in high abundance in chondritic material but low in basalts, including Mars' meteorite Shergotty, should allow determination of the meteoritic concentration.

Boslough (1988) has proposed a second source of meteoritic material in the Martian soils: weathered material derived from shock-activated meteoritic projectiles from the era of heavy bombardment. In addition, concentrations

TABLE X
Meteoritic Component in the Soil of Mars

Element	Infall Rate[a] (ton/year)	Accretion[a,b] (g m^{-2} Gyr^{-1})	Estimated Concentration[b] (weight percent)
C[c]	100–2200	740–15,000	0.07–1.0
Ni	27–590	200–4000	0.02–0.3
S	170–3700	1100–25,000	0.1–1.8
Cr	7–150	50–1000	0.005–0.08
Zn	0.8–18	6–120	0.0006–0.009
Cu	0.3–7	2–50	0.0002–0.004
Ge	0.1–2	0.6–12	0.00006–0.0009
Se	0.05–1	0.4–8	0.00004–0.0006
Mo	0.003–0.06	0.02–0.4	0.000002–0.000029

[a] Calculated assuming the meteoritic component has the CI carbonaceous chondrite composition from Anders and Grevesse (1989).
[b] From meteoritic material only, calculated assuming an indigenous soil production rate of 1 m Gyr^{-1}.
[c] CI carbon value is the orgueil meteorite value from Anders and Grevesse (1989) as no average CI concentration is tabulated by them.

of meteoritic material would be expected in layers of debris from major meteoritic impacts, as seen on Earth at the Cretaceous-Tertiary boundary (Alvarez et al. 1980). Meteoritic material in the soils, as well as surviving fragments of meteorites, may provide a useful resource on Mars.

VI. DIRECT USE OF MARTIAN SOIL MATERIALS

Based on the preceding sections, it is evident that soils having similar physical properties and composition are available globally on Mars. Although the available information about this soil is not as complete as desired, it is still possible to begin designing processes to use materials derived from Martian soils. Using the Martian soil as a resource breaks naturally into two categories: (1) using the soil itself with or without minor processing, and (2) using the soil as an ore from which useful compounds can be extracted. In this section, we examine some possible uses for Martian surface soils directly, with little or no processing.

A. Radiation Protection

One of the long standing concerns about putting people in extraterrestrial environments for long periods of time is the effect that solar and galactic radiation will have on them. Although the thin atmosphere of Mars (20 g cm^{-2} in the zenith direction) provides a significant amount of radiation shielding at the Martian surface (Simonsen et al. 1990), additional radiation shielding of crew habitats will probably be desirable. Also, the occurrence of

occasional high radiation flux levels associated with solar flares will require that crews have access to areas with excess radiation shielding.

One approach to provide extra radiation shielding is to pile loose Martian soil on top of habitat areas. The fact that loose surface soils are widely distributed suggests that conventional techniques can be used to move it. Compared with lunar soils, Martian soils might be even better shields if they contain abundant light elements (such as H, C, N) that are effective at stopping neutrons produced by primary cosmic rays. A layer of Martian soil 1 meter thick would provide adequate radiation shielding for a solar storm shelter (Simonsen et al. 1990).

B. Plant Growth Medium

The properties of Martian soil suggest that it may make a good medium for growing plants. As discussed above, the soil is available planet-wide and has a highly consistent texture and chemical composition. The soil material appears to form a loosely packed porous medium well adapted to mechanically support rooting plants. While spectroscopic evidence appears to rule out clays as a major constituent of the soil, the possibility that the soil may contain some smectite clays (Banin and Rishpon 1979) is relevant for their suitability as a plant growth medium. Smectite clay minerals in the soil will act as an efficient buffering system to stabilize soil pH at the slightly acidic range (pH 5–6) in the absence of carbonate minerals. Smectite minerals will also ensure large reservoir of exchangeable ions due to their high exchange capacity. Fine porosity and excessive swelling characterizing soils rich in smectites may cause low hydraulic conductivity and lead to nonoptimal water/air ratios in the root zone. However, the inferred high content of iron oxyhydroxides in Martian soil may counteract the swelling tendency of smectites, acting as a cementing agent and forming stable microaggregates in the soil.

The productivity of a soil depends on the content of the major nutrient elements and their chemical availability to plants. The information on the abundance in Martian soils of the major nutrient elements needed by plants is shown in Table XI. Of particular interest is the abundance and form of nitrogen compounds. Because the Viking XRFS experiment was not able to detect N, no information is available for nitrogen content in the Martian soil. It may be speculated that some nitrate will be present in the soil as a residue from imported meteoritic organic material after it has been decomposed. Nitrogen is an extremely important soil nutrient for plant growth and is commonly a limiting nutrient for terrestrial biospheres; its presence in terrestrial soils is primarily due to fixation of atmospheric nitrogen by microorganisms, and therefore Martian soil will probably require nitrate fertilizer to enhance its ability to support plants.

The abundance of phosphorus, P, and potassium, K, are also of interest for supporting plants. Phosphorus, along with nitrogen, is one of the most important nutrients that plants need and its abundance in terrestrial soils often limits the soil productivity. The Viking XRFS experiment was not capable of

TABLE XI
Major Nutrient Elements for Plants in Terrestrial and Martian Soils

	Terrestrial Soils[a]		Mars Soils[b]
Element	Concentration Range (ppm)	Selected Average (%)	Selected Average (%)
N	200–4000	0.14	not determined
P	200–5000	0.06	0.30
K	400–30,000	0.83	0.08
Ca	7000–500,000	1.37	4.1
Mg	600–16,000	0.5	3.6
S	30–10,000	0.07	2.9

[a] Adapted from Lindsay (1979).
[b] Following Banin (1989); Banin et al. (1992).

detecting the presence of P, but analysis of the SNC meteorites suggest that it should be present on Mars at greater abundance than in typical terrestrial soils (Table XI). Potassium, however, was not seen by the Viking XRFS experiment because its abundance was beneath the detection threshold of the instrument. This element appears to be at deficiency levels for growing plants and may need to be supplemented in soils. Fertilizers could be produced on Mars from salts found in the soils or in evaporite deposits, or alternatively, crop plants could be selected which were tolerant of Martian soil conditions.

Another major consideration for the suitability of Martian soils in supporting plants is the content of the essential micro-nutrient elements (Table XII). Iron is present at high concentrations in the Martian soil and this along with the noncrystalline forms of the iron-bearing minerals and the slightly acidic reaction of the soil suggests that iron deficiency will not be a limiting factor for plant growth. From similar considerations, manganese will probably not be a limiting factor. Three other essential trace elements (Zn, Cu and Mo) are present in the SNC meteorites at a concentration range similar to that of terrestrial soils, and may not pose problems of either deficiency or toxicity to plants.

Potential interfering elements may be present in the soil. Although aluminum is present at concentration lower than the representative range for terrestrial soils, it may reach toxic levels under the acidic conditions believed to prevail in the Mars soil; under terrestrial conditions aluminum toxicity to agricultural plants is observed in acidic soils. The relatively high concentration of chloride in the Martian soil may impose osmotic stress onto plants and interfere with water availability and uptake. Such problems may be remedied by leaching excess soluble salts from the soil.

C. Making and Using Duricretes

When synthetic Martian soils concocted in the laboratory are thoroughly wetted and then allowed to dry, the high salt content invariably produces an extremely strong material (Boyd et al. 1989). By analogy with the terminology

TABLE XII
Essential Micro-Nutrient Elements for Plants in Terrestrial and Martian Soils

	Terrestrial Soils[a]		Mars Soils[b]	
Element	Concentration Range	Selected Average	SNC Meteorite Range	Shergotty
Fe	7000–550,000	38,000	107,100–210,600	150,000
Mn	20–3000	600	4600–7900	4000
Zn	10–300	50	62–220	62–83
Cu	2–100	30	2.6–<100	26–54
B	2–100	10	not determined	not determined
Mo	0.2–5	2	0.2–0.37	0.37

[a] Adapted from Lindsay (1979).
[b] Following Banin (1989) and Banin et al. (1992).

of duricrust and the resemblance to concrete of the resultant, we term such materials "duricrete."

Boyd et al. (1989) have performed experiments to produce and test duricretes made with simulated Martian materials. Duricretes were made using combinations of water, 12% $MgSO_4$, 1% NaCl, 85% clay (Pennsylvania nontronite or Wyoming bentonite), and 2% Fe_2O_3. Table XIII shows the range of tensile strengths of duricretes made using a variety of techniques. Duricretes can be produced which have strengths comparable to concrete. These experiments also showed that such duricretes will probably crack upon drying, but that with a minimum of fiber or cloth reinforcement, consolidated briquettes or plates of very high tensile strengths can be achieved. Compaction to remove more than half of the water and controlled drying was necessary to prevent cracking when no matrix materials were present.

VII. RESOURCES EXTRACTED FROM SOILS

In this section we consider procedures for extracting useful compounds from Martian soils. In this vein, we concentrate mostly on using the ubiquitous soil as an ore. In the early phase of human activities on Mars, the ability to transport materials from where they occur to where they will be used will be extremely limited. Thus, it will be necessary to use the materials at or near the location of the human base. Table XIV lists a number of useful materials which can be obtained from Martian soils of the the type found at the Viking Lander sites. The following section describes procedures for obtaining many of these compounds.

A. Extracting Bound Water from Minerals

One of the most important resource that humans will need on Mars is water. Water is needed for life support, and will be needed in significant quantities

TABLE XIII
Measured Strengths of Duricretes[a]

Method	Material	Strength (PSI)
Air dry	4-ply nylon mesh	262
	2-ply nylon mesh	146
Compressed, air dry	no matrix	421
	1-ply nylon mesh	270
	2-ply rayon cloth	334
	2-ply nylon cloth	504
	2-ply nylon mesh	584
	1% Kevlar fiber	574
	2% Glass wool	577
Compressed, controlled dry	no matrix	380
Baked	45% sulfur	117
Baked, compressed	45% sulfur	318
Comparison materials	Cement concrete	600
	Brick	800
	Glass, pottery	5000–25,000

[a] Table from Boyd et al. 1989.

on Mars for minerals smelting, materials processing, and manufacturing processes. In addition, hydrogen (derived from water) is a key component in the synthesis of fuels and propellants for powering rover vehicles and rockets for launching off the Martian surface. As discussed above, bound water occurs in the soil although the exact concentration is uncertain. However, Viking results showed that 1% water could be reliably derived from Martian soils by heating to 500°C, a relatively low temperature. This suggests that the soil could be a reliable resource to obtain water on Mars.

For purposes of illustration, Fig. 6 shows a hypothetical system for extracting water from Martian soil. The method used to collect the soil is a track-type caterpillar model loader. The loader fills a dump tank with soil periodically. The dump tank is connected via a conveyer belt to an oven which heats the soil and collects the water. The dry soil is then moved to a slag heap via another conveyer belt.

In order to obtain 1000 kg of water, the system must process 100,000 kg of soil. The vast majority of the energy that must go into this system goes into the actual heating of the soil, 10.30 kW-hr kg^{-1} of water. In addition, a small amount of energy is required to run the bulldozer and the conveyor belts, (0.12 kW-hr kg^{-1}). By way of comparison, Meyer and McKay (1984) describe a conceptual system for extracting water from Martian air using successive cycles of cooling and compression. Here the difficulty is that the mixing ratio of water in Martian air is very small (10 ppm average) and thus a huge volume of air must be processed in order to obtain water. The estimated

TABLE XIV

Materials Available from Martian Soil[a]

Material	Representative Process
H_2O	Heating soil to extract bound water
H_2O_2	Electrolysis of H_2SO_4 solution, vacuum evaporation
O_2	Electrolysis of water, small quantities from humidifying soil
S	From heating of sulfates
Fe	Smelted from iron-oxides in soil
Ti	From titanomagnetite
Al	Molten electrolysis of oxides
Mg	Molten electrolysis of Mg-sulfates
Ceramic	From clays
Glass	From SiO_2
Cement	From silicates, water
Plaster	From gypsum
Brick	From MgO

[a] Table from Meyer and McKay (1989).

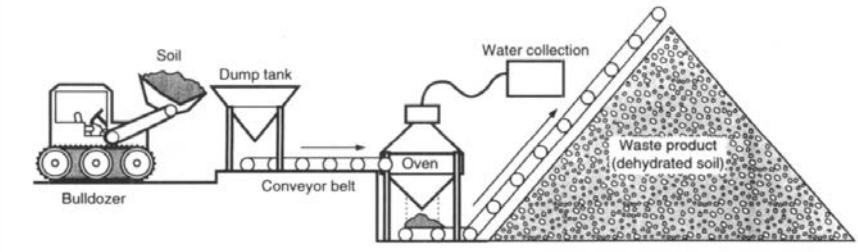


Figure 6. Schematic illustration of a conceptual process to extract bound water from Martian soil.

energy cost to obtain water by cooling and compressing Martian air is 102.8 kW-hr kg^{-1} (Meyer and McKay 1984). Because of the simple technology and the relatively low energy cost, obtaining water from Martian soil is an attractive option.

In addition to the bound water in the ubiquitous Martian soil, water-rich ore deposits may occur on the surface of Mars. Many minerals contain large amounts of bound water that can be extracted by heating. Figure 7 shows the water content of several minerals which have been proposed as candidate components of Martian soil along with the energy required to release the water. Of particular interest is gypsum, which has been detected in all three types of SNC meteorites.

B. Iron Extraction and Smelting

The availability of water and salts on Mars (and with processing, of derived acids and bases) should allow the use of hydrometallurgical treatments which

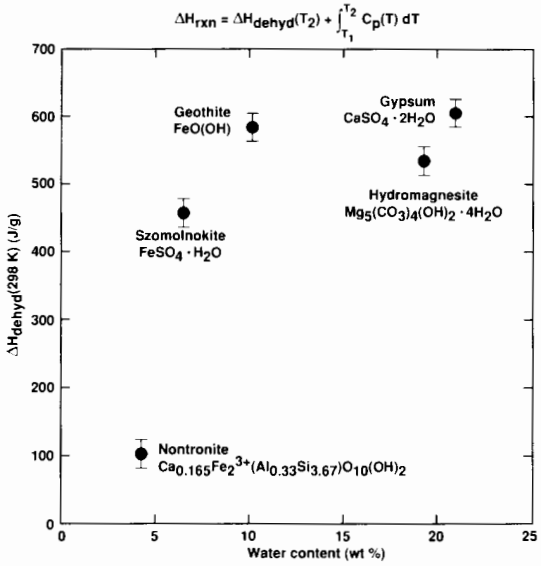


Figure 7. Bound water content of some candidate minerals which may be found on Mars. The heat energy required to extract this water is shown by the equation at the top of the figure.

are similar to those used on Earth (in marked contrast to the situation on the Moon). However, the lack of an oxygen-rich atmosphere and of abundant organic carbon and hydrocarbons will, in detail, necessitate many differences from terrestrial minerals processing. Recycling of reagents (ideally, closed-loop processing) will be necessary owing to the high cost of Mars- or Earth-derived materials, and in order to minimize the environmental impact of surface operations. The major energy sources will be solar, and perhaps nuclear (energy cannot be based on the combustion of organic matter, as on Earth). The major reducing agents for metal production will presumably be hydrogen (H_2), carbon monoxide (CO), and perhaps eventually biologically derived carbon.

The chemical and thermodynamic constraints on metal processing are the same on both Earth and Mars. Geochemical abundances of common elements should be comparable, although conceivable mechanisms for the concentration of desirable elements in ore deposits are much more restricted on Mars (although not nearly so much so as on the Moon). The metals iron (Fe) and copper (Cu) are much easier to reduce from their oxides or other compounds than are other common metals such as aluminum, Al, or magnesium, Mg. Relatively simple chemical methods can therefore be used to recover iron whereas recovery of aluminum and magnesium is typically done by a relatively sophisticated process (molten salt electrolysis) which might also be possible on Mars, given sufficient chlorine and a source of electricity.

Iron oxides can presumably be concentrated from common Martian soils or from local ore deposits, such as weathered sulfide deposits (gossans; see, e.g., Burns 1988) or paler-placer deposits (former alluvial concentrations of dense minerals), should such be found to occur. Iron oxides can presumably be concentrated magnetically, as could unoxidized meteorite fragments (Ni-rich and ready for use) which may occur locally. In theory, aqueous acid leaching of iron-bearing soils followed by neutralization to iron hydroxides and roasting to iron oxide (Fe_2O_3) could be used for soils lacking a separate or crystalline iron oxide phase. Once an iron oxide concentrate is obtained, it can be reduced to native iron by reduction by either hydrogen gas (H_2 obtained by electrolysis of water) or carbon monoxide gas, CO. The latter is attractive because CO is a possible residue of oxygen extraction from the Martian atmospheric CO_2 (Ash et al. 1989) and CO may be produced in large quantities because CO and O_2 can be used as a propellant combination for rockets and other vehicles on Mars (French 1985). For H_2 gas, the equation is:

$$Fe_2O_3 + 3H_2(g) \Longrightarrow 2Fe + 3H_2O(g). \quad (4)$$

(Note that a similar process has been proposed for the reduction of lunar ilmenite, $FeTiO_3$; hematite is, however, easier to reduce than ilmenite.) Using CO gas the corresponding equation is:

$$Fe_2O_3 + 3CO(g) \Longrightarrow 2Fe + 3CO_2(g). \quad (5)$$

On Earth, the reduction is conventionally done using organic carbon (e.g., coal) as the source of CO gas:

$$Fe_2O_3 + 3C \Longrightarrow 2Fe + 3CO(g). \quad (6)$$

In the preindustrial era, charcoal or wood was used and Martian-grown biomass might similarly be used for iron reduction.

In any case, further processing (refining) would be required to make high-strength wrought iron or steel for tools, machinery, vehicles, containers, etc. This involves the introduction of the correct amount of carbon and the addition of alloying elements such as manganese and nickel (see, e.g., Rosenqvist 1983). The ability to smelt and process iron for tools, even in small amounts, could be a valuable asset for a Mars base early on because of the time and transportation costs to obtain any replacement metal part or tool from Earth.

Copper, like iron, can be easily reduced from its oxides. However, unlike Fe, it is not globally distributed in Martian soils. It is geochemically a much rarer element, and it was not detected in the Viking XRFS experiment or in SNC meteorites. Still, it may be locally common in soils formed by the weathering of sulfide concentrations at the bases of lava flows (see, e.g., Burns 1988). The identification of such ores would be valuable, particularly because there are no good substitutes for copper for use in electrical wiring.

C. Sulfur Extraction and Uses

Sulfur is presumably available on the Martian surface in the form of soluble salts and will have a wide variety of uses. One of the most valuable uses for sulfur is in the production of sulfuric acid, H_2SO_4, the most common acid for industrial processes on Earth. One use of this compound will be for metals extraction as mentioned above. Sulfuric acid is typically prepared from sulfur dioxide, SO_2, which could be obtained from the roasting of soil sulfates, for example:

$$CaSO_4 \Longrightarrow CaO + SO_2(g) + 1/2O_2(g). \quad (7)$$

Note that hydrated calcium sulfate (gypsum) is the raw material for plaster, and that it has been identified in the SNC meteorites. Roasting would also yield lime, a good acid-neutralizing base, and an essential ingredient of conventional portland cement. As discussed above, gypsum could also be heated to release up to 20% bound water.

Reactions for $MgSO_4$ would be similar to those given above for $CaSO_4$; MgO would be useful for neutralizing acids. Metallic Mg could be prepared by electrolysis of molten magnesium chloride.

A more common method of sulfuric acid production on Earth is by the roasting and oxidation of metal sulfides such as pyrrhotite, e.g.:

$$2FeS + 7/2O_2(g) \Longrightarrow Fe_2O_3 + 2SO_2(g). \quad (8)$$

Sulfides might be locally present as droplets or masses at the base of lava flows or in other igneous bodies. Other potential uses for sulfur and sulfuric acid include chemical explosives, preparation of phosphate fertilizer (from the mineral apatite), and cleaning of metal fragments.

D. Glass and Ceramics

The probable lack of pure silica sand, or of abundant feldspar, in Martian soils, implies that the preparation of optical-quality glass on Mars will be difficult and expensive. The abundance of iron will make this doubly so; transparent glass must be nearly iron-free. Nevertheless, suitable chemical treatment of Martian soils might allow the preparation of a silica-rich concentrate suitable for glass making, or local silica-rich concentrations might have formed naturally. Even opaque glasses could be used for construction materials and for producing mirrors to reflect light from the surface into buried or underground habitats.

Lower-quality ceramics should be easier to prepare, by baking hydrous materials or by partial fusion of anhydrous materials. Note, however, that kaolinite-rich soils are best for ceramic applications and such soils are probably lacking on Mars. Smectite and palagonite are not known to produce good ceramics because of the expansion properties of smectites.

VIII. SUMMARY AND CONCLUSIONS

We have reviewed the physical and chemical properties of the Martian soil

from the viewpoint of the resources they represent. Martian soil of the type found at the Viking landing sites represents an attractive resource because it appears to be globally distributed with uniform chemical and physical properties. Thus, it is a resource which does not require extensive prospecting. This ubiquitous soil unit is loosely packed, porous and fine-grained. It thus can be easily moved and handled without requiring extensive mining or blasting.

Viking-type soils contain significant amounts of extractable water and metals including iron, magnesium, aluminum and titanium. The soil also contains significant concentrations of sulfate and chloride salts, resources for sulfur and chlorine compounds. Earth-based spectroscopic evidence suggests that iron in the Martian soil is poorly crystalline and the Viking results suggest that, although there is a strongly magnetic phase, it may not be magnetically separable from the bulk of the soil. The best analogs, in terms of spectral properties, to the Martian soil are a class of terrestrial palagonites. In addition, as putative pieces of Mars crustal rocks, the SNC meteorites offer a rich source of materials for study to gain information about the composition of igneous precursors or actual relict grains of Martian soils.

Because there has been no direct determination of the mineralogy of the Martian soil, a great deal of theoretical and experimental work has been done to infer it and a wide range of mineral models have been proposed. However, it must be emphasized that these models remain controversial. Thus, the most crucial information needed to further develop and evaluate concepts for using the Martian soil is to determine directly its mineralogy and this must be a focus of future missions to Mars if they are to be valid precursors to human activities on the surface which rely on the use of local resources.

Relatively little work has been done to develop strategies or processes for using the Martian soil. This chapter has suggested some preliminary strategies and processes for deriving resources from Martian soil, but considerable engineering design work needs to be done to further develop these and other ideas.

Acknowledgments. We thank C. McKay as well as an anonymous reviewer for useful comments and a careful reading of this manuscript.

REFERENCES

Allen, C. C., Gooding, J. L., Jercinovic, M., and Keil, K. 1981. Altered basaltic glass: A terrestrial analog to the soil of Mars. *Icarus* 45:347–369.

Alvarez, L. W., Alvarez, W., Asara, F., and Michel, H. V. 1980 Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 108:1095–1108.

Anders, E., and Grevesse, N. 1989. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* 53:197–214.

Anders, E., Ganapathy, R., Krahenbuhl, U., and Morgan, J. W. 1973. Meteoritic material on the Moon. *The Moon* 8:3–24.

Aronson, J. R., and Emslie, A. G. 1975. Composition of the Martian dust as derived by infrared spectroscopy from Mariner 9. *J. Geophys. Res.* 80:4925–4931.

Arvidson, R. E., Guinness, E. A., Dale-Bannister, M., Adams, J., Smith, M., Christensen, P. R., and Singer, R. 1989*a*. Nature and distribution of surficial deposits in Chryse Planitia and vicinity, Mars. *J. Geophys. Res.* 94:1573–1587.

Arvidson, R. E., Gooding, J. L., and Moore, H. J. 1989*b*. The Martian surface as imaged, sampled, and analyzed by the Viking landers. *Rev. Geophys. Space Phys.* 27:39–60.

Ash, R. L., Werne, J. A., and Haywood, M. B. 1989. Design of a Mars oxygen processor. In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 479–488.

Baird, A. K., Toulmin, P., III, Clark, B. C., Rose, H. J., Jr., Keil, K., Christian, R. P., and Gooding, J. L. 1976. Mineralogic and petrologic implications of Viking geochemical results from Mars: Interim results. *Science* 194:1288–1293.

Baird, A. K., Castro, A. J., Clark, B. C., Toulmin, P., III, Rose, H. J., Jr., Keil, K., and Gooding, J. L. 1977. The Viking X-ray Fluorescence Experiment: Sampling strategies and laboratory simulations. *J. Geophys. Res.* 82:4595–4624.

Ballou, E. V., Wood, P. C., Wydeven, T., Lehwalt, M. E., and Mack, R. E. 1978. Chemical interpretation of Viking Lander 1 life detection experiment. *Nature* 271:644–645.

Banin, A. 1989. Mars soil—A sterile regolith or a medium for plant growth? In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 559–571.

Banin, A., and Margulies, L. 1983. Simulation of Viking biology experiments suggests smectites, not palagonite, as Martian soil analogs. *Nature* 305:523–526.

Banin, A., and Rishpon, J. 1979. Smectite clays in Martian soil: Evidence for their presence and role in Viking biology experimental results. *J. Molec. Evol.* 14:133–152.

Banin, A., Carle, G. C., Cheng, S., Coyne, L. M., Orenberg, J. B., and Scattergood, T. W. 1988. Laboratory investigations of Mars: Chemical and spectroscopic characteristics of a suite of clays as Mars soil analogs. *Origins of Life* 18:239–265.

Banin, A., Clark, B. C., and Wänke, H. 1992. Surface chemistry and mineralogy. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 594–625.

Becker, R. H., and Pepin, R. O. 1984. The case for a Martian origin of the shergottites: Nitrogen and noble gases in EETA 79001. *Earth Planet. Sci. Lett.* 69:225–242.

Beer, R., Norton, R. H., and Martonchik, J. V. 1971. Astronomical infrared spectroscopy with a Connes-type interferometer. II. Mars, 2500–3500 cm^{-1}. *Icarus* 15:1–10.

Bekker, M. G. 1969. *Introduction to Vehicle Systems* (Ann Arbor: Ann Arbor Press).

Bell, J. F., III, and Crisp, D. 1991. Imaging spectroscopy of Mars during 1990. *Eos: Trans. AGU* 72:521.

Bell, J. F., III, McCord, T. B., and Owensby, P. D. 1990. Observational evidence of iron oxides on Mars. *J. Geophys. Res.* 95:14447–14461.

Biemann, K. 1979. The implications and limitations of the findings of the Viking organic analysis experiment. *J. Molec. Evol.* 14:65–70.

Biemann, K., Oró, J., Toulmin, P., III, Orgel, L. E., Nier, A. O., Anderson, D. M., Simmonds, P. G., Flory, D., Diaz, A. V., Rushneck, D. R., Biller, J. E., and LaFleur, A. L. 1977. The search for organic substances and inorganic volatile

compounds in the surface of Mars. *J. Geophys. Res.* 82:4641–4658.
Blackburn, T. R., Holland, H. D., and Ceasar, G. P. 1979. Viking gas exchange reaction: Simulation on UV-irradiated manganese dioxide substrate. *J. Geophys. Res.* 84:8391–8394.
Blaney, D. L. 1991. Infrared imaging of Mars between 2.4 mm and 5.0 mm. *Lunar Planet. Sci.* XXII:111–112 (abstract).
Blaney, D. L., and McCord, T. B. 1990*a*. Earth-based telescopic observations of Mars in the 4.4 μm to 5.1 μm region. *Lunar Planet. Sci.* XXI:99–100 (abstract).
Blaney, D. L., and McCord, T. B. 1990*b*. An observational search for carbonates on Mars. *J. Geophys. Res.* 94:10159–10166.
Bogard, D. D., Nyquist, L. E., and Johnson, P. 1984. Noble gas contents of shergottites and implications for the Martian origin of SNC meteorites. *Geochim. Cosmochim. Acta* 48:1723–1739.
Boslough, M. B. 1988. Evidence for meteoritic enrichment of the Martian regolith. *Lunar Planet. Sci.* XIX:120–121 (abstract).
Boyd, R. C., Thompson, P. S., and Clark, B. C. 1989. Duricrete and composites construction on Mars. In *Case for Mars III*, ed. C. Stoker (San Diego: Univelt), pp. 539–550.
Burns, R. G. 1988. Gossans on Mars. *Lunar Planet. Sci.* XVIII:713–721 (abstract).
Chen, J. H., and Wasserburg, G. J. 1986. Formation ages and evolution of Shergotty and its parent planet from U-Th-Pb systematics. *Geochim. Cosmochim. Acta* 50:955–968.
Clark, B. C. 1984. Chemistry of the Martian surface: Resources for the manned exploration of Mars. In *The Case for Mars*, ed. P. J. Boston (San Diego: Univelt), pp. 197–208.
Clark, B. C., and Van Hart, D. C. 1981. The salts of Mars. *Icarus* 45:370–378.
Clark, B. C., Baird, A. K., Rose, H. J., Toulmin, P., III, Keil, K., Castro, A. J., Kelliher, W. C., Rowe, C. D., and Evans, P. H. 1976. Inorganic analysis of Martian surface samples at the Viking landing sites. *Science* 194:1283–1288.
Clark, B. C., Baird, A. K., Rose, H. J., Jr., Toulmin, P., III, Christian, R. P., Kelliher, W. C., Castro, A. J., Rowe, C. D., Kiel, K., and Huss, G. 1977. The Viking X-ray fluorescence experiment: Analytical methods and early results. *J. Geophys. Res.* 82:4577–4594.
Clark, B. C., Baird, A. K., Weldon, R. J., Tsusaki, D. M., Schnabel, L., and Candelaria, M. P. 1982. Chemical composition of Martian fines. *J. Geophys. Res.* 87:10059–10067.
Clark, R. N., Swayze, G. A., Singer, R. B., and Pollack, J. B. 1990. High-resolution reflectance spectra of Mars in the 2.3 μm region: Evidence for the mineral scapolite. *J. Geophys. Res.* 95:14463–14480.
Coey, J. M., Mrup, S., Madsen, M. B., and Knudsen, J. M. 1990. Titanomagenite in magnetic soils on Earth and Mars. *J. Geophys. Res.* 95:14423–14426.
Conrath, B. J., Curran, R., Hanel, R., Kunde, V., Maguire, W., Pearl, J., Pirraglia, J., Welker, J., and Burke, T. 1973. Atmospheric and surface properties of Mars obtained by infrared spectroscopy on Mariner 9. *J. Geophys. Res.* 78:4267–4278.
Cordell, B. 1985. A preliminary assessment of Martian natural resource potential. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 627–639.
Encrenaz, T., and Lellouch, E. 1990. On the atmospheric origin of weak absorption features in the infrared spectrum of Mars. *J. Geophys. Res.* 95:14489–14594.
Evans, D. L., and Adams, J. B. 1979. Comparison of Viking Lander multispectral images and laboratory reflectance spectra of terrestrial samples. *Proc. Lunar Planet. Sci. Conf.* 10:1829–1834.
Evans, D. L., and Adams, J. B. 1980. Amorphous gels as possible analogs to Martian weathering products. *Proc. Lunar Planet. Sci. Conf.* 11:757–763.

Farmer, V. C. 1974. The anhydrous oxide minerals. In *The Infrared Spectra of Minerals*, ed. V. Farmer (London: Mineralogical Soc. of London), pp. 183–204.

Farmer, V. C., and Russell, J. D. 1964. The infra-red spectra of layer silicates. *Spectrochim. Acta* 20:1149–1173.

Flynn, G. J., and McKay, D. S. 1988. Meteorites on Mars. In *Mars Sample Return Science Workshop*, LPI Tech. Rept. 88-07, pp. 126–127.

Flynn, G. J., and McKay, D. S. 1990. Assessment of the meteoritic contribution to the Martian soil. *J. Geophys. Res.* 95:14497–14509.

French, J. R. 1985. The impact of Martian propellant production manufacturing on early manned missions. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 519–526.

Gibbs, H. J., Hilf, J. W., Holtz, W. G., and Walker, F. C. 1961. Shear strength of cohesive soils. In *Research Conference on Shear Strength of Cohesive Soils* (New York: American Soc. of Civil Engineers), pp. 33–162.

Gooding, J. L., ed. 1990. *Scientific Guidelines for Preservation of Samples Collected from Mars*, NASA TM-4184.

Gooding, J. L., and Keil, K. 1978. Alteration of glass as a possible source of clay minerals on Mars. *Geophys. Res. Lett.* 5:727–730.

Gooding, J. L., Wentworth, S. J., and Zolensky, M. E. 1988. Calcium carbonate and sulfate of possible extraterrestrial origin in the EETA 79001 meteorite. *Geochim. Cosmochim. Acta* 52:909–915.

Gooding, J. L., Aggrey, K. E., and Muenow, D. W. 1990. Volatile compounds in shergottite and nakhlite meteorites. *Meteoritics* 25:281–289.

Gooding, J. L., Wentworth, S. J., and Zolensky, M. E. 1991. Aqueous alteration of the Nakhla meteorite. *Meteoritics* 26:135–143.

Guinness, E. A., Arvidson, E. E., Gehret, D. C., and Bolef, L. K. 1979. Color changes at the Viking landing sites over the course of a Mars year. *J. Geophys. Res.* 84:8355–8364.

Guinness, E. A., Arvidson, R. E., Dale-Bannister, M., Singer, R. B., and Bruckenthal, E. A. 1987. On the spectral reflectance properties of materials exposed at the Viking landing sites. *Proc. Lunar Planet. Sci. Conf.* 17, *J. Geophys. Res. Suppl.* 92:E575–E587.

Hanel, R. A., Conrath, B., Hovis, W., Kunde, V., Lowman, P., Maguire, W., Pearl, J., Pirraglia, J., Prabhakara, C., Schlachman, B., Levin, G., Straat, P., and Burke, T. 1972. Investigation of the Martian environment by infrared spectroscopy on Mariner 9. *Icarus* 17:423–442.

Hargraves, R. B., Collinson, D. W., Arvidson, R. E., and Spitzer, C. R. 1977. The Viking magnetic properties experiment: Primary mission results. *J. Geophys. Res.* 82:4547–4558.

Hargraves, R. B., Collinson, D. W., Arvidson, R. E., and Cates, P. M. 1979. Viking magnetic properties experiment: Extended mission results. *J. Geophys. Res.* 84:8355–8364.

Herr, K. C., and Pimentel, G. C. 1969. Infrared absorptions near three microns recorded over the polar cap of Mars. *Science* 166:496–498.

Horowitz, N., Hobby, G., and Hubbard, J. 1977. Viking on Mars: The carbon assimilation experiments. *J. Geophys. Res.* 82:4659–4662.

Houck, J. R., Pollack, J. B., Sagan, C., Schack, P., and Decker, J. A. 1973. High altitude infrared spectroscopic evidence for bound water on Mars. *Icarus* 18:470–479.

Hubbard, J. S. 1979. Laboratory simulations of the pyrolitic release experiments: An interim report. *J. Molec. Evol.* 14:211–221.

Huguenin, R. L. 1982. Chemical weathering and the Viking biology experiments on Mars. *J. Geophys. Res.* 87:10069–10082.

Huguenin, R. L. 1987. The silicate component of Martian dust. *Icarus* 70:162–188.

Hunt, G. R. 1977. Spectral signatures of particulate minerals in the visible and near-infrared. *Geophysics* 42:501–513.

Hunt, G. R., and Salisbury, J. W. 1970. Visible and near-infrared spectra of minerals and rocks. I. Silicate minerals. *Modern Geol.* 1:283–300.

Hunten, D. M. 1979. Possible oxidant sources in the atmosphere and surface of Mars. *J. Molec. Evol.* 14:57–64.

Hunten, D. M. 1987. Oxidants in the Martian atmosphere and soil. In *Study of Martian Surface and Atmospheric Effects on Mars Rover, Lander and Balloons*, JPL Tech. Rept. D-4657, pp. 4.7–4.13.

Kieffer, H. H. 1976. Soil and surface temperatures at the Viking Landing sites. *Science* 194:1344–1346.

Kieffer, H. H., Martin, T. Z., Peterfreund, A. R., Jakosky, B. M., Miner, E. D., and Palluconi, F. D. 1977. Thermal and albedo mapping of Mars during the Viking primary mission. *J. Geophys. Res.* 82:4249–4291.

Klein, H. P. 1978. The Viking biological experiments on Mars. *Icarus* 34:666–674.

Klein, H. P. 1979. The Viking mission and the search for life on Mars. *Rev. Geophys. Space Phys.* 17:1655–1662.

Laul, J. C. 1986. The Shergotty consortium and SNC meteorites: An overview. *Geochim. Cosmochim. Acta* 50:875–887.

Laul, J. C., Smith, M. R., Wänke, H., Jagoutz, E., Dreibus, G., Palme, H., Spettel, B., Burghele, A., Lipschutz, M. E., and Verkouteren, R. M. 1986. Chemical systematics of the Shergotty meteorite and the composition of its parent body (Mars). *Geochim. Cosmochim. Acta* 50:909–926.

Levin, G. V., and Straat, P. A. 1981. A search for nonbiological explanation of the Viking labeled release life detection experiment. *Icarus* 45:494–516.

Lindsay, W. L. 1979. *Chemical Equilibria in Soils* (New York: Wiley Interscience).

Martin, T. Z. 1985. Data set restoration: The Mariner 6/7 infrared spectrometer. *Bull. Amer. Astron. Soc.* 17:723–724 (abstract).

Martin, T. Z., and Kieffer, H. H. 1985. Improved access to Martian infrared radiometry/spectroscopy datasets. *Lunar Planet. Sci.* XVI:523–524 (abstract).

Mazur, P., Barghoorn, E. S., Halvorson, H. O., Jukes, T. H., Kaplan, I. R., and Margulis, L. 1978. Biological implications of the Viking mission to Mars. *Space Sci. Rev.* 22:3–34.

McCord, T. B., and Westphal, J. A. 1971. Mars: Narrowband photometry, from 0.3 to 2.5 microns, of surface regions during the 1969 apparition. *Astrophys. J.* 168:141–153.

McCord, T. B., Huguenin, R. L., Mink, D., and Pieters, C. 1977. Spectral reflectance of Martian areas during the 1973 opposition: Photoelectric filter photometry 0.33–1.10 μm. *Icarus* 31:25–39.

McCord, T. B., Singer, R. B., Hawke, B. R., Adams, J. B., Evans, D. L., Head, J. L., Mouginis-Mark, P. J., Pieters, C. M., Huguenin, R. L., and Zisk, S. H. 1982. Mars: Definition and characterization of global surface units with emphasis on composition. *J. Geophys. Res.* 87:10129-10148.

McKay, C. P., and Stoker, C. R. 1989. The early environment and its evolution on Mars: implications for life. *Rev. Geophys.* 27:189–214.

McSween, H. Y., Jr. 1985. SNC meteorites: Clues to Martian petrologic evolution? *Rev. Geophys.* 23:391–416.

Meyer, T. R., and McKay, C. P. 1984. The atmosphere of Mars—resources for the exploration and settlement of Mars. In *The Case for Mars*, ed. P. J. Boston (San Diego: Univelt), pp. 209–232.

Meyer, T. R., and McKay, C. P. 1989. The resources of Mars for human settlement. *J. Brit. Interplanet. Soc.* 42:147–160.

Moore, H. J., Hutton, R. E., Scott, R. F., Spitzer, C. R., and Shorthill, R. W. 1977.

Surface materials of the Viking landing sites. *J. Geophys. Res.* 82:4497–4523.

Moore, H. J., Liebes, S., Jr., Crouch, D. S., and Clark, L. V. 1978. *Rock Pushing and Sampling Under Rocks on Mars.* U. S. Geological Survey Prof. Paper 1081.

Moore, H. J., Spitzer, C. R., Bradford, K. Z., Cates, P. M., Hutton, R. E., and Shorthill, R. W. 1979. Sample fields of the Viking landers, physical properties, and aeolian processes. *J. Geophys. Res.* 84:8365–8377.

Moore, H. J., Clow, G. D., and Hutton, R. E. 1982. A summary of Viking sample-trench analyses for angles of internal friction and cohesions. *J. Geophys. Res.* 87:10043–10050.

Moore, H. J., Hutton, R. E., Clow, G. D., and Spitzer, C. R. 1987. *Physical Properties of the Surface Materials of the Viking Landing Sites on Mars.* U. S. Geological Survey Prof. Paper 1389.

Moroz, V. I. 1964. The infrared spectrum of Mars (1.1–4.1μm). *Astron. Zh.* 8:273–281 (in Russian).

Morris, R. V., and Lauer, H. V. 1990. Matrix effects for reflectivity spectra of dispersed nanophase (superparamagnetic) hematite with application to Martian spectral data. *J. Geophys. Res.* 95:5101–5109.

Morris, R. V., Lauer, H. V., Jr., Lawson, C. A., Gibson, E. K., Jr., Nace, G. A., and Stewart, C. 1985. Spectral and other physicochemical properties of submicron powders of hematite (a-), maghemite (g-Fe_2O_3), magnetite (Fe_3O_4), goethite (a-FeOOH), and lepidocrocite (g-FeOOH). *J. Geophys. Res.* 90:3126–3144.

Morris, R. V., Agresti, D. G., Lauer, H. C., Jr., Newcomb, J. A., Shelfer, T. D., and Murali, A. V. 1989. Evidence for pigmentary hematite on Mars based on optical, magnetic, and Mossbauer studies of superparamagnetic (nanocrystalline) hematite. *J. Geophys. Res.* 94:2760–2778.

Morris, R. V., Gooding, J. L., Lauer, H. V., Jr., and Singer, R. B. 1990. Iron mineralogy of a Hawaiian palagonitic soil with Mars-like spectral and magnetic properties. *J. Geophys. Res.* 95:14427–14434.

Mutch, T. A., Arvidson, R. E., Binder, A. B., Guiness, E. A., and Morris, E. C. 1977. The geology of the Viking Lander 2 site. *J. Geophys. Res.* 82:4452–4467.

Orenberg, J., and Handy, J. 1992. Reflectance spectroscopy of palagonite and iron-rich montmorillonite clay mixtures: Implications for the surface composition of Mars. *Icarus* 96:219–225.

Oyama, V. I., and Berdahl, B. J. 1977. The Viking gas exchange experiment results from Chryse and Utopia surface samples. *J. Geophys. Res.* 82:4669–4676.

Oyama, V. I., and Berdahl, B. J. 1979. A model of Martian surface chemistry. *J. Molec. Evol.* 14:199–210.

Pimentel, G. C., Forney, P. B., and Herr, K. C. 1974. Evidence about hydrate and solid water in the Martian surface from the 1969 Mariner infrared spectrometer. *J. Geophys. Res.* 79:1623–1634.

Plumb, R. C., Tantayanon, R., Libby, M., and Xu, W. W. 1989. Chemical model for Viking biology experiments: Implications for the composition of the Martian regolith. *Nature* 338:633–635.

Pollack, J. B., Roush, T., Witteborn, F., Bregman, J., Wooden, D., Stoker, C., Toon, O. B., Rank, D., Dalton, B., and Freedman, R. 1990. Thermal emission spectra of Mars (5.4–10.5 μm): Evidence for sulfates carbonates and hydrates. *J. Geophys. Res.* 95:14595–14627.

Ponnamperuma, C., Shimoyama, A., Yamada, M., Hobo, T., and Pal, R. 1977. Possible surface reactions on Mars: Implications for Viking biology results. *Science* 197:455–457.

Rosenqvist, T. 1983. *Principles of Extracting Metallurgy*, 2nd ed. (New York: McGraw-Hill).

Roush, T. L. 1982. Effects of iron silica gels on spectral reflectance. *Lunar Planet.*

Sci. XIII:661–662 (abstract).

Roush, T. L., Roush, E. A., Singer, R. B., and Lucey, P. G. 1988. Preliminary analysis of recent 2.2–4.2 μm telescopic observations of Elysium, Mars: Implications for the crystallinity and hydration state of surface materials. In *MEVTV Workshop on Nature and Composition of Surface Units on Mars*, LPI Tech. Rept. 88-05, pp. 111–113.

Roush, T. L., Pollack, J., Stoker, C., Witteborn, F., Bregman, J., Wooden, D., and Rank, D. 1989. CO_3^{2-}– and SO_4^{2-}–bearing anionic complexes detected in Martian atmospheric dust. *Lunar Planet. Sci.* XX:928–929 (abstract).

Roush, T. L., Blaney, D. L., and Singer, R. B. 1991*a*. The surface composition of Mars as inferred from spectroscopic observations. In *Remote Geochemical Analysis*, eds. C. Pieters and P. Engelhart (New York: American Chemical Soc.).

Roush, T. L., Witteborn, F., Lucey, P., Graps, A., and Pollack, J. 1991*b*. Thermal infrared observations of Mars (7.5–12.8 μm) during the 1990 opposition. *Lunar Planet. Sci.* XXII:1137–1140 (abstract).

Roush, T. L., Witteborn, F., Bregman, J., Rank, D., Graps, A., and Pollack, J. 1992. Thermal infrared spectra (5.5–9.2 μm) of Mars obtained from the Kuiper Airborne Observatory. *Lunar Planet. Sci.* XXIII:1181–1182 (abstract).

Sharp, R. P., and Malin, M. C. 1984. Surface geology from Viking landers on Mars: A second look. *Geol. Soc. Amer. Bull.* 95:1398–1412.

Sherman, D. M., Burns, R. G., and Burns, V. M. 1982. Spectral characteristics of the iron oxides with application to the Martian bright region mineralogy. *J. Geophys. Res.* 87:10169–10180.

Shorthill, R. W., Moore, H. J., Scott, R. F., Hutton, R. E., Lieves, S., and Spitzer C. R. 1976. The soil of Mars (Viking 1). *Science* 194:91–97.

Simonsen, L. C., Nealy, J. E., Townsend, L. W., and Wilson, J. W. 1990. *Radiation Exposure for Manned Mars Surface Missions*, NASA Tech. Paper 2979.

Singer, R. B. 1980. The dark materials on Mars: I. New information from reflectance spectroscopy on the extent and mode of oxidation. *Lunar Planet. Sci.* XI:1045–1047 (abstract).

Singer, R. B. 1981. Near-infrared spectra: Reflectance of mineral mixtures: Systematic combinations of pyroxenes, olivine, and iron oxides. *J. Geophys. Res.* 86:7967–7982.

Singer, R. B. 1982. Spectral evidence for the mineralogy of high-albedo soils and dust on Mars. *J. Geophys. Res.* 87:10159–10168.

Singer, R. B. 1985. Spectroscopic observation of Mars. *Adv. Space Res.* 5:59–68.

Singer, R. B., and Roush, T. L. 1983. Spectral reflectance properties of particulate weathered coatings on rocks: Laboratory modeling and applicability to Mars. *Lunar Planet. Sci.* XIV:708–799 (abstract).

Singer, R. B., McCord, T. B., Clark, R. N., Adams, J. B., and Huguenin, R. L. 1979. Mars surface composition from reflectance spectroscopy: A summary. *J. Geophys. Res.* 84:8415–8426.

Singer, R. B., Owensby, P. D., and Clark, R. N. 1985. Observed upper limits for clay minerals on Mars. *Lunar Planet. Sci.* XVI:787–788 (abstract).

Singer, R. B., Bruckenthal, E. A., Roush, T. L., and Lucey, P. G. 1986. Mars: spatially resolved spectrophotometry from 2.2–4.2 μm. *Bull. Amer. Astron. Soc.* 18:806 (abstract).

Singer, R. B., Miller, J. S., Wells, K. W., and Bus, E. S. 1990. Visible and near-IR spectral imaging of Mars during the 1988 opposition. *Lunar Planet. Sci* XXI:1164–1165 (abstract).

Sinton, W. M. 1967. On the composition of Martian surface material. *Icarus* 6:222–228.

Soderblom, L. A., and Wenner, D. B. 1978. Possible fossil H_2O liquid-ice interfaces

in the Martian crust. *Icarus* 34:622–637.
Soderblom, L. A., Edwards, K., Eliason, E. M., Sanchez, E. M., and Charette, M. P. 1978. Global color variations on the Martian surface. *Icarus* 34:446–464.
Stubican, V., and Roy, R. 1961. Isomorphous substitution of infra-red spectra of the layer lattice silicates. *Amer. Mineralogist* 46:32–51.
Swayze, G. A., and Clark, R. N. 1990. Infrared spectra and crystal chemistry of scapolites: Implications for Martian mineralogy. *J. Geophys. Res.* 95:14481–14495.
Swindle, T. D., Caffee, M. W., Hohenberg, C. M., Hudson, G. B., and Rajan, R. S. 1984. Noble gases in SNC meteorites. *Meteoritics* 19:318–319.
Toulmin, P., III, Clark, B. C., Baird, A. K., Keil, K., and Rose, H. J., Jr. 1976. Preliminary results of the Viking X-ray Fluoresence Experiment: The first sample from Chryse Planitia, Mars. *Science* 194:81–84.
Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christian, R. P., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.
Wasson, J. T., Boynton, W. V., and Chou, C. L. 1975. Compositional evidence regarding influx of interplanetary material onto the lunar surface. *The Moon* 13:121–141.
Wright, I. P., Carr, R. H., and Pillinger, C. T. 1986. Carbon abundance and isotopic studies of Shergotty and other shergottite meteorites. *Geochim. Cosmochim. Acta* 50:983–991.

THE IGNEOUS CRUST OF MARS: COMPOSITIONAL EVIDENCE FROM REMOTE SENSING AND THE SNC METEORITES

R. B. Singer
University of Arizona

and

H. Y. McSween, Jr.
University of Tennessee

Our knowledge of the composition of unaltered igneous crust on Mars is based on information from a variety of indirect sources. Two primary sources are reflectance spectroscopy and study of the SNC meteorites. Virtually all remote observations indicate that there is abundant unaltered basaltic crustal material exposed at or very near the present Martian surface, concentrated in low-albedo regions. Near-infrared reflectance spectroscopy provides compositional information about Fe^{2+}-bearing igneous minerals such as pyroxenes, which have been consistently observed on the Martian surface. As estimated from spectral analysis, the composition averages relatively high-Fe, low- to medium-Ca clinopyroxene, with some regional variation. The SNC meteorites are a small and diverse group of relatively young igneous rocks generally thought to have originated in the crust of Mars. Shergotty, a basaltic meteorite from this group, is consistent in important mineralogic features with telescopic observations of low-albedo regions on the planet. Because shergottites formed as basalt flows or shallow cumulates it is reasonable that such materials would be exposed on the surface of their parent planet. Spectral analogs to the other two SNC meteorite families, the nahklites and chassignites, have not been detected at the Martian surface, consistent with their petrogenesis as higher-pressure cumulate assemblages. While the SNCs sampled relatively young igneous crust, there is evidence that the ancient crust of Mars is compositionally similar. Abundant ultramafic igneous rock at the Martian surface is neither required by nor consistent with the bulk of observational evidence. There is also no evidence for highly differentiated intermediate or acidic crust on Mars, and significant constraints on its possible abundance.

I. INTRODUCTION

Our knowledge of the composition of unaltered igneous crust on Mars is based on information from a variety of sources, all indirect. These sources include photogeology, remote sensing, *in-situ* chemical measurements of weathered soils, and very likely the SNC meteorites. There have been no direct measurements of the chemistry or petrology of the crust or any other igneous

rocks on Mars. Despite this, the evidence we do have forms a consistent and reasonable picture of the composition of the planet. The purpose of this chapter is to review and synthesize that evidence. Primary emphasis is placed on reflectance spectroscopy and the SNC meteorites. Other relevant sources of information are discussed as appropriate.

A. Global Geology

While Mars is in many ways the most Earth-like of the other planets, there are some significant differences. Mars is a small planet, roughly half the diameter of the Earth and twice the diameter of the Moon. Mars is also considerably less dense than the Earth: the mean density is 3.9 gm cm^{-3}, compared to 5.5 gm cm^{-3} for the Earth and 3.3 gm cm^{-3} for the Moon. Despite the low bulk density of Mars the surface is more iron-rich than the Earth, based on results of remote sensing and *in-situ* chemical measurements of weathered soils. Our knowledge of the Martian interior is quite limited.

Because of its smaller size Mars presumably lost its internal heat faster than the Earth. Nevertheless, there is abundant evidence for volcanism throughout the history of Mars, by some estimates terminating less than half a billion years ago (see, e.g., Carr 1981; Greeley and Spudis 1981) and possibly continuing to the present (Lucchitta 1987). This evidence is primarily geomorphologic, interpreted from orbital imaging obtained by the Mariner 9 and Viking missions. The largest features produced by volcanic activity include voluminous flood lavas and massive shield volcanoes. Other observed volcanic features range down in size to wrinkle ridges, individual lava flows, leveed channels, and small vents. Caution is required when interpreting rock composition from morphologic evidence alone, as many volcanic landforms can be produced from magmas of different compositions (see, e.g., Greeley 1987). Nevertheless, all observed volcanic features on Mars are consistent with lavas of very low viscosity, generally interpreted to be rich in iron and poor in silica (see, e.g., Carr 1981).

Mars has a roughly hemispherical global dichotomy between ancient heavily cratered southern highlands and younger, resurfaced northern lowlands, as shown in Fig. 1. Relative ages (and approximate absolute ages) of Mars surface units are determined by crater densities, as reviewed by Carr (1981). The mean difference in elevation between the highlands and lowlands is 3 to 5 km (Carr 1981). The origin of this dichotomy is not well understood. The oldest remnant crust on Mars is preserved in the highland ancient cratered terrain, the surface surviving the end of late heavy bombardment ~3.9 Gyr (Carr 1981). The oldest identifiable volcanic flows are found in the highland intercrater ridged plains unit, interpreted to have formed immediately thereafter. These two units are mapped together as cu in Fig. 1. Subsequent volcanic plains, in order of decreasing age, are mapped as pc, pm, and pv. All three of these units are estimated to have formed before ~2.5 Gyr (Carr 1981).

One of the most striking large geologic features on Mars is the Thar-

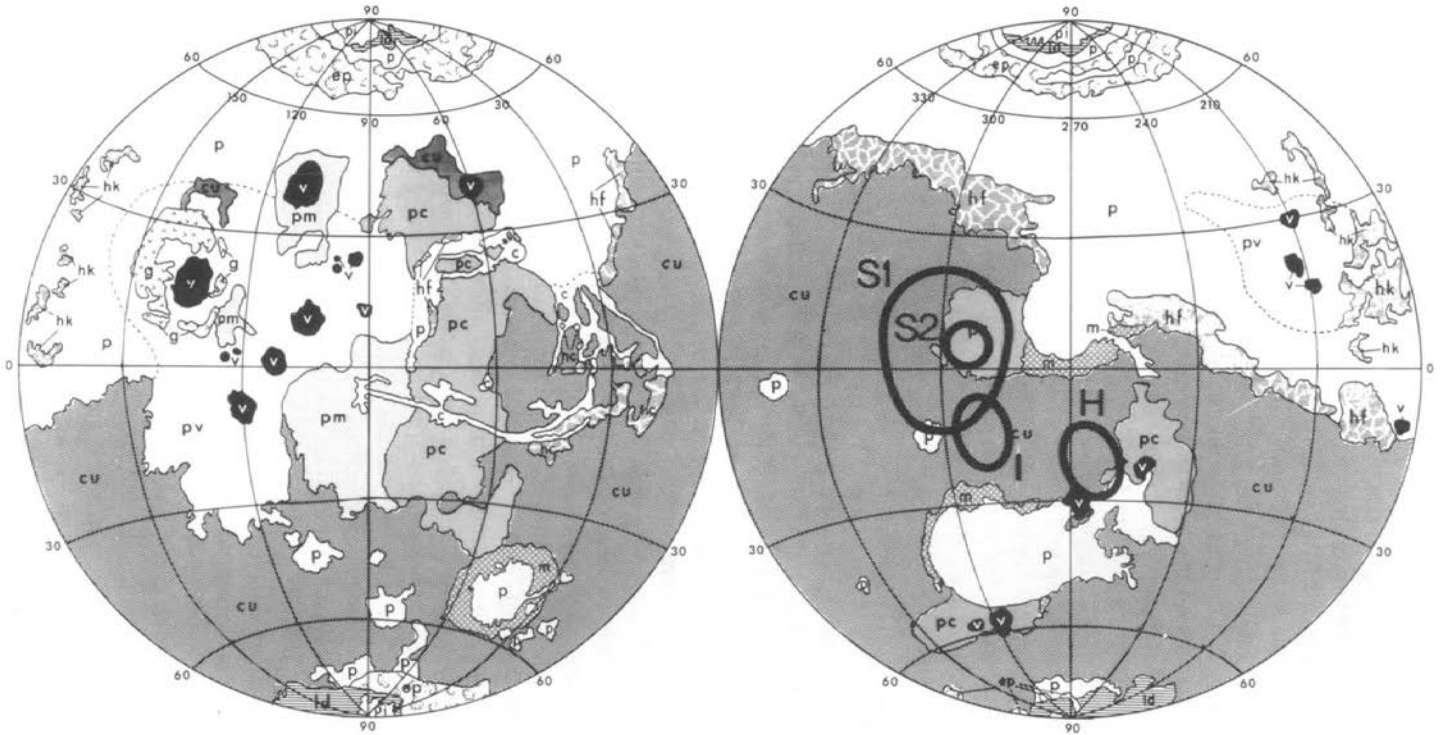


Figure 1. Physiographic units of Mars plotted on a Lambert equal area projection. Note the roughly hemispherical dichotomy between the ancient cratered highlands (cu) generally to the south, and the much younger resurfaced plains (p) to the north. Volcanic plains units, younger than cu but older than p, are shown as pc, pm, and pv, in order of decreasing age. Volcanic constructs of all ages are mapped as v, including the large Tharsis shields. Ages and relationships are discussed further in the text. The four labeled outlines indicate the locations of telescopic spectral observations presented and analyzed in this chapter. H is Hesperia, I is Iapygia, S1 is Syrtis 4/7, and S2 is Syrtis 4/8. Note that Syrtis 4/7 was observed with a larger aperture and overlaps all of Syrtis 4/8 and part of Iapygia, covering both Syrtis Major Planitia (pc) and surrounding older terrain (cu). Syrtis 4/8 lies entirely within Syrtis Major Planitia. (Physiographic map from Mutch et al. 1976.)

sis bulge, centered near 10°S, 110°W. Tharsis is an uplifted dome roughly 5000 km wide and 10 km high, upon which are located a number of enormous shield volcanoes including Olympus Mons. (These volcanoes, as well as other volcanic constructs of various ages are mapped as v in Fig. 1.) The Tharsis bulge is not isostatically compensated, apparently supported by a thick rigid lithosphere or by dynamic pluming from the mantle (see, e.g., Phillips 1978). The surfaces of Olympus Mons and the other large Tharsis shields are very young, perhaps on the order of hundreds of Myr. There is good evidence, however, that these large structures accumulated for a considerable time, beginning as long ago as 2.5 to 3.0 Gyr. (Carr 1981).

Adjacent to Tharsis is a large canyon system, Valles Marineris, on the order of 3000 km long and up to 7 km deep. While there is some evidence for aeolian, fluvial, and possibly volcanic modification of Valles, the overall canyon system is generally felt to be tectonic in origin, an extensional rift valley system (Lucchitta et al. 1991). Plate tectonics has not operated on Mars, nor is there evidence for compressional tectonics (Carr 1981).

B. Color and Albedo

The brightness, color, and spectral properties of Mars are controlled by the composition and physical characteristics of the extreme surface layers. To first order, there are two surface color units: high-albedo heavily weathered orange-tan Fe^{3+}-rich "bright regions," and low-albedo unaltered gray Fe^{2+}-rich "dark regions." Detailed spectral, physical, and compositional relationships are of course more complicated, reflecting compositional variations and extensive surficial mixing of surface materials. Color relationships have been studied at high spatial resolution using multispectral imaging from the Viking Orbiters and Landers (see, e.g., Soderblom et al. 1978; McCord et al. 1982*a*; Guinness et al. 1987; Arvidson et al. 1989). Compositionally, these data provide an indication of degree of oxidation and some broad mineralogic constraints, but cannot provide detailed mineralogic or chemical information because of their limited spectral resolution and coverage. Reflectance spectroscopy provides considerably more compositional information, but because most of these observations have been made from Earth telescopically the spatial resolution is poor, on the order of many hundreds of kilometers.

Surface color and albedo relationships often do not correlate with geomorphologic units and boundaries. While the atmosphere of Mars is quite thin (~7 mbar average), it is responsible for considerable aeolian action. Much of the heavily weathered material is extremely fine grained, and is periodically redistributed by large regional and global dust storms on Mars. That some dark material is also redistributed on more local scales by aeolian action is demonstrated by abundant dark sand dunes on the planet. Nevertheless, much of the dark material observed at the surface is thought to be *in situ* or at least locally or regionally derived. In some cases spectral boundaries do clearly correlate with morphologic boundaries (see, e.g., Geissler et al. 1990; Murchie et al. 1991). The dark regions appear to be composed of regionally

variable combinations of dark soils, blocks, and bedrock. While periodically "dusted" by weathered material, dark surface regions consistently reappear after dust activity (documented over hundreds of years of telescopic observations). Thus all of Mars is not blanketed by weathered dust, and crustal rock is reasonably abundant at or very near the surface in many regions.

C. Viking Lander Chemical Measurements

Direct chemical analyses of fine-grained weathered soils at two locations on Mars were made by X-ray fluorescence spectrometers carried by the Viking Landers. Accurate analyses were obtained for the elements Si, Al, Fe, Mg, Ca, K, Ti, S and Cl. Compared to typical terrestrial continental soils and lunar mare fines, measured Martian fines are higher in Fe, lower in Al, and much higher in S and Cl (see, e.g., B. C. Clark et al. 1982). The average composition measured at the two widely separated landing sites was nearly identical, leading some researchers to interpret a globally uniform soil unit. (Spectral evidence indicates there is some regional compositional variations do exist among weathered soil deposits [see, e.g., Bell et al. 1990; Clark et al. 1990; Erard et al. 1990].) Sulfur and chlorine in the samples, presumed to occur as salts, are generally thought to be associated with slight and somewhat variable induration of the soils (duracrust), and may have originated as volcanic aerosols (Settle 1979). While these chemical results do not permit a unique mineralogic interpretation of either the fine-grained soil or the origin of these fines, the source material is constrained to be predominantly mafic to ultramafic. There has been continued debate about whether these soils originated through isochemical or differential weathering of the source material (see, e.g., Baird and Clark 1981; B. C. Clark et al. 1982).

II. COMPOSITIONAL EVIDENCE FROM SPECTROSCOPY

Various types of remote spectroscopy provide valuable information about the composition of Mars. Prior to spacecraft missions, all studies of the composition of Mars were astronomical, based on spectroscopy of reflected sunlight and/or thermally emitted radiation from the planet's surface. Astronomical observations continue to provide important new information at the present time, as new techniques and instrumentation are available to Earth-based researchers at least a decade before they are flown on spacecraft. An inherent limitation of astronomical observations is poor spatial resolution, typically a few hundred to a thousand kilometers on Mars. There have been a number of near-infrared and thermal-infrared spectrometers flown to Mars on spacecraft, which have produced limited results concerning crustal composition. Petrologically more complete observations are planned for future U. S. and Soviet missions to Mars.

Reflectance spectroscopy provides information primarily about mineralogy and secondarily about chemistry. The visible region of the spectrum (0.4

to 0.7 μm) provides diagnostic information on Mars mainly related to charge-transfer and crystal-field electronic transitions in ferric iron, associated most directly with the heavily weathered bright soils. Ferrous iron, on the other hand, has mineralogically diagnostic crystal-field electronic absorptions in the vicinity of 1 μm and 2 μm. The near infrared ($\sim$0.7 to 2.6 μm) is therefore one of the most productive wavelength regions for studying crustal composition. These wavelengths are especially useful for identifying and analyzing ferrous-iron bearing minerals such as pyroxenes and olivine. The mid- and thermal infrared ($\sim$2.6 μm and greater) generally sense vibrational molecular features, including water-bearing minerals, salts, and silicates (see the Chapter by Stoker et al.). Longward of $\sim$3.5 to 4 μm, radiation received from Mars is dominated by thermally-emitted flux. The 9- to 11- μm region is especially important because it includes Si-O bending fundamentals which provide diagnostic information about the degree of polymerization in silicates. At these wavelengths the entire signal from Mars is thermally emitted radiation, and measurement and analysis techniques are quite different from those for solar reflectance. Spectroscopic observations of weathered dust suspended in the Martian atmosphere by the IRIS instrument onboard Mariner 9 resulted in varied and somewhat conflicting compositional interpretations (Hanel et al. 1972; Hunt et al. 1973; Aronson and Emslie 1975; Toon et al. 1977). There were no IRIS measurements of low-albedo surface regions.

The most direct information we have concerning the composition of crustal igneous materials on Mars comes from near-infrared reflectance spectroscopy, which as previously mentioned is sensitive to Fe^{2+} in pyroxenes and olivine (Burns 1970; Adams 1974,1975). This technique is important and diagnostic for estimating basalt compositions from remote observations. Systematics for pyroxene composition vs the position of the two diagnostic Fe^{2+} absorption bands were first derived by Adams (1974) and recently extended by Cloutis and Gaffey (1991). The energy of the electronic transitions leading to the absorption bands is controlled by the size and shape of crystallographic sites, which in turn depends on cation content. In general, for orthopyroxenes and low-Ca clinopyroxenes the position of both band minima moves to longer wavelength with increasing Fe content. For higher-Ca clinopyroxenes both minima move to still longer wavelengths with increasing Ca content.

Pyroxene was first detected on Mars by Adams and McCord (1969) based on detection of a weak band near 1 μm observed for low-albedo regions. They spectrally matched these observations with a crystalline basaltic rock "weathered" in the laboratory to form a thin ferric-oxide stain. Since that time, many additional observations of low-albedo regions have been made (see, e.g., McCord and Westphal 1971; McCord et al. 1977,1982*b*; Singer et al. 1980, 1990*a*; Bell et al. 1990), consistently showing the 1-μm band. Data presented by McCord et al. (1977) show some regional variability in details of that band, implying compositional variability among Martian basalts. While Huguenin et al. (1978) interpreted those differences in mineralogic detail, Singer (1980) made the case that those data were not suitable for such detailed

interpretations. Observations in 1978 by McCord et al. (1982*b*) produced the first good-quality data to extend through both the 1- and 2-μm regions (0.65–2.55 μm). The pyroxene band near 1 μm is well defined in these data, centered near 0.95 μm, implying an augitic composition[a] (Singer 1980). Within the noise of these data there is not a clearly defined pyroxene band near 2 μm, limiting further compositional analysis.

Spectroscopic data obtained in 1980 by Singer et al. (1980) using the same instrumentation include a number of spectra of low-albedo regions in which both 1-μm and 2-μm pyroxene bands are apparent. An analysis of the pyroxene absorptions in these data was first presented by Singer and Roush (1985). As these remain one of the best sources of information relevant to the composition of igneous rocks on Mars, an updated analysis is presented here. Figure 2 shows spectra from 0.65 to 2.55 μm for four observations of low-albedo regions. The locations of these observed regions are shown in Fig. 1. The data are presented as scaled reflectance, differing from radiance factor by a single multiplicative factor for each spectrum. Data quality varies, as indicated by the error bars; spectra for the regions Hesperia and Iapygia, both observed on 4/10/80 (UT), are the highest quality. Both these spots lie predominantly within unit cu. The two Syrtis Major spectra were obtained on different nights with different size apertures. Syrtis 4/8 (4/8/80 UT) lies entirely within Syrtis Major Planitia, unit pc. Syrtis 4/7 includes all of Syrtis 4/8 and part of Iapygia. The broad absorption with a minimum near 0.95 μm is due to pyroxene on Mars. Most of the complicated absorption structure in the 2-μm region is caused by Martian atmospheric CO_2, significantly masking the broad and weak 2-μm pyroxene band. (A more complete discussion of Mars spectra throughout this wavelength region is provided by Singer [1985].)

To parameterize the positions of these bands linear continua were fit and divided out independently across the 1-μm and 2-μm regions of each of the spectra. Polynomials were then fit to each continuum-removed band, and the minimum determined by taking the first derivative of the polynomial. In the 1-μm region this process is straightforward, as demonstrated in Fig. 3a for the Hesperia data. There is some uncertainty on the long-wavelength side of the band near 1.2 μm caused by slightly imperfect removal of telluric water absorptions from these telescopic observations. These uncertainties were investigated and are reflected in the error estimates for the 1-μm band positions. There is greater uncertainty in the analysis of the 2-μm region because much of it is dominated by absorptions of atmospheric CO_2 on Mars. An example of this analysis process is shown in Fig. 3b. Because there is not yet an accurate way to model and remove the CO_2 absorption in the multiply scattering Martian atmosphere, the polynomial fit is based only on those wavelengths at which CO_2 does not absorb significantly. Because the

[a] The high-Ca pyroxene composition originally favored by Singer (1980) was in error due to a telluric atmospheric artifact. The alternate interpretation of augitic composition is now considered correct by the author.

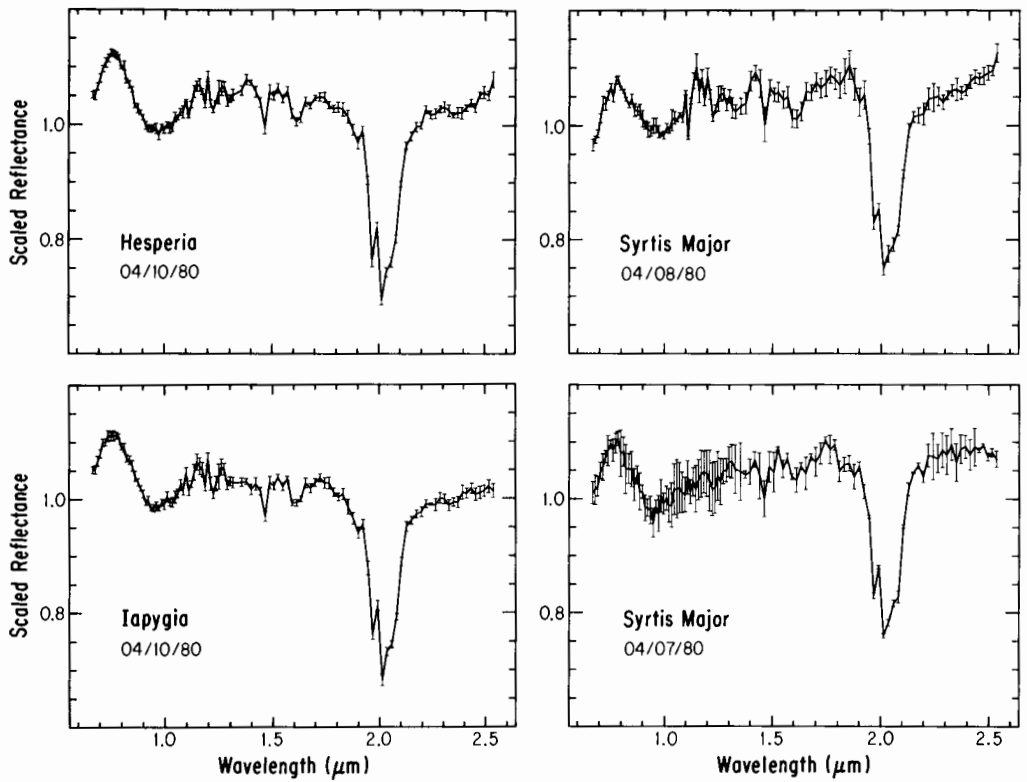


Figure 2. Scaled near-infrared reflectance spectra of four low-albedo regions on Mars. Data were obtained by Singer et al. (1980). These observations are for the spots shown in Fig. 1, near the named locations. Fe^{2+} pyroxene bands with minima just shortward of 1 μm are clearly defined, and show some variation in shape and width. The Fe^{2+} pyroxene bands near 2 μm are more difficult to see because of major interference by Mars atmospheric CO_2 absorptions, also centered near 2 μm.

fit is based on fewer points and the actual band center is in an interpolated region, the resultant band center location is likely to be less certain than for the 1-μm band. Nevertheless, experimentation with various fit parameters gave a convergence of results, with the uncertainty reflected in the error estimates. Figure 4 shows the Mars band-position results overlain on a plot of 1-μm vs 2-μm band positions for a wide variety of terrestrial pyroxenes. The Mars data cluster reasonably closely, in a central region of the trend characteristic of clinopyroxene with low to medium Ca and relatively high Fe. The differences among these four locations on Mars are probably real, although the observational uncertainties allow substantial overlap.

Both the Adams (1974) and Cloutis and Gaffey (1991) pyroxene calibrations were used to analyze these data. The results are similar but not identical. While the two pyroxene calibrations were developed in similar ways there are some technical differences. Additionally, Cloutis and Gaffey (1991) used a greater number of mineral samples, including many zoned and exsolved pyroxenes, adding completeness but inherently complicating subsequent interpretations. Of the telescopically observed regions discussed above, Hesperia, Iapygia, and Syrtis 4/7 yield fairly similar compositional estimates with relatively low Ca of ~10 to 20%[b] and relatively high Fe in excess of about 50%. The Cloutis and Gaffey (1991) calibration consistently yields estimates with somewhat lower Fe contents relative to Adams (1974). In both calibrations the possibility exists, for Iapygia especially, that the optically dominant pyroxene might have Fe as high as 80 to 85% with very low Ca (<5%). All of these interpretations are based on there being a single optically dominant pyroxene, which is a good assumption for many (but not all) terrestrial basalts. The interpretation of the Syrtis 4/8 spectrum raises the possible complications of multiple optically important pyroxene compositions, whether they occur as zoned or exsolved crystals, or as separate grains. Using the Adams (1974) calibration, Syrtis 4/8 is interpreted to have Fe in excess of ~40%, and Ca in the range of about 15 to 30%. This higher Ca content relative to the other Mars observations is consistent with the location of Syrtis 4/8 in Fig. 4, closer to the high-Ca pyroxenes plotted at the upper right. The interpretation of Syrtis 4/8 with the Cloutis and Gaffey (1991) calibration is less unique. While a consistent high-Fe medium-Ca composition is allowed, there are also other possibilities with lower Fe and higher Ca. This is apparently due to the well-documented inclusion in their study of some zoned and exsolved crystals, for which the optical properties do not necessarily combine according to the weighted average of the compositions. Research is continuing to see if a more unique interpretation, possibly indicative of multiple pyroxenes, is possible for Syrtis 4/8.

During the favorable 1988 Mars opposition, Singer et al. (1990*a*) obtained spectral images of Mars from 0.44 to 1.02 μm, combining high-

[b] In this pyroxene system composition is given as mole percent relative to the total amount of Mg, Fe, and Ca. Thus % Ca/Mg+Fe+Ca, and % Fe/Mg+Fe+Ca.

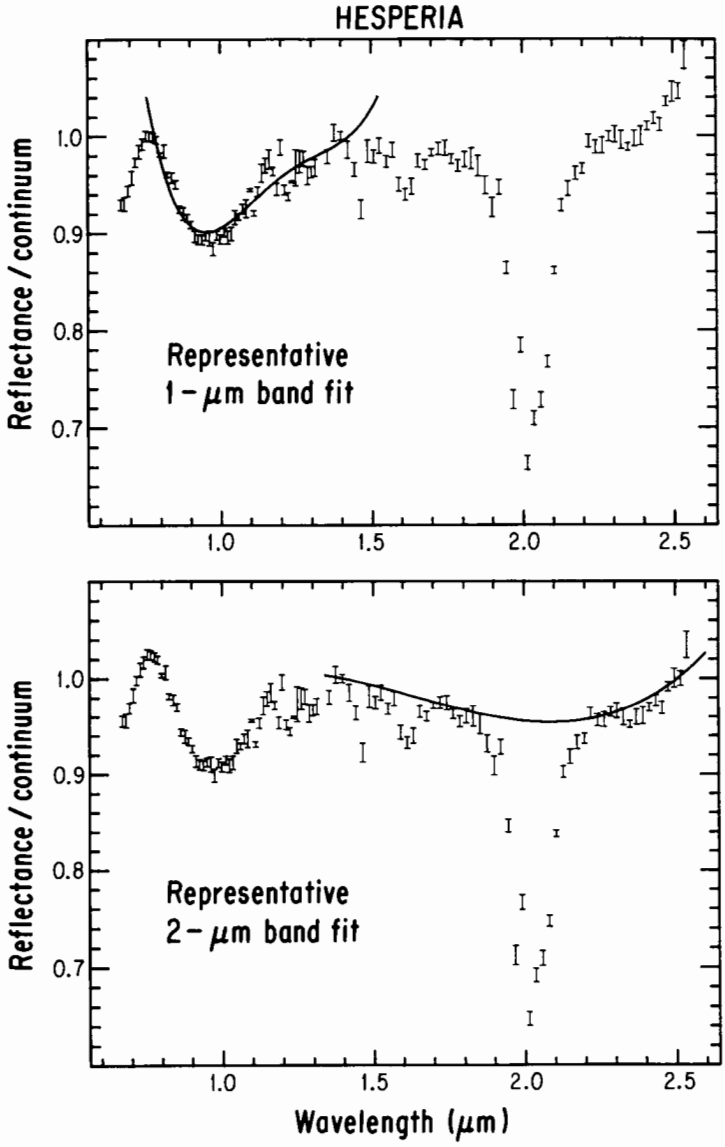


Figure 3. Representative examples of the band-fitting procedure used to determine wavelengths of minimum reflectance, as described in the text. Separate linear continua were removed by division from the 1-μm and 2-μm regions. Polynomials were then fit to the regions of interest to smooth the data and facilitate finding the minima. In the 2-μm region the polynomials were fit only to continuum points known not to have significant atmospheric absorption, and so interpolated across those absorptions. While band shape information was not used in these analyses, the 1-μm band shown here for Hesperia is relatively broad with a long-wavelength shoulder, possibly indicating the presence of some olivine along with the optically-dominant pyroxene.

resolution spectral information with spatial imaging (best-case resolution ~280×150 km). While these data do not cover the long-wavelength side of the 1-μm band, they do provide literally thousands of regional observations of 1-μm band center positions. Most dark regions analyzed so far show a pyroxene band centered variably between 0.93 and 0.97 μm, consistent with the results presented above. Certain low-albedo regions, most notably Acidalia Planitia in the northern hemisphere, are spectrally distinct and show no indication of pyroxene Fe^{2+} absorptions in this wavelength region (Singer et al. 1990*b*).

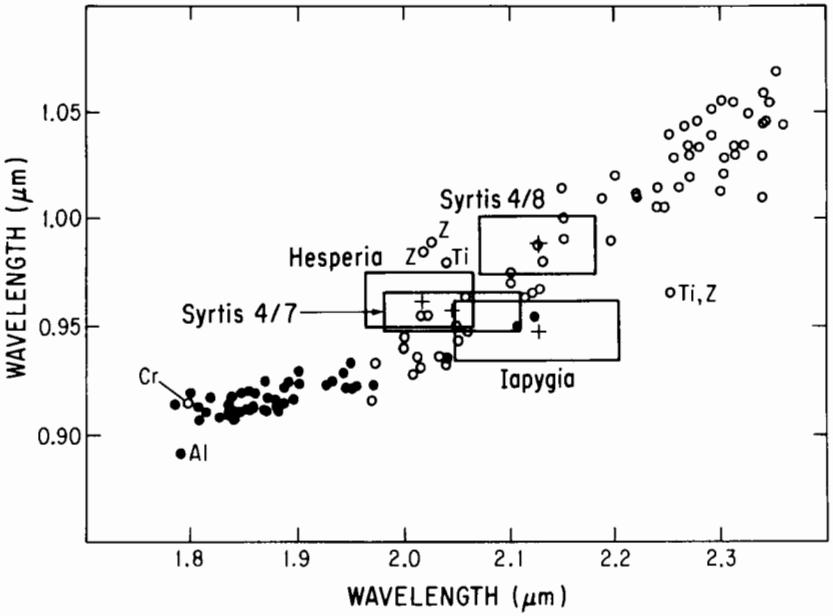


Figure 4. Long- and short-wavelength pyroxene band positions for the four Mars observations presented here are plotted on the "band-band" plot of Cloutis and Gaffey (1990), which is an augmentation of the original calibration of Adams (1974). This and related diagrams presented by those authors allow pyroxene mineralogy and composition to be estimated from spectral reflectance data. The circles represent spectral characteristics of individual terrestrial pyroxene samples measured in the laboratory. The Mars data (X) cluster fairly well on the main trend in a region characterized by augite clinopyroxenes with low to moderate Ca and relatively high Fe. The boxes indicate the degree of uncertainty in the band parameterization of the Martian spectral date.

Mars-orbital near-infrared spectral imaging (0.76 to 3.16 μm) was obtained for some regions on Mars at high spatial resolution (<20 km) in 1989 by the ISM instrument onboard the Soviet Phobos 2 spacecraft (Bibring et al. 1989). Many of those spectra have been analyzed by Murchie et al. (1992) and Mustard et al. (1992), using methods similar to those described above. Those

authors confirm the presence of Fe^{2+} pyroxene bands near 1 and 2 μm for most low-albedo regions observed. The band positions, and therefore implied compositions, are also in good agreement with the results discussed above. Locations for the short wavelength band vary from about 0.93 to 0.98 μm, with some variability in shape and width, possibly due to mixtures of mafic minerals in some regions. Long-wavelength band positions near 2.1 to 2.2 μm are also reported. Data for low-albedo deposits on cratered volcanic plains (unit pc in Fig. 1) are fairly consistent; the major heterogeneities appear in the interior deposits of Valles Marineris, which are quite different in age and geologic setting (S. L. Murchie, personal communication 1992).

Taken together, spatially resolved spectroscopic observations of low-albedo regions on Mars indicate regional variability in pyroxene (and therefore basalt), but within a limited compositional range high in iron and low-to-medium in calcium. The detection of other basaltic mineral phases is more problematic. Magnetite and/or other opaque oxide phases are inferred to be present because of the overall low albedo of the observed materials. If plagioclase is abundant in a basalt and contains at least trace Fe^{2+}, it is often apparent in reflectance spectra as a weak band or shoulder near 1.2 to 1.3 (see, e.g., Adams 1974). Such a feature has not been seen in spectra of Mars. While low plagioclase content is consistent with the measured low Al in Martian soils (B. C. Clark et al. 1982), that mineral's weak spectral feature could be masked by effects of weathered dust and/or other minerals. Because of the potential for identifying degree of silica polymerization from data in the 9 to 11-μm region, thermal infrared spectra from such instruments as the Mars Observer Thermal Emission Spectrometer (TES) might provide additional information about plagioclase abundance (see, e.g., Christensen et al. 1992). Fe^{2+}-bearing olivine, which has a broad crystal-field band complex centered just longward of 1 μm, can often be detected in a basalt because it broadens the 1-μm pyroxene band, adds a shoulder near 1.3-μm, and increase the overall near-infrared spectral slope. Olivine also tends to disturb the relationship between 1- and 2-μm band positions (see, e.g., Adams 1974,1975; Singer 1981) as discussed further below. The occurrence of olivine at the surface of Mars is possibly consistent with some spectra but has not been unambiguously determined. Lastly, mafic glass has near-infrared Fe^{2+} crystal-field bands roughly analogous to pyroxene bands, but which occur in somewhat different locations and are quite weak and broad due to the high degree of structural disorder (see, e.g., Adams 1974,1975). A relatively pure mafic glass deposit on Mars should be identifiable spectroscopically, but in the presence of crystalline pyroxene the glass bands tend to be masked.

III. SURFACE EXPOSURES OF UNALTERED IGNEOUS MATERIALS

A very basic question when discussing the measurement and utilization of unaltered igneous material at the surface Mars of is, "where is it, and how

close to the surface?" While fine-grained weathered dust and soil is nearly ubiquitous on the planet, there are many lines of evidence that crystalline rock is readily accessible. As discussed above, near-infrared spectra of most low-albedo regions indicate the presence of Fe^{2+} in pyroxenes associated with unaltered igneous rock. For these absorptions to be detectable, the crystalline pyroxene must be either freshly exposed, or coated by no more than $\sim 10\ \mu$m of weathered dust (Singer and Roush 1983). It has been observed telescopically for hundreds of years that while regional and global dust storms are relatively common on Mars, the characteristic low-albedo features always return to view after some period of time. Since 1968, pyroxene bands have consistently been observed in spectra of these regions (see Singer et al. 1979, for a historical review). This demonstrates that natural mechanisms must exist and operate for re-exposing crustal materials after they have been subjected to dust fallout. The aerosol dust is very fine-grained ($<5\ \mu$m; Pollack et al. 1979) and presumably adheres loosely to rock surfaces by static charges. Nevertheless there is apparently no permanent bonding mechanism currently operating, or at least such new bonding is balanced by erosion of fresh surfaces. The evidence also implies that rates of weathering of crystalline mafic rock are extremely low under present surface conditions (cf. Burns 1992).

The fact that crystalline pyroxenes (presumably derived from crystalline mafic rocks) are observed for low-albedo regions does not necessarily demonstrate that the surfaces of these regions consist primarily of exposed bedrock. While the systematics have not been fully studied, there are differences in spectral reflectance between a basaltic rock and a powder derived by crushing that rock (see, e.g., Adams and Felice 1967; Singer and Blake 1983). The pyroxene bands are deeper in basalt powders; in a whole rock the 2-μm band in particular is often difficult to detect; and the spectral continuum slope in the near-infrared tends to be more negative for whole rocks. Based on these criteria, it appears that much or most of the low-albedo material observed on the surface of Mars consists of particulate soils derived from igneous rock. (Some fraction of rocks, blocks, and bedrock could also be exposed in these regions, but would require other techniques for their detection.) There are some locations, however, where substantial bedrock is almost certainly exposed, based primarily on morphology viewed in high-resolution orbital images. An example is the chasma walls of Valles Marineris (see, e.g., Geissler et al. 1990; Murchie et al. 1992).

Our best close-up views of the surface of Mars were provided by the two Viking Landers. The two landing sites are similar, both in chemistry and geology; both are "intermediate" between low- and high-albedo regions. The Landers returned monochrome and multispectral images which provide some indirect evidence for mineralogic composition. Rocks of all sizes from pebbles to boulders are exposed at the landing sites, although most of the surface area is covered by soils. By visual appearance and color properties, most of the rocks are volcanic in origin and covered or coated to some thickness by weathered soil (Guinness et al. 1987). The Lander cameras were

unfortunately not capable of detecting pyroxene absorptions; nevertheless, some rock surfaces and soil exposures are quite dark and much less red (oxidized) than typical heavily weathered soil. The darker, "bluer" rock exposures are interpreted as being consistent with basalt, perhaps with a very thin, semi-transparent coating of weathered dust (Guiness et al. 1987). The variety of color and texture of soil exposures might be due to a gradation from drifts composed entirely of fine-grained heavily weathered dust, to darker, less red soils containing a substantial fraction of unaltered mafic rock and mineral fragments.

Lastly, dark aeolian dunes are quite common on the Martian surface, imaged extensively by Mariner 9 and Viking. Based on modeling of the aeolian environment of Mars, as well as thermal inertia results, these dunes consist of fine-grained sand on the order of 125 to 250 μm grain size (Christensen and Kieffer 1979; Christensen 1982). Orbital multispectral imaging has shown these dune deposits to be the darkest and least red materials exposed on the planet's surface (see, e.g., Singer et al. 1984; Geissler et al. 1990). Active saltation of these sand grains is a likely mechanism for efficiently clearing these deposits of weathered dust fallout. The color and albedo properties of these dunes are most consistent with an Fe^{2+}-rich composition such as crystalline basalt sand or basaltic glass (Geissler et al. 1990). Unfortunately, dune deposits are too small to be spatially resolved in existing Earthbased telescopic visible and near-infrared spectroscopic measurements, so we do not yet know if they generally contain crystalline material. At least one location of dark dune material in Juventae Chasma was resolved by the ISM instrument, and pyroxene bands are apparent (S. L. Murchie, personal communication 1991). There is also evidence that some dune materials originated from thick layers of mechanically weak material exposed in the walls of Valles Marineris; this implies basaltic ash rather than hard crystalline basalt as a source (Geissler et al. 1990). Independent of their degree of crystallinity, however, dunes contain perhaps the cleanest and least altered surface exposures of mafic crustal material(s) on Mars.

IV. SNC METEORITES AS SAMPLES OF THE MARTIAN CRUST

A small and diverse group of igneous meteorites, collectively called the "SNC" meteorites (for shergottites, nakhlites, and chassignites), are now generally thought to be Martian samples. All of these meteorites have apparent crystallization ages of 1300 Myr or younger (see, e.g., Nakamura et al. 1982*a*; Nyquist et al. 1984; Jagoutz and Wanke 1986; Jones 1986). These late crystallization ages first prompted the suggestion that SNC meteorites might be from Mars (McSween et al. 1979; Walker et al. 1979; Wasson and Wetherill 1979), because of the difficulty of sustaining igneous activity on an asteroid-sized body for such a long a time. The complex, multi-stage fractionations required to explain the trace element and isotopic characteristics of these meteorites are consistent with a Mars-sized parent body (Shih et al. 1982; Smith et al. 1984;

McKay et al. 1986; Jones 1989; Longhi 1991), as is the calculated size of the gravitational field necessary to effect crystal accumulation in shergottites (Grimm and McSween 1982). A direct link with Mars is provided by the discovery in SNC meteorites of trapped gases with relative elemental abundances and isotopic compositions that are virtually indistinguishable from those of the Martian atmosphere, as measured by Viking spacecraft (Bogard et al. 1984; Becker and Pepin 1984; Carr et al. 1985).

The SNC meteorites have petrographic features indicative of crystallization at shallow crustal levels or, in some cases, as volcanic flows (McSween 1985; Johnson et al. 1991). Moreover, impact models for liberating rocks from the Martian surface (Melosh 1984; Vickery and Melosh 1987) are most effective in ejecting surficial spalls. SNC meteorites may thus provide direct sampling of the Martian surface.

Literature published prior to 1985 on the mineralogy, petrology, and chemistry of the SNC meteorites have been reviewed comprehensively by McSween (1985), with more recent data provided by Laul et al. (1986), Stoffler et al. (1986), Wright et al. (1988), Treiman (1986,1990), Lundberg et al. (1988,1990), Jagoutz (1989), and Johnson et al. (1991). Brief petrographic summaries are given below; details of mineral compositions and modal proportions can be found in McSween (1985).

A. Shergottites

The shergottites (Shergotty, Zagami, EETA79001) are basaltic rocks, consisting primarily of pigeonite, augite, and maskelynite (diaplectic glass formed from plagioclase), with minor magnetite, ilmenite, pyrrhotite, and phosphates. Preferred orientations of mineral grains and the high proportion of pyroxenes (almost 70 volume %) suggest that these meteorites were cumulates or, perhaps more accurately, phenocryst-enriched flows. Melting experiments (Stolper and McSween 1979) suggest that the cores of pyroxenes are cumulus phases, and the iron-enriched pyroxene rims as well as other minerals grew from the ambient liquid. The compositions of coexisting iron-titanium oxides indicate that redox conditions were similar to terrestrial basalts (Stolper and McSween 1979), and the presence of hydrous amphibole in trapped magmatic inclusions (Treiman 1985) indicates that the magma contained water. Although amphibole stability requires modest pressure (>1 kbar), the melt inclusions in which it occurs are in cumulus pigeonite cores which could have formed in a subsurface magma chamber and been transported to the surface during eruption.

The EETA79001 meteorite has an igneous contact between two shergottite lithologies with different mineral proportions and grain sizes, probably representing successive volcanic flows (Steele and Smith 1982; McSween and Jarosewich 1983). One of the lithologies also contains partly resorbed xenocrysts of olivine and orthopyroxene, the vestiges of a disaggregated plutonic rock.

ALHA77005 is a feldspathic harzburgite closely allied with the shergot-

tites. It consists of olivine and orthopyroxene, with subordinate maskelynite, chromite, ilmenite, troilite, and phosphates, and displays cumulate texture. The mineralogy is very similar to the xenocryst assemblage in EETA79001.

The shergottites and ALHA77005 are heavily shocked, resulting in conversion of plagioclase to maskelynite, formation of pockets and veins of impact melt, and other shock features (McSween 1985; Stoffler et al. 1986). Shock metamorphism has apparently affected some radiogenic isotope systems (Shih et al. 1982; Chen and Wasserburg 1986) and contributed to disputes concerning the crystallization ages of these meteorites, which range from 1300 Myr (Nyquist et al. 1984), 360 Myr (Jagoutz and Wanke 1986), to 180 Myr (Jones 1986). Martian weathering has also apparently altered shergottites to some degree. Secondary carbonates and sulfates with unusual stable isotopic compositions in EETA79001 are thought to have formed on the parent planet (Gooding et al. 1988; Wright et al. 1988).

Strontium and neodymium isotope systematics indicate that these meteorites fall in two genetic groups: Shergotty–Zagami, and the two Antarctic meteorites. The two groups cannot have been derived from the same parental magma or source region without some assimilation of a crustal contaminant by Shergotty and Zagami.

B. Nakhlites and Chassigny

The nakhlites (Nakhla, Governador Valadares, Lafayette) are pyroxenites, consisting mostly of augite, with lesser olivine and minor orthopyroxene, plagioclase, alkali feldspar, magnetite, silica, sulfides, and phosphates. Augite crystals have preferred orientations, and the cores of these grains are interpreted as cumulus phases. The unique meteorite Chassigny is a dunite, with cumulus olivine, subordinate augite and orthopyroxene, and minor chromite, plagioclase, alkali feldspar, ilmenite, sulfides, and phosphates. Hydrous amphibole has also been described in Chassigny magmatic inclusions (Floran et al. 1978; Johnson et al. 1991). Shock effects in the nakhlites and Chassigny are not nearly as pronounced as in the shergottites, although Chassigny olivines do exhibit minor shock damage and contain insets of shock-melted glass.

Hydrous amphibole has also been described in Chassigny magmatic inclusions (Floran et al. 1978; Johnson et al. 1991). Based on petrographic criteria and the experimentally determined stability limits for amphibole, Chassigny has been estimated to have crystallized at a pressure between 1.5 and 5 kbar (Johnson et al. 1991). Similar features in nakhlites suggest that this pressure interval may be reasonable for them as well. Therefore, the nakhlite-chassignite association probably solidified within the Martian crust, rather than on the surface.

Nakhla, Governador Valadares, and Chassigny all have unambiguous crystallization ages of 1300 Myr (Papanastassiou and Wasserburg 1974; Nakamura et al. 1982*a*,1982*b*). The nakhlites and Chassigny lie on the same oxygen isotope mass-fractionation line as the shergottites and ALHA77005 (Clayton

and Mayeda 1983; 1986), which is normally taken as evidence that they are from the same parent body. They also share many petrologic characteristics with shergottites. There is no compelling reason to argue that all of the SNC meteorites are not from the same planet, although the trapped gases in Nakhla and Chassigny are somewhat different from shergottites (Ott 1988). The nakhlite-chassignite association does, however, constitute a third genetic group, distinct from the two groups of shergottites in its radiogenic isotopic composition and thus requiring a different parent magma. The calculated rare earth element patterns for the nakhlite-chassignite source region are complementary to that for the shergottites (Longhi 1991). Based on radiogenic isotope systematics, Jones (1989) argued that the Martian mantle was melted twice, once at 1.3 Gyr to yield nakhlite-chassignite parental magma(s), and again at 180 Myr to form shergottite parental magma(s).

V. SYNTHESIS AND DISCUSSION

A. Spectroscopic Comparison of SNC meteorites to the Surface of Mars

McFadden (1987) published visible and near-infrared laboratory reflectance spectra for powdered samples of a number of SNC meteorites. Those results are reviewed here and compared with spectroscopic observations of Mars. The two shergottites measured were Shergotty and ALHA77005. The Shergotty spectrum reproduced in Fig. 5 (McFadden 1987) is dominated by clinopyroxene, with Fe^{2+} band minima located near 0.98 μm and 2.10 μm. These results are similar to those reported for Shergotty by Feierberg and Drake (1980). These parameters plot near the center of the pyroxene band-band diagram as shown in Fig. 6, implying a relatively high-Fe, medium-Ca augite composition. Shergotty is an unequilibrated assemblage, with pyroxene crystals zoned from lower-Fe cumulate cores to high-Fe rims. This is displayed graphically in Fig. 7, which shows individual microprobe analyses of pyroxenes in that meteorite. The spectrum in Fig. 5 shows some indication of the two-pyroxene makeup of Shergotty in the breadth of both bands. The 2-μm band also shows very slight resolution of the individual pyroxene components (cf. Singer 1981). What is plotted in Fig. 6, then, is some average optical measure of pyroxene composition in this basalt. The spectrum of ALHA77005 reflects its composition of primarily olivine and orthopyroxene. The 1-μm region shows a relatively broad composite Fe^{2+} absorption of these two phases, with a minimum near 0.95 μm. Since olivine does not absorb in the 2-μm region the observed band minimum at 1.9 μm is due solely to the orthopyroxene. These positions plot above the pyroxene trend in the band-band diagram in Fig. 6, indicative of abundant olivine shifting the composite 1-μm band position to longer wavelengths than characteristic for a pure orthopyroxene.

McFadden (1987) also measured spectra for powdered samples of Chassigny and Nakhla. Chassigny, as expected from its modal composition, shows a spectral signature of nearly pure olivine: a very broad 3-component band

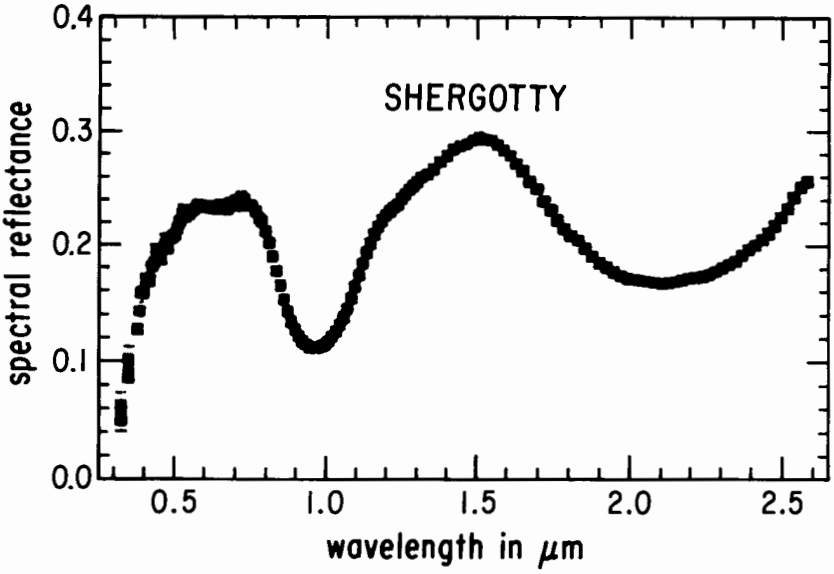


Figure 5. The spectral reflectance of a powdered sample of Shergotty meteorite, measured by McFadden (1987). The spectrum is dominated by the diagnostic spectral signature of pyroxenes: Fe^{2+} crystal-field absorptions centered near wavelengths of 1 μm and 2 μm. The subtle bump near 2.2 μm indicates that there are two compositionally different pyroxenes contributing to this spectrum, consistent with petrologic analyses of Shergotty.

with a minimum near 1.06 μm, with high reflectance beyond 1.7 μm. There is a very weak absorption centered between 1.9 and 2.0 μm, indicative of minor orthopyroxene or pigeonite. The matching pyroxene absorption expected near 0.92 μm is totally masked by the strong olivine band; for this reason Chassigny is not plotted in Fig. 6. The reflectance spectrum for Nakhla shows a combination of olivine and clinopyroxene features. The 1-μm band, an unresolved composite of olivine and pyroxene Fe^{2+} absorptions, has a minimum just longward of 1.0 μm. The 2-μm band, due strictly to clinopyroxene, has a minimum at 2.3 μm. There may also be a very weak indication of a band in the 1.9- to 2.0-μm region, which would be consistent with minor high-iron orthopyroxene or pigeonite. Nakhla plots just below the high-Ca clinopyroxene trend in Fig. 6.

A comparison of Figs. 4 and 6 shows striking agreement between the pyroxene spectral characteristics of Shergotty (McFadden 1987) and those from the telescopic observations Mars discussed above, particularly Syrtis 4/8. Band locations (not plotted here) from ISM orbital observations (Murchie et al. 1992; Mustard et al. 1992) are also in good agreement. When the Shergotty spectrum is analyzed using the calibrations of Adams (1974) and Cloutis and Gaffey (1991) the results are the same as for Syrtis 4/8, including the uniqueness problem discussed above. In the case of Shergotty we know

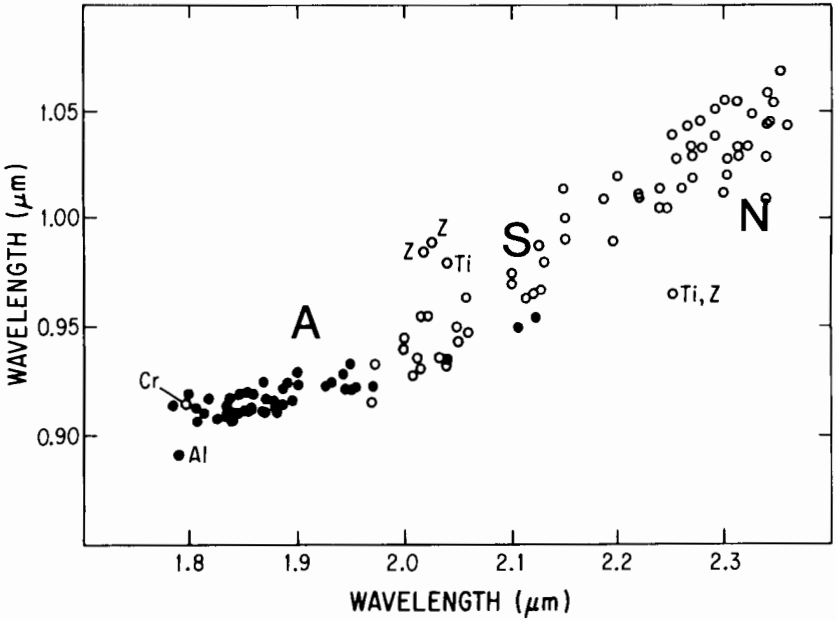


Figure 6. Pyroxene band positions for three SNC meteorites, derived from laboratory spectroscopy by McFadden (1987), are replotted on the Cloutis and Gaffey (1990) "band-band" diagram. A represents ALHA77005, S represents Shergotty, and N represents Nakhla. Shergotty is in good agreement with the Mars observations, especially Syrtis 4/8 (cf. Fig. 4). Chassigny was also measured by McFadden (1987) but is not plotted because its spectrum, like its modal composition, is dominated by olivine.

from direct examination that there are two pyroxene trends differing in Ca content, each compositionally zoned in Fe content (Fig. 7). It is unlikely that all pyroxene compositions physically present interact equally with photons, due to differences in absorption coefficient, grain size, and position within a grain (core vs rim). This complicated situation leads to the interpretation of some "optical average" composition. While the Syrtis 4/8 observation does not have enough spectral detail to indicate possible multiple and/or zoned pyroxenes, analogy to Shergotty suggests that this might be the case for at least this region of Mars.

Materials similar to ALHA77005 are not consistent with the observational data discussed above. In the 1-μm region, ALHA77005 is consistent with measurements at many other locations on Mars (Singer et al. 1990*a*, *b*), but without information about the diagnostic 2-μm region no unique identification of this orthopyroxene-olivine assemblage is possible. Nakhla is not consistent with current remote observations of the Martian surface, nor has a spectral analog of Chassigny been observed.

B. Inferences about the Martian Crust

Given the complexity of igneous processes on a large planet, it would be simplistic to think that the Martian crust consists only of rocks like the SNC

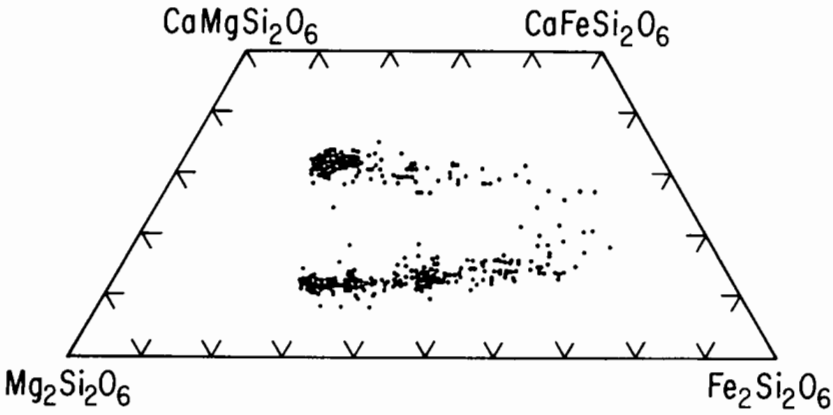


Figure 7. Individual microprobe analyses of pyroxenes from the Shergotty meteorite, plotted in the standard pyroxene ternary diagram (from the Basaltic Volcanism study; Lofgren et al. 1981). $Mg_2Si_2O_6$, the magnesium endmember, is the orthopyroxene enstatite. $Fe_2Si_2O_6$, the iron endmember, is the orthopyroxene ferrosilite. A pure calcium endmember does not exist for pyroxenes, hence the ternary diagram is truncated at 50 mole % Ca. $CaMgSi_2O_6$ is the clinopyroxene diopside, while $CaFeSi_2O_6$ is the clinopyroxene hedenbergite. The trends seen in Fe content of the plotted pyroxenes are because Shergotty is unequilibrated, consistent with rapid cooling after eruption. Both high-Ca and low-Ca pyroxenes in that meteorite are zoned, with high-Fe compositions at the rims. As discussed in the text, this can complicate spectroscopic estimates of composition.

meteorites. Nevertheless, the many lines of evidence for a Mars crustal origin for the SNCs are compelling. The strong spectroscopic similarity between Shergotty and various near-equatorial dark regions on Mars is of particular interest, as this meteorite is inferred from petrologic evidence to have formed as a surface flow or near-surface cumulate. Coverage of the surface of Mars by suitable spectral data is far from complete, but the data in hand imply that basalts varying around a Shergotty-like composition might be commonly exposed at the surface.

Chemical lines of evidence are also strong. There is a striking similarity in the bulk chemical compositions of shergottites and Martian duracrust-free soil at both Viking landing sites (Baird and Clark 1981; B. C. Clark 1983). Furthermore, these meteorites contain magnetite in approximately the same modal abundance as the magnetic oxide component of Martian soil (McSween 1985). Taken at face value, these observations may also suggest that shergottite-like lavas may be a dominant constituent of the surface of Mars, if the XRF-measured soils derived directly from volcanic bedrock. This seems unlikely, however. The Viking Lander XRF measured fine-grained soils exclusively, with no sampling of rocky material. These fine-grained soils are quite bright and "red" in color. There is general consensus that they are heavily weathered (oxidized and hydrated), compositionally and genetically

related to the heavily weathered soils common elsewhere and possibly to the globally distributed aeolian dust (see, e.g., McCord et al. 1977; Guinness et al. 1987; Arvidson et al. 1989). For the weathered soils to reflect such a high degree of chemical similarity to a shergottite-like source rock would require a nearly isochemical process, consistent with a very limited role by liquid water. A less likely alternative is that the soils derived from fractional weathering of a basaltic source[a] with greater plagioclase (higher Al and K) than the shergottites (see, e.g., B. C. Clark 1983). This would require efficient segregation of these residual elements elsewhere on the planet, where they have been detected neither directly nor indirectly.

Given that nakhlite-chassignite cumulates formed within the crust rather than on the surface, they are less likely to be observed on the surface of Mars except possibly in localized areas where the uppermost crust has been breached by impacts or severely denuded by erosion. This is consistent with their apparent lack of detection to date. Orbital near-infrared imaging-spectrometer observations with high spatial resolution ($\sim$1 km/pixel) are necessary to detect any such exposures.

Because the SNC meteorites are cumulates or at least contain cumulus minerals, there have been a number of attempts to estimate the compositions of the parental magmas from which they formed. Various methods employing melting experiments (Stolper and McSween 1979), calculations based on estimates of the proportions of intercumulus liquids (McSween and Jarosewich 1983; Treiman 1986) or parameterization of phase boundaries (Longhi and Pan 1989), and analysis of the compositions of trapped magmatic inclusions (Johnson et al. 1991) have been tried. The results converge on SNC parental magma compositions that are similar in many respects to terrestrial boninite lavas, rich in iron, low in aluminum, and widely varying in calcium. Fractionation of such magmas at low pressure is expected to produce ultramafic cumulates, all of which should be relatively iron-rich and have low feldspar contents. The considerable variations in calcium contents of these magmas lead to different crystallization sequences, permitting various combinations of olivine, augite, and pigeonite (or orthopyroxene) as cumulus phases. The liquids complementary to these cumulates would solidify into iron-rich basalts. Extrusions of these magmas would have low viscosities unless choked with phenocrysts of olivine, augite, or pigeonite (feldspar is unlikely because of the low aluminum content), and vapor bubbles due to concentration of water during crystallization might occur. Thus we can extrapolate from what we have learned about SNC meteorites to predict the general characteristics of the surficial igneous crust on its parent body.

C. Possible Relationship to Ancient and/or Silicic Crust

Volcanism on Mars continued throughout most of the planet's history. The old-

[a] From investigation of alteration kinetics and analogy to terrestrial palagonites, it is generally thought that weathered soils on Mars derived primarily from mafic glass rather than crystalline rock (see, e.g., Allen et al. 1981; Singer 1982)

est remaining crust and volcanic flows are preserved in the ancient highlands, dating back to ~3.9 Gyr. (Carr 1981). The SNC meteorites are comparatively quite young (1.3 to 0.2 Gyr), and therefore tell us about igneous crust much later in the planet's history. The SNCs are not primitive igneous materials; they were generated by a complex sequence of geologic processes. What if any inferences can we draw about the composition of ancient Martian crust from available evidence?

At present there is no evidence for highly differentiated, intermediate-to-acidic crustal rocks on Mars. More conclusive determination of possible intermediate or silicic crustal components requires orbital mapping by gamma-ray and thermal-infrared spectrometers, planned for both the Mars Observer and Mars 94 missions. A number of lines of evidence though, appear to put significant constraints on possible silicic crustal components. First, the SNC magmas were impoverished in Al, and feldspar was a late-crystallizing phase (see, e.g., Stolper and McSween 1979). If this reflects fundamental characteristics of Martian mantle source regions, as many think it does, it seems highly unlikely that Mars ever had ancient crustal components which were anorthositic (such as that on the Moon) or granitic (such as found on the Earth). On Earth, basalt is the dominant type of igneous rock supplied from the mantle. The more highly differentiated Si-rich continental crust is produced by reworking of igneous crust in subduction zones at plate margins, in the presence of water. On Mars, there is no evidence that plate tectonics ever operated, consistent with the apparent absence of Si-rich crust. Second, the overall composition of weathered soils at both Viking landing sites suggest that these soils derived from mafic igneous materials, based largely on the combination of low Al and low K (Toulmin et al. 1977). Those researchers estimated that as little as 10% granitic surface rock would yield K abundances in the weathered soils several times higher than allowed by the Viking XRF measurements. The third line of evidence is spectroscopic. The majority of observations of pyroxenes on Mars (consistent with shergottite-like basaltic compositions) have been in low-albedo regions located in the ancient highlands (cu in Fig. 1), the oldest crust exposed on Mars. It seems probable that some (if not most) of these low-albedo surface exposures consist of locally or regionally derived materials. If that is true, we are spectroscopically sampling rather ancient basaltic crust, and constraining it to be compositionally similar to the more recent basaltic crust sampled by the Shergotty meteorite.

VI. SUMMARY

Three conclusions can be drawn. First, there is abundant, unaltered basaltic crustal material exposed at or very near the present Martian surface over much of the planet, concentrated in low-albedo regions. The physical forms include bedrock, blocks ranging in size from boulders to pebbles, clean dune sand, and soils which include varying proportions of unaltered minerals and rock fragments mixed with weathered materials. The presence of fine-grained

weathered soils and dust on Mars complicates interpretation of remote observations, but does not totally hide crustal material. Crystalline basaltic rock has been directly identified over much of Mars. There is also possible occurrence of unaltered basaltic glass (sideromelane and/or tachylite).

Second, spectroscopically, Shergotty (and possibly other shergottites) are consistent in important mineralogic features with telescopic spectra of low-albedo regions on Mars. While this does not in itself prove that shergottites originated on Mars, it does demonstrate that they are similar in pyroxene mineralogy and chemistry, key petrologic parameters for basalts. It therefore provides yet another piece of evidence, added to the already numerous and compelling arguments, that shergottites do in fact represent Martian crustal rocks. Because shergottites are interpreted to have formed as basalt flows or shallow cumulates, it is reasonable that they could be observed on the surface of their parent planet. Spectral analogs to the other two SNC meteorite families, the nahklites and chassignites, have not been detected at the Martian surface. This is not inconsistent with their petrologic interpretation as higher-pressure cumulate assemblages which solidified within the crust.

A third conclusion, based on the lines of evidence discussed here, is that the crust of Mars is apparently dominantly basaltic, high in Fe, low in Al, and moderate in Ca. While the SNCs sampled relatively young igneous crust, there is evidence that the ancient crust of Mars is compositionally similar. Ultramafic igneous rocks with more iron- and magnesium-rich compositions may occur at some locations on Mars, but such compositions are neither required by nor consistent with the bulk of observational evidence. There is no compelling evidence for highly differentiated intermediate or acidic crustal rocks on Mars, and significant constraints on their possible abundance.

Acknowledgments. We thank T. Roush and S. Murchie for their detailed and helpful reviews of this manuscript.

REFERENCES

Adams, J. B. 1974. Visible and near-infrared diffuse reflectance spectr of pyroxenes as applied to remote sensing of solid objects in the solar system. *J. Geophys. Res.* 79:4829–4836.

Adams, J. B. 1975. Interpretation of visible and near-infrared diffuse reflectance spectra of pyroxenes and other rock-forming minerals. In *Infrared and Raman Spectroscopy of Lunar and Terrestrial Minerals*, ed. C. Karr (San Diego: Academic Press), pp. 91–116.

Adams, J. B., and Felice, A. L. 1967 Spectral reflectance 0.4 to 2.0 microns of silicate rock powders. *J. Geophys. Res.* 72:5705–5715.

Adams, J. B., and McCord, T. B. 1969 Mars: Interpretation of spectral reflectivity of light and dark regions. *J. Geophys. Res.* 74:4851–4856.

Allen, C. C., Gooding, J. L., Jercenovic, M., and Keil, K. 1981 Altered basaltic glass: A terrestrial analog to the soil of Mars. *Icarus* 45:347–369.

Aronson, J. R., and Emslie, A. G. 1975. Composition of the martian dust as derived by infrared spectroscopy from Mariner 9. *J. Geophys. Res.* 80:4925-4931.

Arvidson, R., Guiness, E., Dale-Bannister, M., Adams, J., Smith, M., Christensen, P., and Singer R. 1989. Nature and distribution of surficial deposits in Chryse Planitia and vicinity. *J. Geophys. Res.* 94:1473–1587.

Baird, K. K., and Clark, B. C. 1981. On the original igneous source of Martian fines. *Icarus* 45:113–123.

Becker, R. H., and Pepin, R. O. 1984. The case for a Martian origin for the shergottites: Nitrogen and noble gases in EETA79001. *Earth Planet. Sci. Lett.* 69:225–242.

Bell, J. F., McCord, T. B., and Owensby, P. D. 1990. Observational evidence of crystalline iron oxides on Mars. *J. Geophys. Res.* 95:14447–14462.

Bibring, J.-P., Combes, M., Langevin, Y., Soufflot, A., Cara, C., Drossart, P., Encrenaz, Th., Erard, S., Forni, O., Gondet, B., Ksanfomality, L., Lellouch, E., Masson, Ph., Moroz, V., Rocard, F., Rosenqvist, J., and Sotin, C. 1989. Results from the ISM experiment. *Nature* 341:591–592.

Bogard, D. D., Nyquist, L. E., and Johnson, P. 1984. Noble gas contents of shergottites and implications for the Martian origin of SNC meteorites. *Geochim. Cosmochim. Acta* 48:1723–1739.

Burns, R. G. 1970 *Mineralogical Applications of Crystal Field Theory* (New York: Cambridge Univ. Press).

Burns, R. G. 1992. Rates of oxidation weathering on the surface of Mars. In *Proc. of Martian Surface and Atmosphere Through Time (MSATT) Workshop*, eds. R. M. Haberle and B. M. Jakosky, LPI Tech. Rept. 92-02, pp. 26–27 (abstract).

Carr, M. H. 1981. *The Surface of Mars* (New Haven: Yale Univ. Press).

Carr, R. H., Grady, M. M., Wright, I. P., and Pillinger, C. T. 1985. Martian atmospheric weathering products in SNC meteorites. *Nature* 314:248–250.

Chen, J. H., and Wasserburg, G. J. 1986. Formation ages and evolution of Shergotty and its parent planet from U-Th-Pb systematics. *Geochim. Cosmochim. Acta* 50:955–968.

Christensen, P. R. 1982 Kasei Valles revisited. *Lunar and Planet. Sci.* XIII:96-97 (abstract).

Christensen, P. R., and Kieffer, H. 1979. Moderate resolution thermal mapping of Mars: The channel terrain around the Chryse Basin. *J. Geophys. Res.* 84:8233–8238.

Christensen, P. R., Anderson, D. L., Stillman, C. C., Clark, R. N., Kieffer, H. H., Malin, M. C., Pearl, J. C., Carpenter, J., Bandiera, N., Brown, F. G., and Silverman, S. (1992) Thermal emission spectrometer experiment: Mars Observer Mission. *J. Geophys. Res.* 97:7719–7734.

Clark, B. C. 1983. Correspondence of shergottites and Martian fines. *Lunar Planet. Sci.* XIV:117–118 (abstract).

Clark, B. C., Baird, A. K., Weldon, R. J., Tsusaki, D. M., Schnabel, L., and Candelaria, M. P. 1982. Chemical composition of martian fines. *J. Geophys. Res.* 87:10059–10067.

Clark, R. N., Swayze, G. A., Singer, R. B., and Pollack, J. 1990. High resolution spectra of Mars in the 2.3 μm region: Evidence for mineral scapolite. *J. Geophys. Res.* 95:14463–14479.

Clayton, R. N., and Mayeda, T. K. 1983 Oxygen isotopes in eucrites, shergottites, nakhlites, and chassignites. *Earth Planet. Sci. Lett.* 62:1–6.

Clayton R. N., and Mayeda, T. K. 1986 Oxygen isotopes in Shergotty. *Geochim.*

Cosmochim. Acta 50:979–982.

Cloutis, E. A., and Gaffey, M. J. 1991. Pyroxene spectroscopy revisited: Spectral-compositional correlations and relationships to geothermometry. *J. Geophys. Res.* 96:22809–22826.

Erard, S., Bibring, J.-P., Mustard, J., Formi, O., Head, J. W., Hurtrez, S., Langevin, Y., Pieters, C. M., Rosenqvist, J., and Sotlin, C. 1990. Spatial variations in composition of the Valles Marineris and Isidis Planitia regions of Mars derived from ISM data. *Proc. Lunar Planet. Sci.* 21:437–455.

Feierberg, M. A., and Drake, M. J. 1980. The meteorite-asteroid connection: The infrared spectra of Eucrites, Shergottites, and Vesta. *Science* 209:805–807.

Floran, R. J., Prinz, M., Hlava, P. F., Keil, K., Nehru, C. E. and Hinthorne, J. R. 1978. The Chassigny meteorite: A cumulate dunite with hydrous amphibole-bearing melt inclusions. *Geochim. Cosmochim. Acta* 42:1213–1229.

Geissler, P. E., Singer, R. B., and Lucchitta, B. K. 1990. Dark materials in Valles Marineris: Indications of the style of volcanism and magmatism on Mars. *J. Geophys. Res.* 95:14399–14413.

Gooding, J. L., Wentworth, S. J., and Zolensky, M. E. 1988. Calcium carbonate and sulfate of possible extraterrestrial origin in the EETA 79001 meteorite. *Geochim. Cosmochim. Acta* 52:909–915.

Greeley, R. 1987. Photogeological inferences of martian surface composition. In *MEVTV Workshop on Nature and Composition of Surface Units on Mars*, LPI Tech. Report No. 88-05, pp. 67–68.

Greeley, R., and Spudis, P. D. 1981. Volcanism on Mars. *Rev. Geophys. Space Phys.* 19:13–41.

Grimm, R. E., and McSween, H. Y. 1982. Numerical simulation of crystal fractionation in shergottite meteorites. *Proc. Lunar Planet. Sci. Conf.* 13:385–392.

Guinness, E., Arvidson, R., Dale-Bannister, M., Singer, R. B., and Bruckenthal, E. A. 1987. On the spectral reflectance properties of materials exposed at the Viking Landing sites. *Proc. Lunar Planet. Sci. Conf.*, 17, *J. Geophys. Res. Suppl.* 92:E575–E587.

Hanel, R., Conrath, B., Hovis, W., Kunde, V., Loman, P., Maguire, W., Pearl, J., Pirraglia, J., Prabhadara, C., Schlachman, B., Levin, G., Straat, P., and Burke, T. 1972. Investigation of the martian environment by infrared spectroscopy on Mariner. *Icarus* 17:423–442.

Huguenin, R. L., Head, J. W., and McGetchin, T. R. 1978. Mars: Petrologic units in the Margaritifer Sinus and Coprates Quadrangle. In *Reports of Planetary Geology Program—1977–1978*, NASA TM-79729.

Hunt, G. R., Logan, L. M., and Salisbury, J. W. 1973. Mars: Components of infrared spectra and composition of the dust cloud. *Icarus* 18:459–469.

Jagoutz, E. 1989. Sr and Nd isotopic systematics in ALHA 77005: Age of shock metamorphism in shergottites and magmatic differentiation on Mars. *Geochim. Cosmochim. Acta* 53:2429–2441.

Jagoutz, E., and Wänke, H. 1986. Sr and Nd isotopic systematics of Shergotty meteorite. *Geochim. Cosmochim. Acta* 50:939–953.

Johnson, M. C., Rutherford, M. J., and Hess, P. C. 1991. Chassigny petrogenesis: Melt compositions, intensive parameters, and water contents of Martian(?) magmas. *Geochim. Cosmochim. Acta* 55:349–366.

Jones, J. H. 1986. A discussion of isotopic systematics and mineral zoning in the shergottites: Evidence for a 180 m.y. igneous crystallization age. *Geochim. Cosmochim. Acta* 50:969–977.

Jones, J. H. 1989. Isotopic relationships among the shergottites, the nakhlites and Chassigny. *Proc. Lunar Planet. Sci. Conf.* 4:465–474.

Laul, J. C., Smith, M. R., Wänke, H., Jagoutz, E., Dreibus, G., Palme, H., Spettel,

B., Burghele, A., Lipschutz, M. E., and Verkouteren, R. M. 1986. Chemical systematics of the Shergotty meteorite and the composition of its parent body (Mars). *Geochim. Cosmochim. Acta* 50:909–926.

Lofgren, G. E., Bence, A., Duke, M., Dungan, M., Green, J., Haggerty, S., Haskin, L., Irving, A., Lipman, P., Naldrett, A., Papike, J., Reid, A., Rhodes, J., Taylor, S., and Vaniman, D. 1981. Petrology and chemistry of terrestrial, lunar and meteoritic basalts. In *Basaltic Volcanism on the Terrestrial Planets* (Houston: Lunar and Planetary Inst.).

Longhi, J. 1991. Complex magmatic processes on Mars: Inferences from the SNC meteorites. *Proc. Lunar Planet. Sci. Conf.* 21:695–709.

Longhi, J., and Pan, V. 1989. The parent magmas of the SNC meteorites. *Proc. Lunar Planet. Sci. Conf.* 19:451–464.

Lucchitta, B. K. 1987. Recent mafic volcanism on Mars. *Science* 235:565–567.

Lucchitta, B. K., Clow, G. D., Geissler, P. E., McEwen, A. S., Schultz, R. A., Singer, R. B., and Squyres, S. W. 1992. The canyon system on Mars. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 453–492.

Lundberg, L. L., Crozaz, G., McKay, G. A., and Zinner, E. 1988. Rare earth element carriers in and chronology of the Shergotty meteorite. *Geochim. Cosmochim. Acta* 52:2147–2163.

Lundberg, L. L., Crozaz, G., and McSween, H. Y. 1990. Rare earth elements in minerals of the ALHA77005 shergottite and implications for its parent magma and crystallization history. *Geochim. Cosmochim. Acta* 54:2535–2547.

McCord, T. B., and Westphal, J. A. 1971. Mars: Narrowband photometry from 0.3 to 2.5 μm, of surface regions during the 1969 apparition. *Astrophys. J.* 168:141–153.

McCord, T. B., Huguenin, R. L., Mink, D., and Pieters, C. 1977. Spectral reflectance of Martian areas during the 1973 opposition: Photo-electric filter photometry 0.33–1.10 μm. *Icarus* 31:25–39.

McCord, T. B., Singer, R. B., Hawke, B., Adams, J., Evans, D., Head, J., Mouginis-Mark, P., Pieters, C., Huguenin, R., and Zisk, S. 1982*a*. Mars: Definition and characterization of global surface units with emphasis on composition. *J. Geophys. Res.* 87:10129–10148.

McCord, T. B., Clark, R. N., and Singer, R. B. 1982*b*. Near-IR reflectance spectra of surface regions and compositional implications. *J. Geophys. Res.* 87:3021–3032.

McFadden, L. A. 1987. Spectral reflectance of SNC meteorites: Relationships to martian surface composition. In *MEVTV Workshop on Nature and Composition of Surface Units on Mars*, LPI Tech. Rept. 88-05, pp. 88–90.

McKay, G. A., Wagstaff, J., and Yand, S.-R. 1986. Clinopyroxene REE distribution coefficients for shergottites: The REE content of the Shergotty melt. *Geochim. Cosmochim. Acta* 50:927–937.

McSween, H. Y. 1985. SNC meteorites: Clues to Martian petrologic evolution? *Rev. Geophys.* 23:391–416.

McSween, H. Y., and Jarosewich, E. 1983. Petrogenesis of the Elephant Moraine A79001 meteorite: Multiple magma pulses on the shergottite parent body. *Geochim. Cosmochim. Acta* 47:1501–1513.

McSween, H. Y., Stolper, E. M., Taylor, L. A., Muntean, R. A., O'Kelley, G. D., Eldridge, J. S., Biswas, S., Ngo, H. T., and Lipschutz, M. E. 1979. Petrogenetic relationship between Allan Hills 77005 and other achondrites. *Earth Planet. Sci. Lett.* 45:275–284.

Melosh, H. J. 1984. Impact ejection, spallation, and the origin of meteorites. *Icarus* 59:234–260.

Murchie, S. L., Erard, S., Mustard, J., Bibring, J., Langevin, Y., Head, J., and Pieters,

C. 1991. Spectral properties of interior deposits of Valles Marineris from ISM imaging spectroscopy. *Lunar Planet. Sci.* XXII:945–946 (abstract).

Murchie, S. L., Erard, S., Mustard, J. F., Bibring J.-P., Langevin Y., Head, J. W., and Pieters, C. M. 1992. The geology of the interior deposits of Valles Marineris from Viking images and ISM imaging spectroscopy. *Lunar Planet. Sci.* XXIII:945–946 (abstract).

Mustard, J. F., Erard, S., Bibring, J. P., Langevin, Y., Head, J. W., and Pieters, C. M. 1992. Pyroxene chemistry of the Syrtis Major Volcanic Plateau. *Lunar Planet. Sci.* XXIII:955–956 (abstract).

Mutch, T. A., Arvidson, R. E., Head, J. W., Jones, K. L., and Saunders, R. S. 1976. *Geology of Mars* (Princeton, N. J.: Princeton Univ. Press).

Nakamura, N., Unruh, D. M., Tatsumoto, M., and Hutchison, R. 1982*a*. Origin and evolution of the Nakhla meteorite inferred from the SmNd and U-Pb systematics and REE, Ba, Sr, Rb abundances. *Geochim. Cosmochim. Acta* 46:1555–1573.

Nakamura, N., Komi, H., and Kagami, H. 1982*b*. Rb-Sr isotopic and REE abundances in the Chassigny meteorite. *Meteoritics* 17:257–258. Nyquist, L. E., Wooden, J., Bansal, B., Weismann, H., and Shih, C.-Y. 1984. Sr and Nd isotopic systematics of EETA79001. *Meteoritics* 19:284 (abstract).

Ott, U. 1988. Noble gases in SNC meteorites: Shergotty, Nakhla, Chassigny. *Geochim. Cosmochim. Acta* 52:1937–1948.

Papanastassiou, D. A., and Wasserburg, G. J. 1974. Evidence for a late formation and young metamorphism in the achondrite Nakhla. *Geophys. Res. Lett.* 1:23–26.

Phillips, R. J. 1978. *Report on the Tharsis Workshop*, NASA TM-79729, pp. 334–336.

Pollack, J. B., Colburn, D. S., Flaser, M., Kahn, R., Carlston, C. E., and Pidek, D. 1979. Properties and effects of dust particles suspended in the martian atmosphere. *J. Geophys. Res.* 84:2929–2945.

Settle, M. 1979. Formation and deposit of volcanic sulfate aerosols on Mars. *J. Geophys. Res.* 84:8343–8354.

Shih, C.-Y., Nyquist, L. E., Bogard, D. D., McKay, G. A., Wooden, J. L., Bansal, B. M., and Weismann, H. 1982. Chronology and petrogenesis of young achondrites, Shergotty, Zagami, ALHA77005: Late magmatism on a geologically active planet. *Geochim. Cosmochim. Acta* 46:2323–2344.

Singer, R. B. 1980. The dark materials on Mars: II. New mineralogic interpretation from reflectance spectroscopy and petrologic implications. *Lunar Planet. Sci.* XI:1048–1050 (abstract).

Singer, R. B. 1981. Near-infrared spectral reflectance of mineral mixtures; systematic combinations of pyroxenes, olivine, and iron oxides. *J. Geophys. Res.* 86:7967–7982.

Singer, R. B. 1982. Spectral evidence for the mineralogy of high albedo soils and dust on Mars. *J. Geophys. Res.* 87:10159–10168.

Singer, R. B. 1985. Spectroscopic observations of Mars. *Adv. Space Res.* 5:59–68.

Singer, R. B., and Blake, P. L. 1983. Effects of mineral grain size and physical particle size on spectral reflectance of basalts. *Lunar Planet. Sci.* XIV:706–707 (abstract).

Singer, R. B., and Roush, T. L. 1983. Spectral reflectance properties of particulate weathered coating on rocks: Laboratory modeling and applicability to Mars. *Lunar Planet. Sci.* XIV:708–709 (abstract).

Singer, R. B., and Roush, T. L. 1985. Analysis of Martian crustal petrology. *Bull. Amer. Astron. Soc.* 17:737 (abstract).

Singer, R. B., McCord, T. B., Clark, R. N., Adams, J. B., and Huguenin, R. L. 1979. Mars surface composition from reflectance spectroscopy: A summary. *J. Geophys. Res.* 84:8414–8426.

Singer, R. B., Clark, R. N., and Owensby, P. D. 1980. Mars: New regional near-

infrared spectrophotometry (0.65–2.50 μm) obtained during the 1980 apparition. *Bull. Amer. Astron. Soc.* 12:680 (abstract).

Singer, R. B., Cloutis, E., Roush, T. L., Mouginis-Mark, P. J., Hawke, B. R., and Christensen, P. R. 1984. Multispectral analysis of the Kasei Vallis-Lunae Planum Region on Mars. *Lunar Planet. Sci.* XV:74–75 (abstract).

Singer, R. B., Miller, J. S., Wells, W. K., and Bus, E. S. 1990*a*. Visible and near-IR spectral imaging of Mars during the 1988 opposition. *Lunar Planet. Sci.* XXI:1154–1155 (abstract).

Singer, R. B., Miller, J. S., and Wells, W. K. 1990*b*. Observed variation in martian crustal composition. *Bull. Amer. Astron. Soc.* 22:1061.

Smith, M. R., Laul, J. C., Ma, M.-S., Huston, T., Verkouteren, R. M., Lipschutz, M. E., and Schmitt, R. A. 1984. Petrogenesis of the SNC (shergottites, nakhlites, chassignites) meteorites: Implications for their origin from a large dynamic planet, possibly Mars. *Proc. Lunar Planet. Sci. Conf.* 15:612–630.

Soderblom, L. A., Edwards, K., Eliason, E., Sanchez, E., and Charette, M. 1978. Global color variations on the Martian surface. *Icarus* 34:446–464.

Steele, I. M., and Smith, J. V. 1982. Petrography and mineralogy of two basalts and olivine-pyroxene-spinel fragments in achondrite EETA79001. *Proc. Lunar Planet. Sci. Conf.* 13:375–384.

Stoffler, D., Ostertag, R., Jammes, C., Pfannschmidt, G., Sen Gupta, P. R., Simon, S. B., Papike, J. J., and Beauchamp, R. H. 1986. Shock metamorphism and petrography of the Shergotty achondrite. *Geochim. Cosmochim. Acta* 50:889–903.

Stolper, E. M., and McSween, H. Y. 1979. Petrology and origin of the shergottite meteorites. *Geochim. Cosmochim. Acta* 43:1475–1498.

Toon, O. B., Pollack, J. B., and Sagan, C. 1977. Physical properties of the particles composing the martian dust storm of 1971–1972. *Icarus* 30:663–696.

Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christian, R., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.

Treiman, A. H. 1985. Amphibole and hercynite spinel in Shergotty and Zagami: Magmatic water, depth of crystallization, and metasomatism. *Meteoritics* 20:229–243.

Treiman, A. H. 1986. The parental magma of the Nakhla achondrite: Ultrabasic volcanism on the shergottite parent body. *Geochim. Cosmochim. Acta* 50:1061–1070.

Treiman, A. H. 1990. Complex petrogenesis of the Nakhla (SNC) meteorite: Evidence from petrography and mineral chemistry. *Proc. Lunar Planet. Sci. Conf.* 20:273–280.

Vickery, A. M., and Melosh, H. J. 1987. The large crater origin of the SNC meteorites. *Science* 237:738–743.

Walker, D., Stolper, E. M., and Hays, J. F. 1979. Basaltic volcanism: The importance of planet size. *Proc. Lunar Planet. Sci. Conf.* 20:1995–2015.

Wasson, J. T., and Wetherill, G. W. 1979. Dynamical, chemical, and isotopic evidence regarding the formation locations of asteroids and meteorites. In *Asteroids*, ed. T. Gehrels (Tucson: Univ. of Arizona Press), pp. 926–974.

Wright, I. P., Grady, M. M., and Pillinger, C. T. 1988. Carbon, oxygen and nitrogen isotopic compositions of possible martian weathering products in EETA 79001. *Geochim. Cosmochim. Acta* 52:917–924.

WATER ON MARS: ITS HISTORY AND AVAILABILITY AS A RESOURCE

BRUCE M. JAKOSKY
University of Colorado

and

AARON P. ZENT
NASA Ames Research Center

Water will be an important resource for future missions to explore Mars. As such, some understanding is required of the abundance and distribution of water in the Martian near-surface environment. As no technique has been used to date to map near-surface water in the Martian regolith or polar regions, our understanding of where the water is today depends on interpretations of observations which relate to the global history of water over geologic time and to the seasonal cycle of water in the atmosphere; *in situ* measurements are available at only two locations. These observations and the inferences from them are summarized in this chapter, along with the current best guesses as to where water can be found and what future observations might be used to locate water. Based on our current understanding of the Martian environment, the only sure source of water is in the polar regions; water or ice in the nonpolar regolith may be unavailable for mining or may be nonexistent. Additional observations might provide a more optimistic basis for finding water.

The existence of water on Mars is undisputed today. Direct measurements have been made of water vapor in the atmosphere and in the regolith; additional remote-sensing observations suggest abundant water ice in the polar regions and possibly in the high-latitude regolith. Geomorphological features are observed at a wide variety of locations on the Martian surface which suggests a formation by liquid water flowing on the surface, either under climatic conditions similar to those occurring at present or under possibly different conditions during earlier epochs. A variety of observations also suggest that water has played an important role in the evolution of Mars throughout geologic history. Major scientific issues remain, however, and involve how much water has been outgassed to the surface since Mars' formation and its subsequent evolution.

These questions are important in the discussion of the availability of water as a resource on Mars. The use of water will presumably be required on future exploration missions, especially those involving humans. While water is clearly available at the polar regions in significant abundance, water in nonpolar regions would be much more readily accessible. The amounts of

water in the nonpolar regolith are uncertain, however; our estimates depend on the interpretation of data which pertains to the outgassing history of water and to its subsequent evolution over time as well as on inferences based on the seasonal cycling of water into and out of the atmosphere.

We will consider water which might be present in the following reservoirs: (i) polar deposits, as ice either immediately at the surface or beneath a veneer of dust in some locations; (ii) high-latitude regolith, where water ice might be stable within the top meter or so of the surface or deeper; and (iii) global regolith, where water might be present either as adsorbed water intermixed with the regolith materials or as a stable or transient ice or brine deposit. In order to determine how much water might be in these various reservoirs, we will look at the amounts of water outgassed to the surface and the subsequent loss from the surface environment by different processes; although the results are necessarily speculative, they provide a valuable constraint on the amounts of water which might be present. In addition, we can infer the existence of water in some of these locations based on our observations of the seasonal cycle of atmospheric water, where changes in the amount of water in the atmosphere suggest the reversible exchange of water with some of these non-atmospheric reservoirs.

Section I contains a discussion of the volatile inventory of Mars, with an emphasis on the amounts of water incorporated into the accreting planet and subsequently outgassed to the surface; processes by which water can be lost, such as escape of H and O to space, will also be discussed. Section II describes observations of the near-surface environment which pertain to the presence of water in the atmosphere and regolith and its accessibility. The final section, Sec. III, contains a summary of the current information on the availability of water, along with some discussion of the types of measurements which can be used to search for water or to improve our understanding of the history of water.

I. VOLATILE INVENTORY ON MARS

This section contains discussion of the amounts of volatiles, specifically water, which have been outgassed to the surface of Mars over geologic time and the subsequent evolution of the volatiles. Evidence as to the amounts of water present at the surface come from two sources: geologic inferences as to how much water was required to form geomorphologic features at the surface and geochemical inferences as to the amounts of water present in the atmosphere or outgassed to the surface. Geologic constraints will be discussed first as they form a very distinct and unique set of boundary conditions for the geochemical models; they will be presented briefly as they are discussed in more detail in the chapter by Baker et al. A detailed description of the geological inferences can also be found in Carr (1986).

A. Geological Evidence

The major geologic evidence which pertains to the abundance and distribution of water comes from observations of water-related features on the Martian surface (see, e.g., Carr 1986). Of special interest are the outflow (catastrophic flood) channels, the valley networks (runoff channels), and the so-called terrain softening at high latitudes. In addition, water can be emplaced at the surface via volcanic eruptions, which can be mapped over the surface for each of the earlier geologic epochs (Greeley 1987).

The outflow channels appear to have been formed by catastrophic eruption of subsurface liquid water to the surface (see, e.g., Baker 1982). Carr (1979,1986) estimated the total volume of water required to have carved them, making the assumptions that the water carried as much debris as was possible and that each molecule of water flowed through the system once. Under these assumptions, he estimated that a global equivalent layer of about 35 m H_2O was required; the water presumably erupted to the surface over a period of billions of years (Baker 1982). The fate of this water is uncertain, but it may have ponded in the lowlands in the northern hemisphere. If it ended up at low-enough latitudes (where the temperatures are warmer and sublimation into the atmosphere would be more rapid), then much of the water would have sublimed into the atmosphere and ultimately been cold-trapped into the polar regions; alternatively, at higher latitudes and covered with a layer of dust or debris which would decrease sublimation, much of it could still be there today (Carr 1990). Carr (1986) also argues that these floods drained only a tenth of a putative global subsurface aquifer system, and that another 300 m of water resides beneath the surface in regions other than Chryse; this water presumably has never made it to the surface, however, and is unlikely to be available as a resource.

The valley networks consist of well-developed drainage networks that presumably involved erosion by sapping from a subsurface reservoir and may have involved runoff of surface precipitation (Baker 1982; Carr 1981). The valleys appear predominantly on the older, heavily cratered terrain, implying formation under somewhat different conditions than occur at the present. The nature of the difference is uncertain, however. One suggestion is the existence of an early greenhouse atmosphere that would have raised the temperature to above the melting point of water (see, e.g., Pollack 1979; Pollack et al. 1987), although the physics of this process has recently been questioned (Kasting 1991). Other possibilities include melting of water beneath a layer of ice emplaced during periods which had an axial obliquity different from that at the present time (Clow 1987; Jakosky and Carr 1985) or melting during early epochs when the geothermal heat flux would have caused the melting isotherm to be near to the surface (Squyres 1989). Under any of these mechanisms, the amounts of water involved are very uncertain due to the presumed recycling of water through the system many times during the erosion process.

Recently, Goldspiel and Squyres (1991) examined the drainage pattern

of the valley networks and located a number of closed depressions into which they debouch; these locations might have contained standing ponds of water at some time. Again, however, the amount of water which may have flowed into them is uncertain, as is the timing of how much water would have been present at a given time; it is possible that subsurface ice might remain in these locations today.

At latitudes poleward of about 40° in each hemisphere, topography appears to have a muted or softened texture (Squyres and Carr 1986). This appearance might be related to the presence of near-surface ground ice, which, if present in sufficient abundance, might allow the regolith to relax and flow over geologic time. The occurrence of ice at these latitudes is consistent with models of the evolution of an ice-filled regolith, with lower latitudes losing ice via sublimation into the atmosphere and higher latitudes retaining ice due to their lower temperatures (Fanale et al. 1986). The high fraction of ice required for flow is a problem, however. Additionally, no observations require the existence of ice in the near-surface regolith at the present.

Additional features such as rampart craters, debris flows, regions of possible dissolution of the regolith by groundwater, etc., may also be indicative of the localized occurrence of near-surface water (see, e.g., Carr 1986; Carr and Schaber 1977; Schaeffer 1990). At least in the case of rampart craters, however, entrainment of atmospheric volatiles is a viable alternative to the release of subsurface volatiles as a mechanism for producing the observed morphology (see, e.g., Schultz and Gault 1979).

Volcanic eruptions will supply volatiles to the atmosphere from gases dissolved in the high-temperature magma. Photogeological analysis has been used to determine the total volume of volcanic materials erupted to the surface as a function of time since the earliest epochs (Greeley 1987; Tanaka et al. 1988; Greeley and Schneid 1991). Based on geochemical evidence from the SNC meteorites, pieces of basaltic lava flow thought to have been ejected from the Martian surface (see, e.g., McSween 1985), erupting lava probably contained about 1% water by mass (Treiman 1985; Scambos and Jakosky 1990). Combining these estimates, the equivalent of a global layer of water up to about 80 m thick might have been emplaced at the surface from volcanic eruptions (Greeley 1987; Greeley and Schneid 1991). This estimate includes the possible release of water to the surface from unobserved intrusive volcanism.

Finally, we note the possibility that the Martian pole has wandered over geologic time (see, e.g., Murray and Malin 1973; Schultz and Lutz 1988), although such wander is by no means proven (Ward 1992). Had it occurred, then regions where water might have ponded (as discussed above) or might have been stable with respect to sublimation into the atmosphere (as described in the next section) would not necessarily be the same regions where water is stable today; water could be present as relict features where it is not currently stable, or could be absent in regions where it would currently be stable.

B. Geochemical Evidence

Geochemical arguments for the volatile inventory and outgassing history come in two basic areas. First, some data pertain to the total volatile inventory of the planet, including those species incorporated during or immediately subsequent to accretion. Second, other data tell us about the outgassing history of the volatiles or about their evolution subsequent to outgassing to the surface. Fanale et al. (1992) have recently summarized the geochemical arguments. Each of these areas is discussed below.

Several geochemical arguments have been made as to how much water would have been incorporated into Mars during its formation. The release of water from the interior to the surface or atmosphere is not generally well constrained by these arguments, however. Additionally, the range of amounts of water incorporated into the accreting Mars vary widely, depending on which data is used in the analysis and what model assumptions are made. The following discussion is not exhaustive, but is meant to be representative of the range of values discussed for the amounts of water outgassed.

Anders and Owen (1977) used estimates of the abundance of volatile elements at the Martian surface, as inferred from the Viking Lander measurements, as a guide to the availability of volatiles as a whole. They assumed that the volatiles were brought in by primitive meteorites, and tried to estimate their amount by scaling from the terrestrial abundances of volatiles. They argued that ^{36}Ar, as a noble gas that might be equally incorporated into the Earth or Mars, could be used as an indicator of total volatile abundance, and that ^{40}Ar, which forms from the decay of ^{40}K in the planet's interior, could be used as an indicator of the relative efficiency of outgassing. The total abundance of K was estimated from a comparison of volatile elements on Mars, Earth, and in meteorites. They estimated that Mars would have incorporated during its formation a total of about 10 m of water (as a global equivalent layer if all of it were outgassed).

The relative abundances of the noble gases was questioned, however, by Pollack and Black (1979). They argued that the pattern of noble gases (Ne, Ar, Kr and Xe) on the terrestrial planets was indicative of a common source, and that the hundred-fold decrease in ^{36}Ar from Venus to Earth, and additional hundred-fold decrease from Earth to Mars, suggested some differential incorporation into the accreting planets from the solar nebula. They suggested that, had the solar nebula been isothermal, then pressure gradients from Venus to Earth to Mars would cause the noble gases to adsorb in different amounts onto planetesimals and could account for their varying abundances. Volatiles such as nitrogen, carbon dioxide, or water would be incorporated in amounts less dependent on the solar nebula pressure and would be present in approximately equal abundances on a per mass basis. In their model, as much as several hundred meters of water might have been incorporated into the interior.

An alternative model was suggested by Dreibus and Wänke (1985,1987),

based on the geochemistry of the SNC meteorites. They assumed that accretion was from two sources of material which are analogous to Fe meteorites (volatile poor) and to carbonaceous chondrite meteorites (volatile rich). The relative abundances of these components, and the subsequent incorporation of volatiles, was determined from the trace elemental abundances in the SNCs. The amount of water incorporated into the planet was determined by scaling the amount of water in the volatile-rich component by the relative depletion of another volatile element (chlorine) and then by the relative solubility in melt of Cl and H_2O, assuming that the water in the planet was incorporated by equilibrium between melt and atmosphere during a globally molten stage. They estimated that Mars contained between the equivalent of about 50 and 150 m H_2O.

The subsequent outgassing history of the water can be inferred from the ^{40}Ar abundance. The measurements within the SNC meteorites allow an inference to be made of the global ^{40}K abundance within the Martian interior. Decay of ^{40}K to ^{40}Ar (in part) occurs with a half-life of about 1.25 Gyr; as a result, had outgassing been early, relatively little ^{40}Ar would have been available for outgassing, while, had it been late, a larger amount would have been present and outgassed. The amount of ^{40}Ar in the atmosphere at present suggests either relatively inefficient or relatively early outgassing of the interior (see, e.g., Anders and Owen 1977; Dreibus and Wänke 1987). When combined with the history of outgassing as inferred from the volcanic and flooding history, a relatively inefficient outgassing is suggested (Scambos and Jakosky 1990).

Unfortunately, the strength of the conclusions reached from the geochemical arguments is uncertain. Estimates for the total abundance of water range between ten and perhaps a few thousand meters, depending on the assumptions of the various models. Additionally, the fraction of the total water which has outgassed to the surface is uncertain. It is likely that the geologic constraints as to how much water has been present at the surface are less ambiguous. The question remains, however, as to what the subsequent fate of this water has been and how much of the outgassed water remains in the vicinity of the Martian surface. These issues can be addressed to some extent by analysis of the stable isotopic clues to the evolution of the volatiles.

Geochemical clues to the subsequent history of water come from two sources—observations of the ratio of $^{18}O/^{16}O$ and of D/H in the Martian atmosphere. These ratios relate to the evolution of volatiles because of the fractionation which occurs as a result of almost any process which involves either exchange of atmospheric volatiles with a non-atmospheric reservoir or loss of one or more species to space. The ratio $^{18}O/^{16}O$ has been measured in atmospheric CO_2 and H_2O, and D/H has been measured in atmospheric H_2O (Nier and McElroy 1977; Bjoraker et al. 1989; Owen et al. 1988); in addition, these same ratios have been measured in various components of the SNC meteorites (see, e.g., Carr et al. 1985; Wright et al. 1986,1988; Clayton and Mayeda 1988; Kerridge 1988). The observations have been synthesized,

along with estimates of the fractionation which results from various exchange and evolution processes by Jakosky (1991).

The observed ratio of $^{18}O/^{16}O$ in atmospheric CO_2 is within about 5% of the terrestrial value (Nier and McElroy 1977). In atmospheric H_2O, however, the ratio is about 10% less than terrestrial (Bjoraker et al. 1989). This difference is readily explained as resulting from fractionation of oxygen between CO_2 and H_2O if the oxygen is chemically exchanging between these two molecules (Jakosky 1991). Although processes such as condensation of ice or formation of carbonates can readily fractionate the oxygen over geologic time (Jakosky 1991), the major fractionation process is probably nonthermal escape of oxygen to space, with the lighter isotope escaping more readily than the heavier one (McElroy and Yung 1976). The lack of a significant fractionation requires the dilution of atmospheric oxygen by exchange with a large non-atmospheric reservoir of oxygen; on the basis of a variety of geologic and geochemical arguments, Jakosky (1991) argued that the oxygen reservoir was probably water ice from the polar caps rather than adsorbed CO_2 in the regolith. The polar caps would need to contain the equivalent of a global layer of ice 15 to 30 m thick in order to have sufficient oxygen; they probably do contain this much water ice, and exchange of this ice through the atmosphere is probably sufficiently rapid that it can act as the buffering agent for atmospheric oxygen.

D/H is both simpler and more complex than oxygen. The observed factor of 5 enhancement in atmospheric D/H relative to the terrestrial value (Owen et al. 1988; Bjoraker et al. 1989) can be explained only by thermal escape of hydrogen to space, with the hydrogen coming from photodissociation of atmospheric water and with the H atoms escaping more rapidly than the heavier D atoms. The magnitude of the enhancement requires the loss of more than 85% of the total available water in the system (Owen et al. 1988; Yung et al. 1988). Unfortunately, the time scale on which this loss occurs and the amount of water implied to have been lost is uncertain. Owen et al. (1988) suggested that water loss occurred early in Martian history, when the solar ultraviolet flux and the atmospheric water content were greater than at present, and that the amount of water lost was the equivalent of a global layer up to hundreds of meters thick. At the other extreme, Yung et al. (1988) determined that, if the current rate of water loss had been constant over geologic time, then only about 3 m of water would have been lost, and the remainder would be exchanging with a non-atmospheric reservoir containing only about 10 cm H_2O. Other alternatives included the loss during periods of high atmospheric water content at high axial obliquity, with loss of about 50 m H_2O (Jakosky 1990), or the continual replenishment of atmospheric water by continued outgassing with the consequence that the amount of water lost and the time scale over which it was lost is indeterminate (Carr 1990).

Additionally, there appears to be a contradiction between the D/H and the $^{18}O/^{16}O$ interpretations. The oxygen data suggest that the most likely scenario involves the exchange of atmospheric water with a non-atmospheric

reservoir (most probably the polar caps), and that a relatively small fraction of the total water has been lost from the system. On the other hand, the D/H data suggest that most of the water has been lost from the system. Jakosky (1991) suggests that these can be reconciled by incorporating the different time scales for loss of each species; in this case, the D/H and $^{18}O/^{16}O$ are set on different time scales and therefore tell us different things about volatile evolution. If this interpretation is correct, then the D/H represents loss of water either very early in Martian history or very recently, in the last 10^5 to 10^7 yr; in the latter instance, the loss is of a small surface reservoir and is not representative of the total water inventory. This hypothesis is consistent with the idea that a relatively small fraction of the outgassed water has been lost over geologic time.

An alternative viewpoint has been suggested by F. P. Fanale (personal communication, 1991). If the regolith contains the required 15 to 30 m of ground ice, then the oxygen in the ice can buffer or dilute the atmospheric oxygen; if the water molecules do not exchange with atmospheric water, the buffering might still occur via the exchange of oxygen in CO_2 as a carrier gas as long as the bulk of the water is in diffusive contact with the atmosphere. In this scenario, atmospheric CO_2 and H_2O equilibrate isotopically, and the CO_2 diffuses into the regolith and equilibrates with H_2O in the regolith. This model has the advantage of effectively decoupling the fractionation of $^{18}O/^{16}O$ from that of D/H. By this scenario, the amount of water which has been lost over geologic time is a small fraction of the ground-ice inventory and at the same time a large fraction of the water which does exchange with the atmosphere. While these competing theories have very different implications for the exchangeable water, they both require that a small fraction of the total inventory has been lost.

C. Summary

The above discussion of outgassing and loss of water at the surface and of the volatile inventory is very uncertain due to the lack of firm constraints on the system as a whole. We can summarize the results as follows. The amount of water supplied to the surface is probably on the order of the equivalent of a global layer 100 m thick. This is based on the supply of approximately 50 m H_2O from global volcanism and approximately 50 m H_2O from the outflow channels; of course, the uncertainties in this number gloss over questions involving (i) the extent to which the same water molecules might have traveled through both reservoirs, or (ii) the degree to which the latter number might actually represent a minimum amount of water required. Certainly, the amount of water outgassed to the surface can be as high as several hundred meters of water or more. The subsequent loss of water from the system is to some extent equally uncertain. Approximately 30 m of water could be stored in the polar caps (see below); this water probably exchanges with the atmosphere on geologic time scales. Alternatively, the 30 m of water could reside within the global regolith. Somewhere between a few meters

and perhaps several hundred meters of water has been photodissociated and lost by escape to space. Unfortunately, this leaves very uncertain the amount of water still in the near-surface regolith. Likely reservoirs for this water are (i) ponded in the northern lowlands (primarily from the outflow channels or at the terminus of some of the valley networks), (ii) physically or chemically bound within the regolith, or (iii) as ice or liquid in the global or high-latitude regolith. Several of these reservoirs will be discussed in following sections.

II. CURRENT STATE OF WATER

The existence of water in the near-surface layer is inferred from observations of the seasonal cycle of water vapor in the atmosphere in conjunction with other observations of the physical and chemical properties of the regolith. The relevant observations and analyses are described in this section. The following subsections include discussion of inferences which can be made directly from the seasonal cycle of atmospheric water, observations pertinent to water in the residual polar caps, and estimates of the amount of water stored in either the global or the high-latitude regolith. In each section, discussion will generally be limited to what can be deduced about water which the observations require to be present in the near-surface region; additional speculation will be presented briefly about what the observations allow but do not require. The final section of the chapter will include a description of observations which can allow us to better measure the actual water content of the regolith.

A. Inferences from Observations of Atmospheric Water

Figure 1 shows a summary of the observations of atmospheric water vapor as a function of latitude and season, made from the Viking Orbiter spacecraft. A full discussion of the seasonal water cycle can be found in review papers by Jakosky and Haberle (1992) and Jakosky (1985). The major focus of this section will be on the evidence in the observations for a time-variable global abundance of water in the atmosphere and the implied seasonal exchange with (and therefore the existence of) non-atmospheric reservoirs of water. Detailed discussion of the specific reservoirs is contained in the following subsections.

Figure 2 shows the globally integrated abundance of atmospheric water vapor as a function of season. Notice that it varies with time by about a factor of 2; it is this variation that requires seasonal supply and removal from the atmosphere. The possible seasonal reservoirs for the water are the seasonal polar caps, the residual polar caps, and water in the regolith.

The peak column abundance occurs over the north polar residual cap during the northern summer season (Fig. 1). This requires that the polar region be a source of water vapor at that time. The seasonal CO_2 polar cap disappeared at around $L_s = 80^\circ$ in the north (see, e.g., Paige and Ingersoll 1985), so the water is not being released from the retreating CO_2 frost. Rather, detailed water vapor and thermal mapping suggest that the residual polar ice deposits consist of water ice, and that this water sublimes into the atmosphere

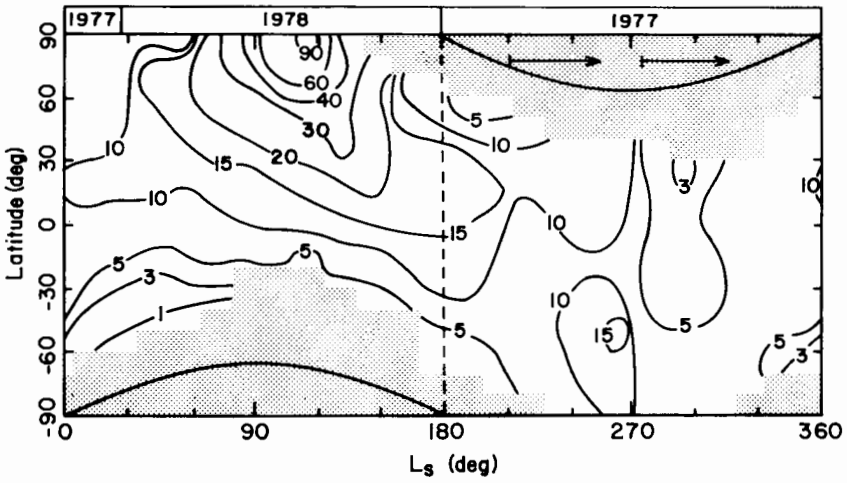


Figure 1. Summary of the column abundance of water within the atmosphere as a function of latitude and season, as measured from the Viking spacecraft. Values are in units of precipitable microns, equivalent to about 10^{-4} g cm^{-2}. The data are described by Jakosky and Farmer (1982), and the figure is taken from Jakosky and Haberle (1992).

when heated to temperatures near 205 K during the summer (Kieffer et al. 1976; Farmer et al. 1976). There is some question as to the total amount of water which actually sublimes into the atmosphere from the polar ice (Haberle and Jakosky 1990), but the existence of the ice deposits is not at issue. Additionally, it is not possible to distinguish between the amount of water sublimed from the polar ice versus that sublimed from ice or other water located within the circumpolar debris deposits but not part of the residual cap itself.

A similar large atmospheric water-vapor abundance is not seen in the south polar region during southern hemisphere summer (Fig. 1). Although some atmospheric water might be masked from view by atmospheric dust, the explanation is more likely that the south polar cap did not lose its seasonal cover of CO_2 frost (see, e.g., Kieffer 1979). As the surface thus remains near the CO_2 frost-point temperature of around 150 K, it acts as an efficient cold trap for atmospheric water; therefore, it is likely that there is a deposit of water ice of unknown thickness underlying the CO_2 frost (Jakosky and Farmer 1982; Jakosky 1985). Although the south pole was also covered with CO_2 frost during the summer observed by the Mariner 9 spacecraft (Paige et al. 1990), it might lose its CO_2-frost covering during some years and expose the underlying water-ice cap (Jakosky and Barker 1984; Jakosky and Haberle 1990). The thickness of the residual CO_2 frost deposit is uncertain, but it is unlikely to be substantial (see, e.g., Murray and Malin 1973; Jakosky and Haberle 1990).

The amount of water vapor in the atmosphere increases prior to the expo-

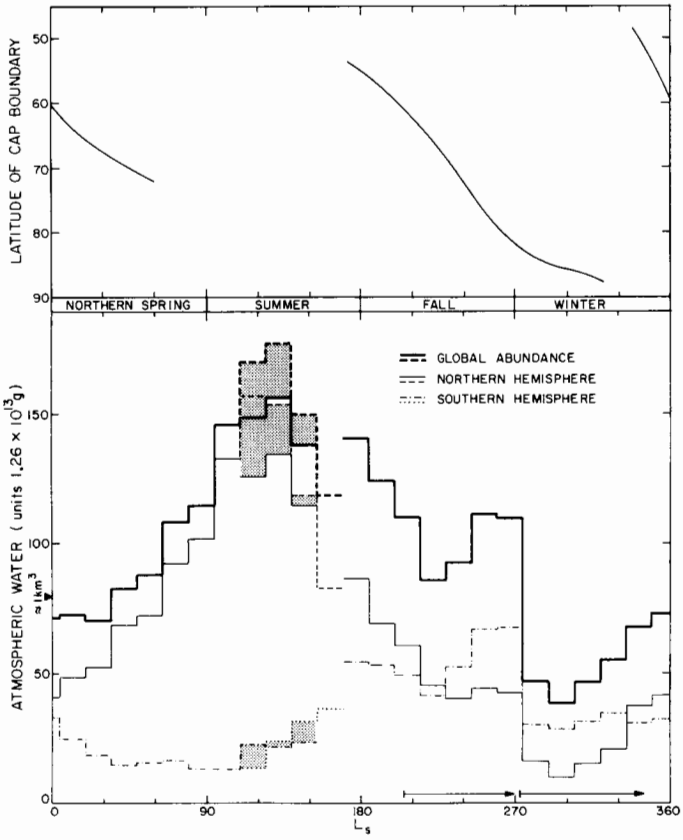


Figure 2. Globally integrated abundance of water within the Martian atmosphere as a function of season; northern hemisphere and southern hemisphere values are shown as well as total global values (figure from Jakosky and Farmer 1982).

sure of the residual polar cap (see Figs. 1 and 2), consistent either with release from the retreating cap itself or from the underlying regolith; distinction between these two sources is difficult (Jakosky 1983*b*). Exchange with water adsorbed within the regolith is expected if the regolith is not impermeable to diffusion; any unbonded or loosely bonded particulate material will meet this requirement. Models of the exchange with the regolith suggest that it is capable of supplying a large fraction of the seasonal variation of atmospheric water (Jakosky 1983*a*, *b*).

The roles of supply of water from the regolith and supply of water from the polar caps are intimately connected to that of transport of water through the atmosphere. For example, models of the transport of water away from the north polar residual cap during the summer season (Haberle and Jakosky 1990) found that the strength of the atmospheric circulation was not sufficient to remove the water as rapidly as it appears to be transported in the observations

shown in Fig. 1. They required that much of the nonpolar water which appears after $L_s = 80°$ had diffused into the atmosphere from the nonpolar regolith. It is likely that both the regolith and the polar caps act as seasonal sources of atmospheric water vapor; quantifying their relative contributions will require additional, more-detailed observations of the global and seasonal distribution of atmospheric water. Further discussion of the physical state and distribution of water within the regolith follows below.

B. Water in the Polar Caps

The residual north polar cap is composed primarily of water ice, at least at the surface. This conclusion is based on: (i) the high albedo of the polar cap, which is indicative of the presence of frost; (ii) temperatures of the surface material during the summer season, which are too high to be consistent with the presence of CO_2 frost on the surface in equilibrium with the CO_2 vapor in the atmosphere (Kieffer et al. 1976); and (iii) the appearance of relatively large amounts of water in the atmosphere over the residual cap during summer, approximately consistent with sublimation of water into the atmosphere from a surface ice deposit at the observed temperatures (Farmer et al. 1976; Haberle and Jakosky 1990). Issues which will be discussed in this section include the thickness of the polar ice deposits, the relative proportions of ice and dust in the deposits, and the possibility of water ice being present near the surface in the south polar region. The possible presence of ice within the regolith equatorward of the polar cap will be discussed in the next section.

The north polar region contains debris deposits poleward of about +80° latitude (Soderblom et al. 1973). The major features consist of layered deposits surrounding the residual ice deposits. Additional features include plains units which surround the layered terrain and a circumpolar sand dune field (Soderblom et al. 1973; Cutts 1973; Cutts et al. 1976; Squyres 1979; Tsoar et al. 1979; Carr 1981). The layered deposits appear to consist of alternate layers of dust and ice or of dust and ice mixed in varying proportions; these layers are generally thought to be deposited and eroded in response to quasi-periodic changes in the climate due to (primarily) the changing axial obliquity (see, e.g., Ward 1974; Pollack and Toon 1982).

The amount of ice present in the polar regions is somewhat uncertain. Based on the observations which require the presence of ice at the pole, as described above, the thickness of the water-ice deposit need only be on the order of 1 mm. Water ice of this thickness can explain: (i) the observed high albedo, owing to the efficient scattering of sunlight by only a thin layer (see, e.g., Clark 1980); and (ii) the total amount of water sublimed from the residual cap during the course of the summer season observed by the Viking spacecraft (Haberle and Jakosky 1990). The fact that discrete layers of, presumably, ice and dust are observed, with the individual layers estimated as being approximately 15 to 30 m thick (see, e.g., Cutts 1973; Blasius et al. 1982), suggests that the surface layer of ice might be of similar thickness. Additionally, the fraction of the total volume of the frost which actually

consists of frost, as opposed to dust, is uncertain. Based on the observed albedo of the cap and the assumption that dust and water ice were deposited simultaneously in the polar regions, Toon et al. (1980) estimate that the deposits would consist of 15% dust and 85% water ice. Given the large uncertainty in the net transport into the polar regions of both dust and water ice, and the possible separation of the dust from the ice as a result of summertime sublimation of water ice (see, e.g., Jakosky and Haberle 1992), however, this estimate has to be considered as extremely uncertain. In addition, Kieffer (1990) points out that the grain growth which occurs in the summertime would decrease the observed scattering from ice grains, such that the observed albedo would require dust to be present at a level of only one part per thousand. This issue has not yet been resolved.

The total thickness of the layered deposits was estimated by Dzurisin and Blasius (1975), based on the limited observations from which the surface topography could be determined, by fitting a smooth plane to the perimeter of the deposit and determining the central thickness. Their thickness estimate is 4 to 6 km. Malin (1986) re-analyzed their topographic contours in order to allow them to close, and estimated a total volume of the polar deposits of 2.1×10^{21} cm^3; the uncertainty in this estimate is probably a factor of 2 due to the uncertainty in the underlying topography. Additionally, Malin (1986) analyzed Viking Orbiter gravity data to determine the total mass of the deposits, and obtained a value of about 2×10^{21} g; this is the mass that needs to be added to the fourth-order gravity field in order to produce the observed accelerations, and also may be in error if the fourth-order field does not adequately represent the underlying surface. Based on these values, he estimates a density of about 1 g cm^{-3}, with an uncertainty of perhaps 50%. Assuming a mixture of ice and silicates, therefore, Malin (1986) estimates that about 95% of the volume of the deposits is water ice, with the uncertainties allowing values no lower than about 55%; this estimate, of course, applies to the whole thickness of the deposits rather than to the uppermost near-surface layer.

Finally, turn again to the question of water ice in the south polar residual cap. As discussed, the residual cap observed from both Mariner 9 and Viking consisted of CO_2 frost (Paige et al. 1990; Kieffer 1979). As the frost acts as a cold trap, it should have water ice mixed in with or lying beneath the CO_2 frost, although the amounts which might be present are clearly uncertain. The residual CO_2 frost is probably no thicker than perhaps several tens of meters, based on its observed morphological appearance (see, e.g., Murray and Malin 1973; Pollack and Toon 1982). It could be as thin as only centimeters, and might even disappear entirely in some years to expose the underlying water ice (Jakosky and Barker 1984; Jakosky and Haberle 1990).

C. Water in the Regolith

Water in the Martian regolith can occur as vapor, as ice, incorporated into the crystalline structure of weathering products, potentially in solid and liquid

solutions, or as a surface species adsorbed on geologic materials. The possible occurrences within the near-surface regolith are discussed in this section.

Water of Hydration. Water molecules can be incorporated into the structure of crystalline geologic materials, such as birnessite, a common manganese mineral on Earth, or ferrihydrite. Such materials are said to contain water of hydration; that is, water molecules are constituents of the unit cell of the crystalline material. Often, water in the silicate melt has had one of its protons removed, such that the primary igneous minerals that pecipitate from the melt contain abundant hydroxyl groups (e.g., hornblend, amphiboles, etc.). It is relatively rare that water molecules are wholly incorporated into primary igneous minerals, however, and water of hydration is most often seen in secondary minerals—those that have formed as a result of chemical weathering at or near the planet's surface. A prominent example of hydrated secondary minerals on Earth is gypsum ($CaSO_4 \cdot 2H_2O$). In some instances, water may be incorporated into solid solutions, when, for example, water is present in a silicate melt that is quenched too rapidly for exsolution of the trapped H_2O. In such materials, water is frozen in a glassy, poorly ordered molecular structure that characterizes the bulk silicate melt.

It is interesting to note in connection with the possibility of water being trapped in a solid solution that spectral evidence suggests that the materials that constitute the Martian surface show poor crystallinity and some evidence that water has been incorporated. As we shall see, however, evidence is equivocal about the precise nature of that water.

Among the earliest evidence that relates directly to the abundance of water in the Martian surface is Earth-based telescopic data (Houck et al. 1973). Their spectroscopic data contained absorption features due to water in the atmospheres of Earth and Mars, as well as within the Martian surface materials; although they were able to suggest that their spectra indicated that the abundance of water was around 1% by mass, they could not demonstrate that the water was chemically bound, as in a phyllosilicate structure, or whether it was adsorbed onto the surfaces, or both.

The Viking Gas Chromatograph Mass Spectrometer (GCMS) instrument observed the evolution of water from samples of the Martian regolith, and Biemann et al. (1977) and Anderson and Tice (1979) suggested that around 0.2% by weight was evolved during heating of the sample to 500° C. Because most of the water was evolved between 250 and 500° C, it is likely that it was water of hydration. The results from both Viking Landers were broadly similar, suggesting between 0.1 and 1.0% water by weight was distributed more or less uniformly across the planet. This result is in agreement with apparent spectral similarity of Mars across its surface (see, e.g., Bell and McCord 1989).

Data from the Soviet Phobos spacecraft appear to show substantial spatial variation in the hydration state of the surface. Observing the Isidis-Syrtis region, Erard et al. (1991) suggest that bright dust in the Isidis region is considerably more hydrated, with a band depth of 40% in the 2.9 μm OH

band, than the darker Syrtis region, where the band depth was only 25%; of course, particle size or compositional variations may also contribute to this difference.

Employing water of hydration as a resource would be more or less difficult depending on the form the water took. In general, the more ordered the crystal site the more energy is required in order to break the electronic bonds of the molecule and extract it to a free-gas stream where it can be collected. The general procedure for acquiring water of hydration for use would probably be to heat the material while simultaneously flushing with a carrier gas of near-zero relative humidity. More specific design will have to await more accurate definition of the state of the water in the Martian regolith.

Adsorbed Water. The presence of adsorbed water on Mars can be inferred with a very high degree of confidence. Adsorption is the net accumulation of a substance at the interface between two contiguous phases (Sposito 1984). In this case, the phases are represented by the Martian regolith and atmosphere. Adsorption is enhanced under conditions of low temperature and high relative humidity, both of which occur in the Mars atmosphere; at the same time, the regolith presents an enormous amount of interfacial area to the atmosphere. The physical adsorption of water in such an environment is assured, and only the relative amounts are in question.

Adsorption occurs via two rather different mechanisms, physical and chemical. In physical adsorption, van der Waals forces bind the adsorbate molecule to the adsorbent. Because water is a molecule with a strong permanent dipole, it is considerably more susceptible to adsorption than nonpolar molecules such as CO_2 (see, e.g., Carter and Husain 1974). In chemical adsorption, or chemisorption, an actual bond is formed between the surface species and the adsorbent. This requires changes in the orbital clouds around the adsorption site, and is favored at high temperatures. Chemisorption is almost irreversible, and is not expected to play a significant role in storing water on Mars. Although the composition of the adsorbent is important, its specific surface area is far more important. The specific surface area of a sample is the combined surface area of all the particles in the sample, including the effects of grain roughness and surface irregularities; this is an operational definition, and unique and different surface areas are often found for a material when different molecules are used as surface probes.

Laboratory adsorption experiments exposing water to adsorbents of broad geologic interest have been performed at low temperatures and high relative humidities, but there is no report of them being carried out at conditions that truly reflect the Martian surface (Mooney et al. 1952; Fanale and Cannon 1971). Nonetheless, it is possible to quantitatively estimate the amount of adsorbed H_2O on a variety of geologic materials at Mars-like conditions (Fanale and Cannon 1974,1978; Zent et al. 1986). These calculations suggest that 0.1 to 1% H_2O by weight might be common for materials near the Martian surface (see Fig. 3). However, should a substantial portion of the regolith be composed of smectites, a clay mineral rich in aluminium, then substantially

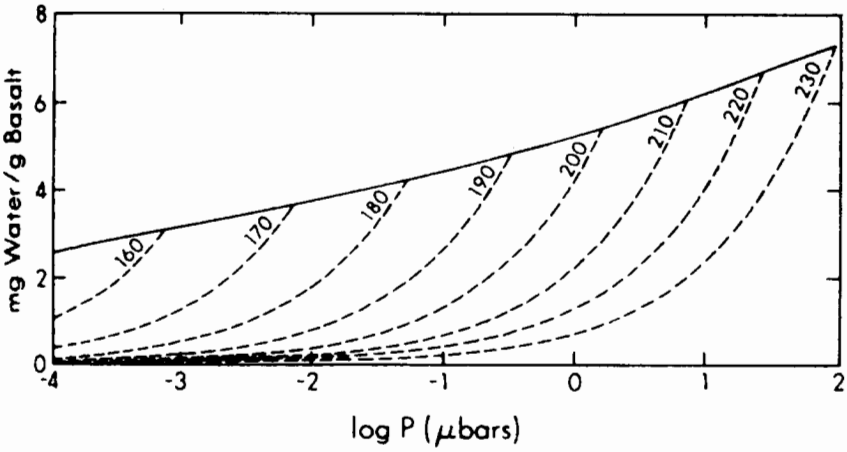


Figure 3. Adsorption isotherm of water on basalt, calculated from the data of Fanale and Cannon (1971) at temperatures of 273 K and 293 K. Each isotherm shows how much water is adsorbed onto the regolith grains as a function of water vapor partial pressure, and is truncated at the vapor pressure over ice; addition of more water to the system would then produce ice rather than additional adsorbate.

more water can be stored; the layers in smectite clays can actually pull apart in order to accommodate the adsorption of water molecules, thus increasing their specific surface area. Smectites are often called "swelling clays" because of the increase in volume associated with adsorbing up to several tens of percent of their original mass. If there are significant amounts of smectites on Mars, adsorbed water could be a very abundant resource. If the regolith on Mars is 100 m thick, for example, then the equivalent of a few meters of liquid water can be stored as adsorbate in the regolith; if the same amount of regolith were comprised of smectite clays, several tens of meters might be adsorbed. The presence of smectite clays in the regolith, however, is controversial and not universally accepted (see Banin et al. 1992).

It is interesting to note that the Viking GCMS provided significant evidence only for water of hydration, although the sample which was analyzed was stored at temperatures which are high enough that any adsorbed water would have been lost to the atmosphere. Physical adsorption is expected to be ubiquitous, however, although direct evidence that would indicate the abundance of adsorbed water as unequivocally as the GCMS indicates some kind of structural H_2O is lacking. One unknown that may influence H_2O adsorption is the predominance of CO_2 in the vapor phase. Although it is expected on theoretical grounds that CO_2 will not strongly influence H_2O adsorption, more direct evidence would be welcome.

Spatial variations in the amount of adsorbed water in the regolith are expected to occur, but little evidence for it exists to date. It is expected that adsorptive coverage might be appreciably greater as one moves closer to the

north pole in particular, however, because relative humidity increases and average temperature decreases in that direction.

Extracting adsorbed water as a resource will present considerably fewer difficulties than water of hydration, because the water all resides on the physical surface of the adsorbing material, and is directly accessible to the free gas or pore vapor. In fact, it is often possible in terrestrial laboratory work to remove much of the adsorbed water simply by purging the headspace with dry nitrogen. Heating the sample expedites the evolution of water.

Brines. Brines, which occur commonly on Earth, are concentrated aqueous solutions. They are formed by exposing solute-bearing waters to an arid environment, often by the evaporation or freeze drying of naturally occurring waters such as lakes or shallow marine bays. On Mars, pure liquid water is currently unstable because the minimum partial pressure of water vapor above the liquid phase is 6 mbar; although ice can melt under 6 mbar mechanical pressure, such as might be supplied by CO_2, it will rapidly vaporize and be lost to the atmosphere. However, concentrated solutions of electrolytes exhibit depression of both the freezing point and equilibrium vapor pressure. Freezing-point depression allows temperatures which occur in the Martian regolith to satisfy the melting requirement. Almost all natural brines on Earth reach their eutectic temperature, the point at which they solidify without undergoing any additional compositional change, at around 218 K. However, even the lowest equilibrium vapor pressures over brines are on the order of ten times greater than the observed water vapor abundance in the Martian atmosphere (Brass 1980); therefore, even brines are not currently stable anywhere on Mars that is in diffusive communication with the atmosphere. However, there are severe kinetic limitations to the rate at which water molecules might be extracted from a brine through a cold, porous, and strongly adsorbing medium such as the Martian regolith (Farmer 1976). The kinetic bottleneck admits the possibility that unstable brines might persist for geologically significant periods of time in the subsurface. Zent and Fanale (1986) calculated that any brine shallow enough to melt on an annual basis would be entirely lost by sublimation to the atmosphere in probably no more than several times 10^4 yr. However, brines at depths of several hundred meters could probably survive throughout much of Martian geologic history (Fanale et al. 1986).

The actual existence of brine on Mars at present, however, is problematic. As with adsorbed water, there is adequate geologic evidence to infer that brines must have formed at some point in the past in association with the evaporation or freezing of pooled groundwaters. Viking imagery shows ample evidence of erosional valleys, due to groundwater sapping and precipitation runoff, that empty into depositional basins (see, e.g., Goldspiel and Squyres 1991). Reaction rates even in cold water are such that groundwaters all contain solutes leached from the country rock. At some point, any putative pooled groundwaters were probably subjected either to evaporative loss or to temperatures below 273 K or both. In either event, water molecules would have been preferentially removed from the solution, leaving an increasingly

concentrated brine. Specific gravity increases with concentration, and it would have been possible for brines to seep into the substrate, mixing with or displacing less-dense water. Similar mechanisms operate on Earth, both in evaporative bays (Sonnenfeld and Perthuisot 1989) and in rare situations where shallow bays are cut off from the open ocean and brine formation is the result of freezing ocean waters (Herut et al. 1990).

The observed enrichment of sulphur and chlorine in crusts or clods of soil at the Viking landing sites (Toulmin et al. 1977) suggests the possibility of evaporitic deposition from a brine at some point in the past; however, data pertaining to the present situation is more difficult to come by. Zisk and Mouginis-Mark (1980) examined radar data collected by Downs et al. (1973,1975) and suggested that the Solis Planum region of Mars became more absorbing of 13.6-cm radiation as summer approached; they suggested melting in a subsurface brine as a possible explanation because ice reflects radar wavelength radiation, whereas it is strongly absorbed by liquid water. Other investigators (Jakosky 1985) suggested that the brine hypothesis was not an optimal explanation for the whole ensemble of observations of the Martian surface. A comparison of the original data with thermal models of the Martian surface led Zent et al. (1990) to suggest that the data showed only a few instances of what might have been verifiable changes in reflectivity, and that these were not compatible with any brine that would have been stable in the bulk Martian regolith (Clark and van Hart 1981). In fact, some reflectivity increases appear to take place well after the surface had begun to cool.

Future investigations into the existence and utility of Martian brines will likely center on theoretical investigations of their preservation at depth. Consideration of the amount of water needed to erode even some of the more modest valley networks on Mars suggest that an amount of water equal to several years worth of Mississippi River discharge may have pooled in each of several depositional basins mapped by Goldspeil and Squyres (1991). Preservation of even 0. 01% of that water as subsurface brines would indicate that several cubic kilometers of brine would exist at depth.

Ice. Water molecules will be stable as ice whenever the temperature is below the saturation temperature appropriate for the partial pressure of water vapor present as gas. The stability criterion is expressed in this manner because the water vapor partial pressure is generally determined over time scales longer than a year by exchange of water with the polar caps and transport through the atmosphere (Jakosky 1985). Water ice can occur either within the atmosphere, if water-laden atmosphere is allowed to cool, or within the regolith, if the subsurface temperatures allow saturation with respect to the gas within the regolith pore spaces; the former cannot be in general a significant reservoir for water, so we will concentrate on the latter.

Leighton and Murray (1966) and then Farmer and Doms (1979) pointed out that the regolith at high latitudes during winter will be cool enough that water vapor present in the atmosphere will condense. Further, depths lower than several tens of centimeters would not heat up sufficiently during the

summer that the ice would be unstable with respect to sublimation, such that these regions could contain water ice which would be stable at all times of year. Another way of saying this is that the water vapor gas number density within the pore spaces at depth is controlled by the number density within the atmosphere; at depth, diffusion of water would be sufficiently slow that the pore-space number density would equal the annual average atmospheric number density. In the high latitudes, then, the subsurface temperature never rises above the value for which this number density would saturate. Therefore, ice is stable within these locations. Figure 4 shows the regions of stability of ice, as calculated by Farmer and Doms (1979).

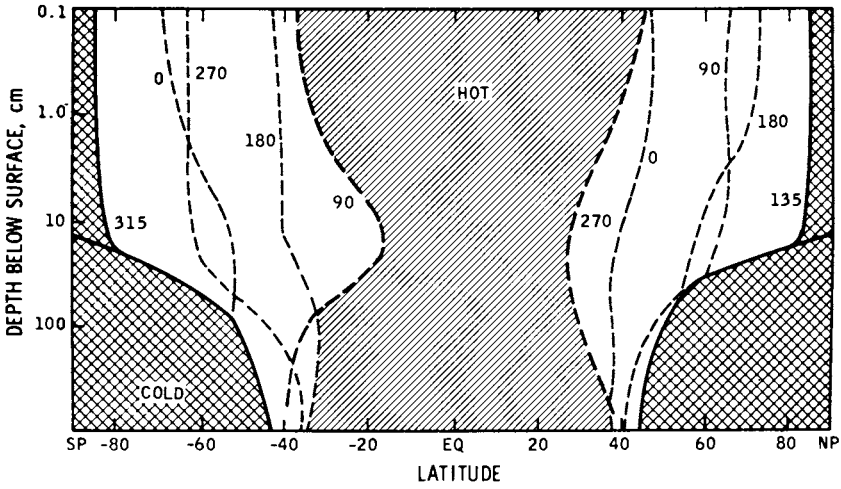


Figure 4. Regions of stability of ice within the Martian regolith, as calculated by Farmer and Doms (1979). Cross-hatched regions are where the temperature is below the saturation temperature at all seasons, and correspond to locations where ice should be stable all year. Clear regions at mid- and high-latitudes are where ice might be stable only during some parts of the year; these locations would not be expected to accumulate ice. In the equatorial region, ice would not be stable at any part of the year at any depth.

There are several caveats that need to be applied to this discussion of water ice. First, the regions where ice is stable as shown in Fig. 4 were calculated for the average global surface properties (albedo and thermal inertia). In fact, the temperature at depth varies significantly as a function of these properties (see Jakosky and Muhleman 1980), and these properties also vary tremendously from location to location; as a result, the regions where water ice will be stable should show tremendous spatial variability. Detailed calculations demonstrate this variability (Mellon and Jakosky 1993; Paige 1992).

Second, the fact that ice is stable at a given location does not guarantee that it is actually present. Water is emplaced within the near-surface regolith by diffusion down the gradient of number density or, equivalently for ice

formation at saturation, down the gradient of temperature. A minimum in temperature will occur at the depth where the seasonal thermal wave diminishes and the geothermal heat flow becomes important; at this depth (meters to tens of meters), ice will be most stable. If water is present at deeper locations, it will not rise above this level. Water ice at shallower locations, which would be consistent with equilibrium with the atmosphere, would have diffused in from the atmosphere, would be a relict of its original emplacement, or would have been emplaced by drainage of liquid from the surface. Assuming equilibrium between the atmosphere and subsurface, the amount of water actually present within the regolith will depend on the ability of water molecules to diffuse from the atmosphere into the regolith. There are locations where the regolith may act as a significant barrier against diffusion (see, e.g., Jakosky and Christensen 1986), such that water will not tend to diffuse into the regolith.

Clearly, ice within the high-latitude regolith may be the most accessible and usable source of water. There are large uncertainties as to where even this water is, however, as well as how much water might be present. Further numerical models, based on data currently in hand, can be used to address some of the issues related to volatile inventory and evolution as well as the stability of water in various forms within the regolith; other issues, however, will not be resolved until appropriate remote sensing or *in situ* sampling of the subsurface can be done.

III. SUMMARY AND FUTURE OBSERVATIONS

In the following, we summarize the inferences to date regarding the distribution and availability of water in the Martian near-surface environment, and we suggest possible future measurements which would better allow us to determine where water might exist. The available evidence suggests the following conclusions about the location and availability of water as a resource on Mars. These are roughly in decreasing order of confidence in the interpretation, with the 5th point marking the boundary between what is known with some confidence and what is more speculative.

1. There is abundant water in the atmosphere, primarily in the form of vapor with a seasonally and spatially variable mixing ratio of approximately 10^{-4}.
2. The north polar residual cap contains water ice, probably in significant abundance, which is readily accessible at the surface.
3. The south polar residual cap contains water ice near the surface, probably under a thin (centimeters to tens of meters) layer of residual CO_2 frost.
4. The global near-surface regolith contains adsorbed and chemically bound water, in amounts between 1 and 30% by mass.
5. The high-latitude, near-surface regolith (top ten meters) may contain a significant quantity of water ice which is stable against sublimation into the atmosphere.
6. These high-latitude deposits might extend to a depth of a kilometer or

more, based on the present stability of ice and its assumed presence from early in Martian history.

7. The northern lowlands and various closed depressions into which the outflow channels and valley networks debouched, respectively, may contain large deposits of water ice near the surface and buried beneath a debris or dust layer of unknown thickness.
8. A large fraction of the water at the surface, and possibly all of the ice except that currently in the polar regions, may have been lost to space via photodissociation and thermal and non-thermal escape of the constituent atoms.

Clearly, although water is an invaluable resource for potential future exploration missions, especially those involving humans, we know very little with certainty about its geographical distribution. Based on our current understanding of what is known as fact versus what is inferred or surmised from the limited observations, water can be obtained from the following sources: (i) from the north-polar cap, in essentially unlimited quantities and with minimal mining or processing efforts; (ii) from the nonpolar regolith, where extraction from the regolith would be an energy- and effort-intensive mining process; or (iii) from the atmosphere.

There are observations which can be made from orbital or *in situ* spacecraft which could considerably lessen the uncertainties in where the water resides and what its abundance is. A brief discussion of these observations follows.

First, we will discuss observations from the Mars Observer mission, launched in late 1992; the spacecraft will be in a near-polar orbit and will observe the surface and atmosphere for a nominal mission of one Martian year (Albee and Palluconi 1990). Although none of the instruments is designed explicitly to search for water, several of them will make observations which can significantly improve our understanding of water on Mars. The atmospheric sounder is a limb-scanning spectrometer called the Pressure Modulator Infrared Radiometer (PMIRR). By obtaining global information on the distribution of water vapor through the Mars atmosphere, it should provide better information on the seasonal sources and sinks of atmospheric water vapor and, therefore, on the distribution and accessibility of non-atmospheric reservoirs of water (McCleese et al. 1986). The Gamma-Ray Spectrometer (GRS), using both neutron and gamma-ray modes, will spatially map the near-surface hydrogen abundance, which is closely related to the water concentration; as such, it will determine the global distribution of near-surface (approximately top meter of regolith) reservoirs of water with a horizontal resolution of $\sim$500 km (see, e.g., Boynton et al. 1992). The Thermal Emission Spectrometer (TES) will map surface temperature and constrain water-ice stability and sublimation rates, both in polar and nonpolar regions, and will aid in understanding the near-surface distribution of water ice (Christensen et al. 1992). Finally, the Mars Observer Camera (MOC) will obtain high-resolution images which

may show small-scale geologic features related to the abundance and history of water (Malin et al. 1992); historically, this type of data has been of the most value in defining our view of Mars.

Additional observations could be done to aid the search for water. Electromagnetic sounding is capable of detecting the presence of either liquid water or solid ice within the regolith, based on its electrical properties as a function of wavelength (see, e.g., Olhoeft and Strangway 1974). The depth probed within the regolith depends on the frequency of the detected radiation, and sounding the uppermost meters to hundreds of meters is possible. The horizontal scale which is observed depends both on the wavelength and on whether sounding is done from an orbiter or a landed vehicle; clearly, use of an orbiter is required for global mapping.

Further information on the volatile inventory and its loss history would help constrain the total amount of water present within the global near-surface environment. Such information could come from measurements of the present escape fluxes of oxygen, hydrogen and carbon; from measurement of the isotopic ratios within the polar ice deposits; and from isotopic measurements within well-characterized rock samples, which would have a direct bearing on both the volatile history and geologic evolution of the surface. Returned samples could be analyzed directly for the presence of water in various forms and for the physical properties which would be conducive to the presence of water.

In addition, any mission which relied on the presence of water would probably require obtaining drill cores of the uppermost meters to tens of meters at a variety of sites, in order to look for water directly, either as a transient liquid or brine or as solid ice; no remote-sensing observations are so convincing as to be absolute in the absence of real ground-truth measurements.

Without these additional measurements, we are left with the conclusion that the only sure source of water in amounts significant as a resource is the polar ice deposits. All other locations either contain water in abundances too small to be usefully accessed or may contain no water at all within distances which are accessible from the surface.

Acknowledgments. Discussions with F. Fanale and M. Mellon were appreciated, as were comments from R. Zurek and F. Fanale on an early version of the manuscript. This research was supported in part by the NASA Planetary Atmospheres Program through a grant to the University of Colorado.

REFERENCES

Albee, A. L., and Palluconi, F. D. 1990. Mars Observer's global mapping mission. *EOS: Trans. AGU* 71:1099; 1107.

Anders, E., and Owen, T. 1977. Mars and Earth: Origin and abundance of volatiles. *Science* 198:453–465.

Anderson, D. M., and Tice, A. R. 1979. The analysis of water in the Martian regolith. *J. Molec. Evol.* 14:33–38.

Baker, V. R. 1982. *The Channels of Mars* (Austin: Univ. of Texas Press).

Banin, A., Clark, B. C., and Wänke, H. 1992. Surface chemistry and mineralogy. In *Mars,* eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 594–625.

Bell, J. F., III, and McCord, T. B. 1989. Mars: Near-infrared comparative spectroscopy during the 1986 opposition. *Icarus* 77:21–34.

Biemann, K., Oró, J., Toulmin, P., III, Orgel, L. E., Nier, A. O., Anderson, D. M., Simmonds, P. G., Flory, D., Diaz, A. V., Rushneck, D. R., Biller, J. E., and Lafleur, A. L. 1977. The search for organic substances and inorganic volatile compounds in the surface of Mars. *J. Geophys. Res.* 82:4641–4658.

Bjoraker, G. L., Mumma, M. J., and Larson, H. P. 1989. Isotopic abundance ratios for hydrogen and oxygen in the Martian atmosphere. *Bull. Amer. Astron. Soc.* 21:991 (abstract).

Blasius, K. R., Cutts, J. A., and Howard, A. D. 1982. Topography and stratigraphy of Martian polar layered deposits. *Icarus* 50:140–160.

Boynton, W. V., Trombka, J. I., Felman, W. C., Arnold, J. R., Englert, P. A. J., Metzger, A. E., Reedy, R. C., Squyres, S. W., Wänke, H., Bailey, S. H., Bruckner, J., Callas, J. L., Drake, D. M., Duke, P., Evans, L. G., Haines, E. L., McCloskey, F. D., Mills, H., Shinohara, C., and Starr, R. 1992. Science applications of the Mars Observer gamma ray spectrometer. *J. Geophys. Res.* 97:7681–7698.

Brass, G. W. 1980. Stability of brines on Mars. *Icarus* 42:20–28.

Carr, M. H. 1979. Formation of Martian flood features by release of water from confined aquifers. *J. Geophys. Res.* 84:2995–3007.

Carr, M. H. 1981. *The Surface of Mars* (New Haven, Conn.: Yale Univ. Press).

Carr, M. H. 1986. Mars: A water-rich planet? *Icarus* 68:187–216.

Carr, M. H. 1990. D/H on Mars: Effects of flood impacts, volcanism and polar processes. *Icarus* 87:210–227.

Carr, M. H., and Schaber, G. G. 1977. Martian permafrost features. *J. Geophys. Res.* 82:4039–4054.

Carr, M. H., Grady, M. M., Wright, I. P., and Pillinger, C. T. 1985. Martian atmospheric carbon dioxide and weathering products in SNC meteorites. *Nature* 314:248–250.

Carter, J. W., and Husain, H. 1974. The simultaneous adsorption of carbon dioxide and water vapour by fixed beds of molecular sieves. *Chem. Eng. Sci.* 29:267–273.

Christensen, P. R., Anderson, D. L., Chase, S. C., Clark, R. N., Kieffer, H. H., Malin, M. C., Pearl, J. C., Carpenter, J., Bandiera, N., Brown, F. G., and Silverman, S. 1992. Thermal emission spectrometer experiment: The Mars Observer mission. *J. Geophys. Res.* 97:7719–7734.

Clark, B. C., and Van Hart, D. C. 1981. The salts of Mars. *Icarus* 45:370–378.

Clark, R. N. 1980. The surface condensates on Mars observed by Viking: Frost layers several tenths of a millimeter thick. *Lunar Planet. Sci.* XI:160–161 (abstract).

Clayton, R. N., and Mayeda, T. K. 1988. Isotopic composition of carbonate in EETA 79001 and its relation to parent body volatiles. *Geochim. Cosmochim. Acta* 52:925–927.

Clow, G. D. 1987. Generation of liquid water on Mars through melting of a dusty snowpack. *Icarus* 72:95–127.

Cutts, J. A. 1973. Nature and origin of layered deposits of the Martian polar regions.

J. Geophys. Res. 78:4231–4249.

Cutts, J. A., Blasius, K. R., Briggs, G. A., Carr, M. H., Greeley, R., and Masursky, H. 1976. North polar regions of Mars: Imaging results from Viking 2. *Science* 194:1329–1337.

Downs, G. S., Goldstein, R. M., Green, R. R., Morris, G. A., and Reichley, P. E. 1973. Martian topography and surface properties as seen by radar: The 1971 opposition. *Icarus* 18:8–21.

Downs, G. S., Reichley, P. E., and Green, R. R. 1975. Radar measurements of martian topography and surface properties: The 1971 and 1973 oppositions. *Icarus* 26:273–312.

Dreibus, G., and Wänke, H. 1985. Mars, a volatile-rich planet. *Meteoritics* 20:367–381.

Dreibus, G., and Wänke, H. 1987. Volatiles on Earth and Mars: A comparison. *Icarus* 71:225–240.

Dzurisin, D., and Blasius, K. R. 1975. Topography of the polar layered deposits of Mars. *J. Geophys. Res.* 80:3286–3306.

Erard, S., Bibring, J.-P., Mustard, J., Forni, O., Head, J. W., Hurtrez, S., Langevin, Y., Pieters, C. M., Rosenqvist, J., and Sotin, C. 1991. Spatial variations in composition of the Valles Marineris and Isidis Planitia regions of Mars derived from ISM Data. *Proc. Lunar Planet. Sci.* 21:437–455.

Fanale, F. P., and Cannon, W. A. 1971. Adsorption on the Martian regolith. *Nature* 230:502–504.

Fanale, F. P., and Cannon, W. A. 1974. Exchange of adsorbed H_2O and CO_2 between the regolith and atmosphere of Mars caused by changes in surface insolation. *J. Geophys. Res.* 79:3397–3402.

Fanale, F. P., and Cannon, W. A. 1978. Mars: The role of the regolith in determining atmospheric pressure and the atmosphere's response to insolation changes. *J. Geophys. Res.* 83:2321–2325.

Fanale, F. P., Salvail, J. R., Zent, A. P., and Postawko, S. E. 1986. Global distribution and migration of subsurface ice on Mars. *Icarus* 67:1–18.

Fanale, F. P., Postawko, S. E., Pollack, J. B., Carr, M. H., and Pepin, R. O. 1992. Mars: Epochal climate changed and volatile history. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 1135–1179.

Farmer, C. B. 1976. Liquid water on Mars. *Icarus* 28:279–28.

Farmer, C. B., and Doms, P. E. 1979. Global seasonal variation of water vapor on Mars and the implications for permafrost. *J. Geophys. Res.* 84:2881–2888.

Farmer, C. B., Davies, D. W., and LaPorte, D. D. 1976. Mars: Northern summer ice cap-water vapor observations from Viking 2. *Science* 194:1339–1341.

Goldspiel, J. M., and Squyres, S. W. 1991. Ancient aqueous sedimentation on Mars. *Icarus* 89:392–410.

Greeley, R. 1987. Release of juvenile water on Mars: Estimated amounts and timing associated with volcanism. *Science* 236:1653–1654.

Greeley, R., and Schneid, B. D. 1991. Magma generation on Mars: Amounts, rates, and comparisons with Earth, Moon, and Venus. *Science* 254:996–998.

Haberle, R. M., and Jakosky, B. M. 1990. Sublimation and transport of water from the north residual polar cap on Mars. *J. Geophys. Res.* 95:1423–1437.

Herut, B., Starinsky, A., Katz, A., and Bien, A. 1990. The role of seawater freezing in the formation of subsurface brines. *Geochim. Cosmochim. Acta* 54:13–21.

Houck, J. R., Pollack, J. B., Sagan, C., Schaack, D., and Decker, J. A., Jr. 1973. High altitude infrared spectroscopic evidence for bound water on Mars. *Icarus* 18:470–480.

Jakosky, B. M. 1983*a*. The role of seasonal reservoirs in the Mars water cycle. I.

Seasonal exchange of water with the regolith. *Icarus* 55:1–18.
Jakosky, B. M. 1983*b*. The role of seasonal reservoirs in the Mars water cycle. II. Coupled models of the regolith, the polar caps, and atmospheric transport. *Icarus* 55:19–39.
Jakosky, B. M. 1985. The seasonal cycle of water on Mars. *Space Sci. Rev.* 41:131–200.
Jakosky, B. M. 1990. Mars atmospheric D/H: Consistent with polar volatile theory? *J. Geophys. Res.* 95:1475–1480.
Jakosky, B. M. 1991. Mars volatile evolution: Evidence from stable isotopes. *Icarus* 94:14–31.
Jakosky, B. M., and Barker, E. S. 1984. Comparison of groundbased and Viking Orbiter measurements of Martian water vapor: Variability of the seasonal cycle. *Icarus* 57:322–334.
Jakosky, B. M., and Carr, M. H. 1985. Possible precipitation of ice at low latitudes of Mars during periods of high obliquity. *Nature* 315:559–561.
Jakosky, B. M., and Christensen, P. R. 1986. Are the Viking lander sites representative of the surface of Mars? *Icarus* 66:125–133.
Jakosky, B. M., and Farmer, C. B. 1982. The seasonal and global behavior of water vapor in the Mars atmosphere: Complete global results of the Viking atmospheric water detector experiment. *J. Geophys. Res.* 87:2999–3019.
Jakosky, B. M., and Haberle, R. M. 1990. Year-to-year instability of the Mars south polar cap. *J. Geophys. Res.* 95:1359–1365.
Jakosky, B. M., and Haberle, R. M. 1992. The seasonal behavior of water on Mars. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 969–1016.
Jakosky, B. M., and Muhleman, D. O. 1980. The longitudinal variation of the thermal inertia and of the 2.8 centimeter brightness temperature of Mars. *Astrophys. J.* 239:403–409.
Kasting, J. F. 1991. CO_2 condensation and the climate of early Mars. *Icarus* 94:1–13.
Kerridge, J. F. 1988. Deuterium in Shergotty and Lafayette (and on Mars?). *Lunar Planet. Sci.* XIX:599–600 (abstract).
Kieffer, H. H. 1979. Mars south polar spring and summer temperatures: A residual CO_2 frost. *J. Geophys. Res.* 84:8263–8288.
Kieffer, H. H. 1990. H_2O grain size and the amount of dust in Mars' north polar residual cap. *J. Geophys. Res.* 95:1481–1493.
Kieffer, H. H., Chase, S. C., Jr., Martin, T. Z., Miner, E. D., and Palluconi, F. D. 1976. Martian north pole summer temperatures: Dirty water ice. *Science* 194:1341–1344.
Leighton, R. B., and Murray, B. C. 1966. Behavior of carbon dioxide and other volatiles on Mars. *Science* 153:136–144.
Malin, M. C. 1986. Density of Martian north polar layered deposits: Implications for composition. *Geophys. Res. Lett.* 13:444–447.
Malin, M. C., Danielson, G. E., Ingersoll, A. P., Masursky, H., Veverka, J., Ravine, M. A., and Soulanille, T. A. 1991. In *The Mars Observer Camera*, submitted.
McCleese, D. J., Schofield, J. T., Zurek, R. W., Martonchik, J. V., Haskins, R. D., Paige, D. A., West, R. A., Diner, D. J., Locke, J. R., Crisp, M. P., Willis, W., Leovy, C. B., and Taylor, F. W. 1986. Remote sensing of the atmosphere of Mars using infrared pressure modulation and filter radiometry. *Appl. Opt.* 25:4232–4245.
McElroy, M. B., and Yung, Y. L. 1976. Oxygen isotopes in the Martian atmosphere: Implications for the evolution of volatiles. *Planet. Space Sci.* 24:1107–1113.
McSween, H. Y., Jr. 1985. SNC Meteorites: Clues to Martian petrologic evolution? *Rev. Geophys.* 23:391–416.
Mellon, M. T., and Jakosky, B. M. 1993. Geographic variations in the thermal and

diffusive stability of ground ice on Mars. *J. Geophys. Res.*, in press.
Mooney, R. W., Keenan, A. G., and Wood, L. A. 1952. Adsorption of water by montmorillonite. I. Heat of desorption and application of BET theory. *J. Amer. Chem. Soc.* 74:1367–1371.
Murray, B. C., and Malin, M. C. 1973. Polar volatiles on Mars—Theory versus observation. *Science* 182:437–443.
Nier, A. O., and McElroy, M. B. 1977. Composition and structure of Mars' upper atmosphere: Results from the neutral mass spectrometers on Viking 1 and 2. *J. Geophys. Res.* 82:4341–4349.
Olhoeft, G. R., and Strangway, D. W. 1974. Electrical properties of the surface layers of Mars. *Geophys. Res. Lett.* 1:141–1434.
Owen, T., Maillard, J.-P., deBergh, C., and Lutz, B. L. 1988. Deuterium on Mars: The abundance of HDO and the value of D/H. *Science* 240:1767–1770.
Paige, D. A. 1992. The thermal stability of near-surface ground ice on Mars. *Nature* 356:43–45.
Paige, D. A., and Ingersoll, A. P. 1985. Annual heat balance of Martian polar caps: Viking observations. *Science* 228:1160–1168.
Paige, D. A., Herkenhoff, K. E., and Murray, B. C. 1990. Mariner 9 observations of the south polar cap of Mars: Evidence for residual CO_2 frost. *J. Geophys. Res.* 95:1319–1335.
Pollack, J. B. 1979. Climatic change on the terrestrial planets. *Icarus* 37:479–553.
Pollack, J. B., and Black, D. C. 1979. Implications of the gas compositional measurements of Pioneer Venus for the origin of planetary atmospheres. *Science* 205:56–59.
Pollack, J. B., and Toon, O. B. 1982. Quasi-periodic climate changes on Mars: A review. *Icarus* 50:259–287.
Pollack, J. B., Kasting, J. F., Richardson, S. M., and Poliakoff, K. 1987. The case for a warm, wet climate on early Mars. *Icarus* 71:203–224.
Scambos, T. A., and Jakosky, B. M. 1990. An outgassing release factor for non-radiogenic volatiles on Mars. *J. Geophys. Res.* 95:14779–14787.
Schaeffer, M. W. 1990. Geochemical evolution of the northern plains of Mars: Early hydrosphere, carbonate development, and present morphology. *J. Geophys. Res.* 95:14291–14300.
Schultz, P. H., and Gault, D. E. 1979. Atmospheric effects of Martian ejecta emplacement. *J. Geophys. Res.* 84:7669–7687.
Schultz, P. H., and Lutz, A. B. 1988. Polar wandering on Mars. *Icarus* 73:91–141.
Soderblom, L. A., Malin, M. C., Cutts, J. A., and Murray, B. C. 1973. Mariner 9 observations of the surface of Mars in the north polar region. *J. Geophys. Res.* 78:4197–4210.
Sonnenfeld, P., and Perthuisot, J. P. 1989. *Brines and Evaporites*, AGU Short Course in Geology, vol. 3 (Washington, D. C.: American Geophysical Union).
Sposito, G. 1984. *The Surface Chemistry of Soils* (New York: Oxford Press).
Squyres, S. W. 1979. The evolution of dust deposits in the Martian north polar region. *Icarus* 40:244–261.
Squyres, S. W. 1989. Paper presented at Fourth International Conference on Mars, Jan. 10–13, Tucson, Ariz.
Squyres, S. W., and Carr, M. H. 1986. Geomorphic evidence for the distribution of ground ice on Mars. *Science* 231:249–252.
Tanaka, K. L., Isbell, N. K., Scott, D. H., Greeley, R., and Grant, J. E. 1988. The resurfacing history of Mars: A synthesis of digitized, Viking-based geology. *Proc. Lunar Planet. Sci. Conf.* 18:665–678.
Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A., and Bilski, K. 1980. The astronomical theory of climatic change on Mars. *Icarus* 44:552–607.

Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christian, R. P., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.

Treiman, A. H. 1985. Amphibole and hercynite spinel in Shergotty and Zagami: Magmatic water, depth of crystallization, and metasomatism. *Meteoritics* 20:229–243.

Tsoar, H., Greeley, R., and Peterfreund, A. R. 1979. Mars: The north polar sand sea and related wind patterns. *J. Geophys. Res.* 84:8167–8180.

Ward, W. R. 1974. Climatic variations on Mars. I. Astronomical theory of insolation. *J. Geophys. Res.* 79:3375–3386.

Ward, W. R. 1992. Long-term orbital and spin dynamics of Mars. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 298–320.

Wright, I. P., Carr, R. H., and Pillinger, C. T. 1986. Carbon abundance and isotopic studies of Shergotty and other shergottite meteorites. *Geochim. Cosmochim. Acta* 50:983–991.

Wright, I. P., Grady, M. M., and Pillinger, C. T. 1988. Carbon, oxygen and nitrogen isotopic compositions of possible Martian weathering products in EETA 79001. *Geochim. Cosmochim. Acta* 52:917–924.

Yung, Y. L., Wen, J.-S., Pinto, J. P., Allen, M., Pierce, K. K., and Paulsen, S. 1988. HDO in the Martian atmosphere: Implications for the abundance of crustal water. *Icarus* 76:146–159.

Zent, A. P., and Fanale, F. P. 1986. Possible Mars brines: Equilibrium and kinetic considerations. *Proc. Lunar Planet. Sci. Conf.* 16, *J. Geophys. Res.* 91:D349–D445.

Zent, A. P., Fanale, F. P., Salvail, J. R., and Postawko, S. E. 1986. Distribution and state of H_2O in the high-latitude shallow subsurface of Mars. *Icarus* 67:19–36.

Zent, A. P., Fanale, F. P., and Roth, L. E. 1990. Possible Martian brines: Radar observations and models. *J. Geophys. Res.* 95:14531–14542.

Zisk, S. H., and Mouginis-Mark, P. J. 1980. Anomalous region on Mars: Implications for near-surface liquid water. *Nature* 288:735–738.

WATER RESOURCES AND HYDROGEOLOGY OF MARS

VICTOR R. BAKER, VIRGINIA C. GULICK and
JEFFREY S. KARGEL
University of Arizona

From the perspectives of energy expenditure, ice is the most economical water resource to target for exploration on Mars. Theoretical stability criteria indicate the planetary-scale potential for ground ice poleward of about 40° latitude. Geologic indicators can constrain the exploration. Particularly useful in this regard are fluidized ejecta blankets, periglacial features, and relict glacial landforms. The relationship of such geomorphological indicators to the modern water resources is dictated by the processes responsible for water cycling in the Martian past and the extension of those processes to the present. The geomorphological evidence indicates extensive water cycling in the geologic past, involving either (1) atmospheric components leading to the formation of large ancient lakes or an ocean, or (2) hydrothermal activity. Exploration strategies can develop around the resource potential of hydrated minerals, hydrothermal systems, and ground ice based on an evolving practical experience as resources are discovered.

I. INTRODUCTION

The use of water resources has traditionally been accomplished through the problem-solving approaches of hydrological engineering (National Research Council 1991). Designs for water utilization are created by the development and application of scientific principles. For Earth the operation of a global hydrological cycle ensures that some water is always available for exploitation by these approaches. Mars is the only other planet besides Earth on which a dynamic hydrological cycle, complete with valley-forming streams (Baker 1982), is known to have operated. However, unlike Earth, the lack of basic information on the presently available Martian resource of water presents a puzzle. Because science is the intellectual pursuit adapted to dealing with such puzzles, it is a necessary prelude to effective water-resource engineering on Mars.

There is no single, valid scientific method of puzzle-solving (Conant 1951). The puzzle of water on Mars can be addressed in terms of the best available theories of water storage and transfer in the modern Martian environment consistent with whatever measurements are presently available. This approach, combined with geochemical inferences concerning the total available water inventory (Squyres 1984; Carr 1987) is reviewed in the Chapter by

Jakosky and Zent. Our approach emphasizes the analyses of specific features of the extant Martian landscape in order to interprete the causative processes required for their existence. Where such processes require causation by water activity, they constitute indicators of a past or present resource of that water. The reasoning process here is abductive, defined as the inference of causative factors from observed effects in a manner consistent with known physical principles. Abduction contrasts with both deduction and induction, which relate effects to theoretical principles consistent with known causative factors. Abductive approaches are common in geology (von Engelhardt and Zimmermann 1982), and we will use them to provide a preliminary assessment of water resources on Mars that is complementary to that in the Chapter by Jakosky and Zent.

There is another useful by-product of the geologic approach to Martian hydrology. Genetic implications do not merely attach to individual Martian landforms; they also relate to whole assemblages of landforms. The larger conceptual schemes required to produce observed landform assemblages, in turn, may be used in a theoretical mode, to predict otherwise unforeseen aspects of the resource. We use this approach to comment on resources, including those other than water, that may be implied by the geologically indicated processes of water cycling on the planet.

II. ECONOMIC ISSUES

Our major concern in this chapter is with geologic issues of resource exploration and utilization. We briefly discuss certain relevant economic and engineering issues, as these will also affect the direction of any exploration strategy.

Presupposing a demand for Martian water, there are five requirements for the existence of exploitable deposits of ice or chemically bound water: (i) the necessary physical (geologic) conditions causing the formation of water-bearing deposits must have existed; (ii) the necessary conditions for the persistence of these deposits must have existed continuously since formation; (iii) the deposit must be accessible; (iv) water must be sufficiently abundant to allow economical exploitation; and (v) the cost (in energy utilization, hardware, and dedicated human resources) of deriving water locally must be less than the cost of delivering this material from Earth.

Important geologic processes related to exploitable water resources may include pluvial (rainfall), fluvial (surface flow), and lacustrine/marine recharge of ground water aquifers, glacial deposition of ice, and hydrothermal and atmospheric hydration of minerals, among other processes.

The persistence of a hydrous deposit depends on the continuity of the conditions necessary for its stability. For hydrated silicate minerals, this is not a major limiting factor, because dehydration is dependent on thermal metamorphism by deep burial or contact with magma, and these processes are not as common on Mars as they are on Earth. The persistence of ice deposits

is more of a problem, because ice is presently unstable over large areas of Mars, and our lack of knowledge regarding environmental conditions in the geologic past makes it difficult to determine the distribution and persistence of such deposits.

Energy utilization, including energy devoted to water extraction and processing, will be an important aspect of mission design. Table I gives the energy costs of generating liquid water from ice and hydrated minerals, and for dissociating this water into hydrogen and oxygen useful for rocket propulsion and respiration. The water extraction energy includes energy required to heat the ore from 210 K (Martian surface temperature) to the melting or dehydration point of the hydrous mineral, plus the enthalpy of fusion or dehydration. Equilibrium dehydration temperatures (in K) and estimated or observed dehydration enthalpies (in kJ $mole^{-1}$) are, respectively 540, 80.8 (brucite); 773, 207 (illite); and 663, 178 (kaolinite). Average heat capacities (in $J K^{-1}$ $mole^{-1}$) over the temperature interval from 210 K to the melting or dehydration point of each phase are observed or estimated as 81 (brucite), 363 (kaolinite), 796 (illite); average heat capacity for basalt, 0.90 $J K^{-1} g^{-1}$ over the interval from 210 K to 540 K, was calculated as the average mixture heat capacity of forsterite, diopside and anorthite. Kaolinite dehydration was approximated as a 1-stage dehydration to andalusite at 663 K rather than including the intermediate formation of pyrophyllite. Mineral stoichiometries are brucite $Mg(OH)_2$, kaolinite $Al_2Si_2O_5(OH)_4$ and illite $K_2Al_4MgSi_7O_{20}(OH)_4$. Thermodynamic and phase equilibrium data sources JANAF tables: Ernst (1976), Fyfe et al. (1958), Robinson and Haas (1983).

These energy costs assume 100% efficiency and a close approach to equilibrium dehydration. Certainly neither assumption would be realized, so actual energy costs would be greater than tabulated. However, Table I provides a useful basis for energy cost comparisons.

Ice (even very impure ice) is a more energy-economical resource than hydrous rocky minerals. The energy cost of dissociating water, perhaps by electrolysis, into H_2 and O_2 is about 13,000 $J g^{-1}$. For rocket propellant use, one must add the energy needed for liquefaction. Total energy requirements thus exceed by an order of magnitude the energy to melt ice-rich permafrost. Therefore, the purity of ice deposits does not substantially impact the total energy cost of propellant production using Martian ice. On the other hand, the energy cost of water extraction from phyllosilicates is comparable to, or exceeds, the cost of, water dissociation, so the grade of chemically bound water ores is the critical factor in determining the level of energy utilization for propellant production.

Certain hydrated salts, however, contain much more water than phyllosilicates, and liberate this water under comparatively gentle heating. Probable salt hydrates on the Martian surface include epsomite ($MgSO_4$-$7H_2O$), gypsum ($CaSO_4$-$2H_2O$) and hydrohalite (NaCl-$2H_2O$). The energy cost of utilizing these minerals would approach the economy of using ice.

The energy cost of utilizing impure ores could be reduced if mechanical

TABLE I

Energy Required to Extract Martian Water

Material	Water Extraction Energy (J g^{-1})	Extraction + Dissociation Energy (J g^{-1})	Mass of Ore (g) to Yield 1 g H_2O
Ice, pure	450	13,500	1.0
Ice, disseminated, 30 vol. %	850	13,900	7.3
Ice, disseminated, 10 vol. %	1,990	15,000	25
Brucite ($Mg(OH)_2$), pure	5,960	19,000	3.1
10 mass % brucite + 90% basalt	14,600	27,600	31
Kaolinite, pure	9,490	23,000	7.2
10 mass % kaolinite + 90 % basalt	36,500	50,000	72
Illite, pure	18,200	31,700	21.8
10 mass % illite + 90% basalt	126,000	139,000	218

means are available to separate hydrous from anhydrous minerals. As hydrous phases are usually softer than typical anhydrous silicates, crushing and sifting might yield fines with a high fraction of hydrous components; or magnetic segregation may effectively separate anhydrous ferrous iron minerals from ice and most hydrous silicates.

The extractable water content of a hydrous deposit is an important factor in determining the required quantity of unprocessed ore (Table I). Utilization of phyllosilicates requires 1 to 2 orders of magnitude more ore than utilization of pure ice to yield an equivalent amount of water. In summary, as a source of liquid water, ice is considerably more economical than phyllosilicates and other hydrated minerals from the perspective of energy expenditure.

III. THEORETICAL GROUND-ICE POTENTIAL

When exploring for potential ground-ice reserves on Mars, certain criteria must be considered.

1. Where is ground ice physically stable?
2. In what geologic units would liquid water have at one time been present?
3. Is the region, unit, or landform located where water would have been confined or ponded naturally and eventually frozen in place?
4. Could the ground ice have remained there over geologic time?

A considerable effort by many researchers has provided a theoretical understanding of the stability of ice on the surface and in the regolith of Mars (see Chapter by Jakosky and Zent). We summarize the essential theoretical results in Fig. 1, based on the collective efforts of others (Farmer and Doms 1979; Zent et al. 1986; Fanale et al. 1986). This figure shows the strong tendency toward desiccation in the tropics and subtropics, and the stability of ice at very shallow depths in polar and middle latitudes. Current environmental conditions at latitudes equatorward of about 30° to 40° on Mars would result in a geologically rapid removal of ice in the near surface zone given a reasonable range of possible diffusion rates. Ice may have survived over geologically lengthy periods in this latitude belt, but only at depths on the order of a hundred meters. Access to these zones might be possible at recent landslides, which could expose deeply buried ice, or possibly through caves such as lava tubes and karst caverns in Tharsis, Elysium, Noctis Labyrinthus, and parts of Valles Marineris.

Ice is expected to have remained stable in the near surface poleward of about 40° latitude (Fig. 1). However, the topmost few centimeters to perhaps a meter, may be desiccated even at these latitudes, except for the residual north polar cap, because summer daytime temperatures exceed the water frost point. Diurnal and annual thermal waves are damped in this thin surficial layer; in the regolith below this surficial layer, temperatures poleward of 40° latitude remain continually below the frost point. At these latitudes, atmospheric water vapor diffuses into regolith pore spaces and condenses as adsorbed and frozen

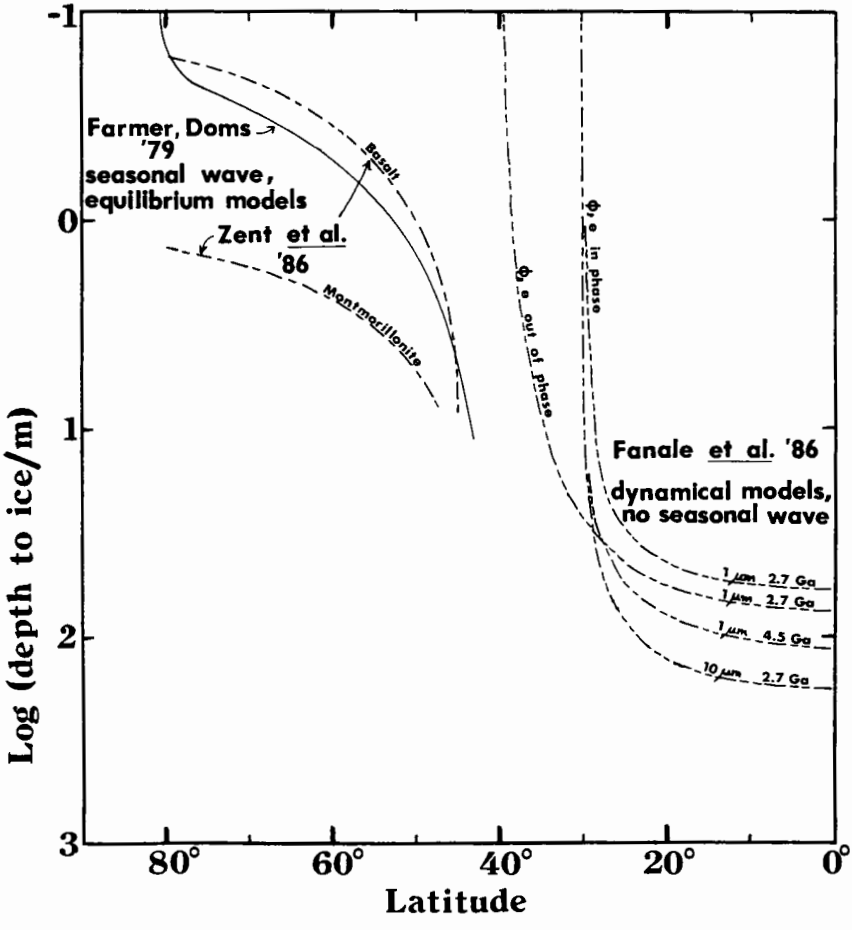


Figure 1. Theoretical regions of ground ice stability or metastability in the Martian crust. Sources are from Farmer and Doms (1979), Zent et al. (1986) and Fanale et al. (1986).

water. The time scales for diffusive water vapor transport and equilibration between the regolith and atmosphere are geologically short, thus making equilibrium a reasonable approximation. Therefore vast quantities of widely distributed, ice-rich permafrost theoretically should exist at high and middle latitudes. However, subsurface condensates would be finely disseminated through rocky impurities, increasing the energy and processing requirements for utilization of this potential resource compared to segregated ice.

In general, the higher the volume fraction of ice in a given geologic unit, the more economical the retrieval of that ice. However, the permeability and pore size of the material are also important considerations. For example, fully saturated, clays can store between 40% and 70% of their total volume

with water, sands between 25% and 50%, gravels between 25% and 40%, fractured basalt between 5% and 50% and crystalline rock between 0% and 10% (Freeze and Cherry 1979). Ice retrieval would be easiest in those units (sands, gravels and fractured basalt) which are also highly permeable, as most of the energy required in extracting the water is that used in melting the ice. Once melted, resulting water flow needs only to be collected. Although clays can potentially store significant quantities of water, only a fraction of the water contained in clays is actually free to flow. Because pore sizes are extremely small, adhesive/cohesive forces of the grains are extremely high. Therefore, leakage of free interstitial water out of clay can occur only if the clay remains fully saturated, and saturated conditions can only be maintained by active recharge. A high yield of water from clay might be most efficiently obtained by drying and vapor cold trapping, a process which would carry a significant additional energy cost of about 2400 $J\,g^{-1}$ water. Thus, a fractured basalt with a lower volume fraction of water can be a more desirable source than a higher water volume fraction of clay.

The envelope of model results shown in Fig. 1, while accounting for climatic variations expected due to the obliquity-eccentricity cycle of Mars, is limited to consideration of the range of climatic conditions during the current geologic epoch (i.e., a cold, dry Mars). Other conditions have not been incorporated, including the possible occurrence of ancient geologic/climatic conditions potentially well beyond the scope of the models and either more or less favorable to the origin and persistence of ground ice. Therefore, it is important that we seek multiple independent geologic indications of the distribution of ground ice.

IV. INDICATORS OF GROUND ICE

Visible and near-infrared reflectance spectroscopy have been used to detect water in surface material of Mars. Sinton (1967) is credited with discovering water on Mars, by observing a strong absorption near the 3 μm region. McCord et al. (1978,1982) noted absorption in the 1.4 to 2.0 μm regions and attributed these bands to the presence of water ice and mineral hydrates. Clark (1978) and McCord et al. (1982) simulated the reflectance of Mars with spectra typical of basalts and their oxidation products and added an ice spectra. These results showed that water is present on the Martian surface in a variety of phases, as ice sheets, as interstitial ice in surface regolith, as water bound in mineral hydrates. These and other historic detections of water in Martian surface material are summarized by Singer et al. (1979).

However, direct measurement of Martian ground water has not been achieved. Several regions of hypothesized near-surface liquid water were thought to have been identified using Earth-based radar data of the southern equatorial region of Mars (Downs et al. 1975). Two such regions were detected by Huguenin et al. (1979), Solis Lacus (25° S, 85° W) and Noachis-Hellespontes (30° S, 315° W). Another major region, centered at 16° S, 90° W,

as well as two smaller areas to the northwest of Hellas were reported by Zisk and Mouginis-Mark (1980,1982). These regions are characterized by very high radar reflectivities and unusual smoothness. Huguenin et al. (1979) and Huguenin and Clifford (1980) proposed that these "oasis" areas are outgassing regions of upwardly migrating brine/ice melts. While these interpretations have been contested (see, e.g., Jakosky and Farmer 1982; Jakosky 1985), recent studies by Roth et al. (1985,1986), who looked at the entire radar data set, concluded that the brine/ice melt explanation is still the most likely cause of the anomalous radar signatures. For Mars' present surface conditions, hypersaline solutions are stable at temperatures as low as 210 K (Brass 1980), but Zent et al. (1990) consider shallow brines out of equilibrium with the environment and unlikely as causes of the unusual radar reflectivities.

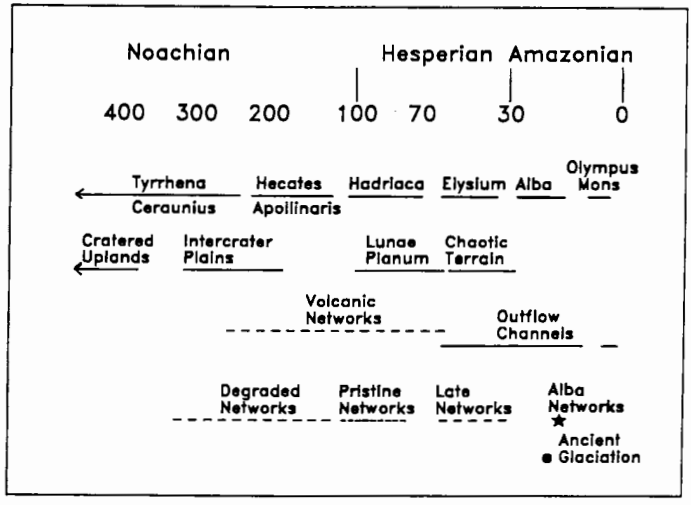


Figure 2. Relative ages of different age surfaces on Mars. Solid lines are crater densities (number of craters greater than 8 km diameter per 10^6 km^2) taken from Barlow (1988). Dashed lines represent the age of the surface on which the various valleys have formed. The age of the Alba networks is the age of the surface on which the valleys have formed. Both the age of the Alba networks and the period of ancient glaciation were originally calculated from craters >1 km and were extrapolated to the Barlow relative age data. Figure from Gulick and Baker (1990*b*), as modified from Gulick and Baker (1990*a*).

Indirect evidence abounds for Martian water. This water seems not only to have existed at the surface several times during Mars' geologic history, but it was actively being recycled for geologically significant time periods. Ice deposits subsequently produced by this activity, provide an especially promising resource. Such deposits may exist on a variety of scales.

A. Channels and Valleys

The formation of valley networks on surfaces of all ages (Fig. 2) is perhaps the best evidence for hydrological cycling of water. Dynamical cycling of water within the surface and subsurface environments on the order of at least 10^5 yr is required to form such valleys. High latitude fluvial deposits may contain embedded ice. Valleys which debouch into impact craters would be of special interest. The fractured crater floor, in combination with the inevitable accumulation of various sedimentary and fluvial deposits would have collected and stored water as liquid and later as ice. The formation of the outflow channels would have required catastrophic discharges of at least 10^7 m^3 s^{-1} lasting several days to several tens of weeks (Baker et al. 1991). These outflows emptied into the northern plains, perhaps forming a temporary ocean (Gulick and Baker 1989*a*; Parker et al. 1990*a*). Given the correct conditions, some of this water may have remained as ice in the higher latitudes. Detailed paleoflood studies of these channels (Tanaka and Chapman 1992) suggest multiple flood episodes. If this was the case, then much of this water eventually would have infiltrated into the subsurface and formed multiple ice-rich layers. Sedimentary and volcanic debris would have been deposited between each episode. Such mantling units would have helped to reduce transfer of water vapor to the atmosphere and confine water from subsequent flood episodes to shallow, subsurface locations.

Understanding Mars' fluvial and paleohydrological history is critical to the eventual location of ground water or ice reserves. Such studies yield information about the source of the water, how it was deposited, and whether the hydrogeological conditions were such that water was likely to flow, become trapped, eventually frozen, and remain at this location in the present climate.

B. Fluidized Ejecta Blankets

Nearly every impact crater on Mars with a diameter between 4 and 40 km and having fresh ejecta is distinguished on the basis of ejecta morphology from craters on the Moon (Carr et al. 1977; Strom et al. 1992). Martian craters have morphologies indicating fluidized, ground-hugging emplacement of ejecta, while lunar craters have ejecta clearly emplaced by ballistic trajectories. Fluidization of Martian ejecta has been explained by (1) entrainment of atmospheric gases in ejecta (Schultz and Gault 1979,1982,1984); and (2) entrainment of volatile-rich rock in the ejecta (Carr et al. 1977; Mouginis-Mark 1979). In the latter case, the distribution of fluidized ejecta craters would provide valuable observations constraining the distribution of subsurface volatiles.

Several morphological aspects of fluidized ejecta are dependent on latitude (Fig. 3). Such a correlation favors evidence supporting fluidization primarily by condensed volatiles, rather than by atmospheric gas, although the atmosphere would probably contribute. These latitudinal dependencies probably relate to a temperature control on the volatiles, but it is not known whether this concerns the physical state (solid, liquid, or adsorbed), abun-

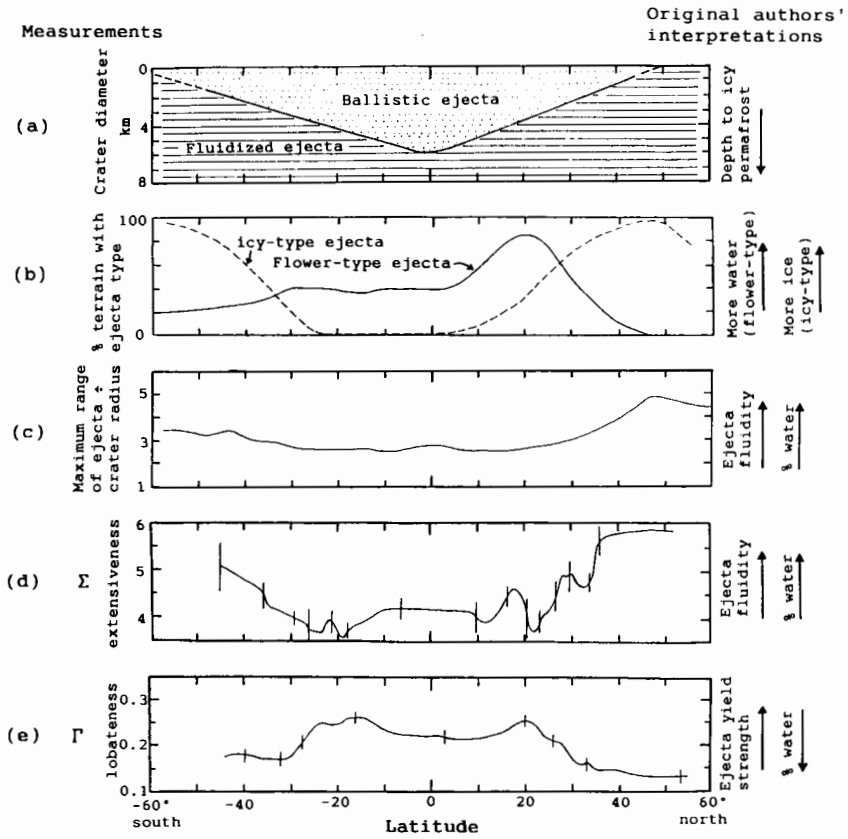


Figure 3. (a) Crater transition diameter between craters displaying fluidized ejecta and those lacking fluidized ejecta as a function of latitude, interpreted to indicate the depth to ice-bearing strata (after Kuzmin 1980). (b) Distribution of "flower-type" (= "water-type") and "icy-type" fluidized ejecta as a function of latitude (after Johansen 1979). (c) Maximum radial extent of fluidized ejecta normalized by crater radius, as a function of latitude (after Costard 1989). (d) Moving box-car average extensiveness (defined in text) of fluidized ejecta as a function of latitude, and representative 1-σ error bars showing statistical deviation of the mean (from Kargel 1989 and unpublished data). (e) Moving box-car average lobateness (defined in text) of fluidized ejecta as a function of latitude. Data adjusted to eliminate crater size dependency before taking the average (after Kargel 1989).

dance, molecular form (CO_2 or H_2O), and/or depth of the responsible volatile agents in the upper crust. However, Mouginis-Mark (1987) noted the presence of small channels draining some fluidized ejecta blankets, and concluded that water at least contributed to ejecta fluidization.

Mouginis-Mark (1979), Johansen (1979), Costard (1989), Barlow and Bradley (1990), and others have presented qualitative morphologic classifications of fluidized ejecta, documenting a wide range of morphologies. Kuzmin

(1980), Kargel (1986,1989), Costard (1989), and Bridges and Barlow (1989), among others, attempted to quantify the morphology and distribution of fluidized ejecta. These studies are briefly summarized below.

Kuzmin (1980) observed a relationship between latitude and the minimum sizes of craters displaying fluidized ejecta. He interpreted this relationship to indicate a latitudinal control on the depth to the ice-bearing permafrost table, this depth shallowing to higher latitudes (Fig. 3a) in qualitative agreement with theoretical expectations (Fig. 1).

Johansen (1979) observed a strong latitudinal influence based on a subjective classification of crater ejecta morphology (Fig. 3b). "Flower-type" ejecta, named for a resemblance to petalous composite flowers, are characterized by generally high lobateness, and "icy-type" ejecta are characterized by generally low lobateness. Johansen noted that lobateness varies inversely with latitude. This relation was also confirmed by Kargel's (1986,1989) and Costard's (1989) subsequent analyses.

Mouginis-Mark (1979) concluded that the so-called "pancake" class of Martian crater ejecta exhibits a pronounced latitudinal control. As "pancake" ejecta probably are erosional residual forms and may be excluded from a discussion of fluidization, one could conclude that fluidized ejecta morphologies do not display a latitudinal control. However, the distributions of several of Mouginis-Mark's classes of fluidized ejecta do seem to show considerable latitudinal control (see Mouginis-Mark 1979, Fig. 4).

Costard (1989) devised a very simple classification scheme, and noted a tendency for highly lobate flower-like ejecta to occur at low latitudes and more circular ejecta to occur poleward of about 30°. Costard also noted a tendency for low-latitude flower-type ejecta blankets to be emplaced closer to the crater rim than the more distant circular ejecta found at high latitudes (Fig. 3c), an aspect confirmed in Fig. 3d.

Kargel (1986,1989) defined two dimensionless parameters, the ejecta "lobateness" and "extensiveness," and found that these aspects vary with latitude and crater size. Lobateness Γ is a measure of the sinuosity of fluidized ejecta perimeters, and is defined as

$$\Gamma = \frac{P}{(4\pi A_e)^{\frac{1}{2}}} - 1 \tag{1}$$

where P is the ejecta perimeter, and A_e is the area enclosed by the trace of the perimeter. A circular ejecta blanket has $\Gamma = 0$. Extensiveness is the size of the ejecta blanket normalized by the size of the crater, thus providing an indication of the relative fluidity of the ejecta. Σ is defined as

$$\Sigma = \frac{(A_e - A_c)}{A_c} \tag{2}$$

where A_c is the area enclosed by the crater rim.

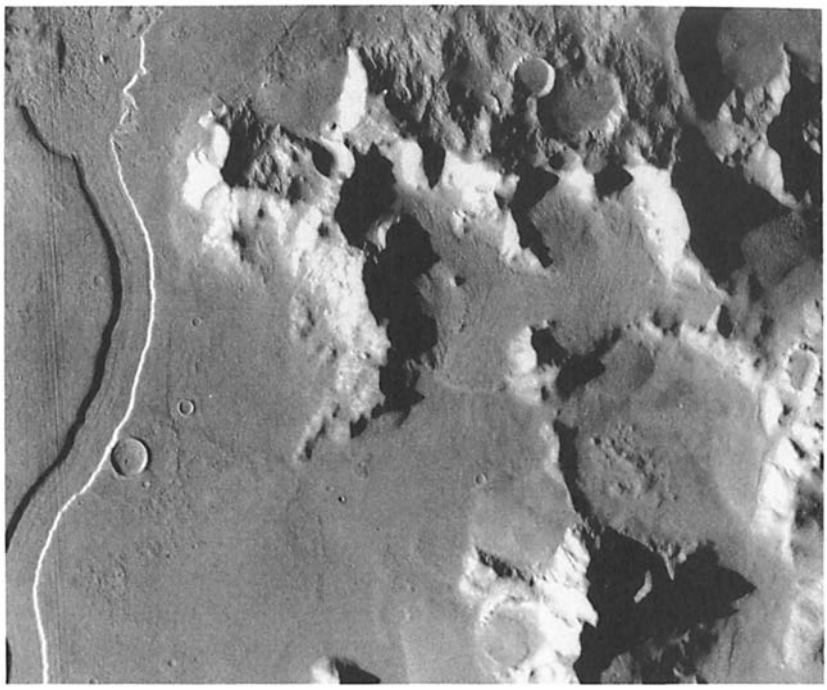


Figure 4. Lobate-debris apron and cirque-like source east of Hellas.

Lobateness (Γ) is highly dependent on crater size as shown by Kargel (1986), and supported by Bridges and Barlow (1989). After accounting for the crater size dependency observed for Γ, significant latitudinal dependencies are revealed in Kargel's data. Figures 3d and 3e show filtered Γ and Σ data as functions of latitude for 361 craters larger than 9 km. The crater sample is distributed globally, although Hesperian Ridged Plains are over-represented and cratered uplands are under-represented.

Extensiveness Σ shown in Fig. 3d, reflects the relative mobility, and apparently the rheology, of the ejecta. The average Σ of ejecta is a minimum in tropical regions of Mars and increases between latitudes of $\pm 20°$ and $\pm 40°$, which seems to imply a corresponding increase in the water content of ejecta and of the Martian upper crust towards higher latitudes. This transition is consistent with the theoretical conclusions and interpreted observations of others (Chapter by Jakosky and Zent; Farmer and Doms 1979; Zent et al. 1986; Fanale et al. 1986; Johansen 1979; Kuzmin 1980; Squyres and Carr 1986).

Figure 3e shows a latitude-lobateness relationship inverse to that for Σ, with low-latitude ejecta tending to be more lobate than middle-and high-

latitude ejecta. Johansen (1979) noted that her subjectively defined "flower-type ejecta," which are characterized by high lobateness and generally occur at low latitudes, also tend to display prominent flow-front ridges. These ridge features are absent or subdued in most of her "icy-type" ejecta located at higher latitudes.

Barlow and Bradley (1990) observed a latitudinal control on the frequency of occurrence of single- and double-lobed fluidized ejecta blankets. They speculated that single-lobed ejecta blankets are formed by impacts into ice-rich targets while double-lobed ejecta are formed by the penetration of the impactor beneath the cryosphere into the ground-water zone. A poleward deepening of the melting geotherm would then explain their observations. Bridges and Barlow (1989) concluded that lobateness increases with crater size, but they did not replicate Kargel's observation that lobateness varies significantly with latitude.

Lacking ground truth, the mechanistic interpretation of fluidized ejecta morphologies is rather speculative. However, in a general sense, it seems probable that the varied appearance of ejecta blankets relates to ejecta rheologies during emplacement. Latitudinal controls on ejecta morphologies would probably reflect the abundance, depth, and/or physical state of subsurface water/ice. The preferred occurrence at low latitudes of ejecta having low extensiveness, high lobateness, and prominent distal ridges is best explained by a relatively low water content, either due to low water concentrations in the regolith, or to a deep water/ice table. The preferred occurrence at middle and high latitudes of ejecta characterized by high extensiveness, low lobateness, and subdued distal ridges is consistent with relatively high water contents, due to either high concentrations of water/ice, or to a shallow water/ice table. Low water contents at low latitudes might produce ejecta with high effective viscosity (hence, low extensiveness) and high effective yield strength. This would cause ejecta to interlock suddenly and pile up at the terminus of the ejecta flow, thus forming a distal ridge. High water contents would result in lower viscosities, greater extensiveness, and a smaller tendency for the ejecta to cease motion suddenly at low strain rates. The high lobateness of low-latitude ejecta could be a result of small heterogeneities in liquid content manifested in considerable differences in effective viscosity, a phenomenon observed at high solid/liquid ratios in laboratory suspensions of crystals in liquid (Kargel et al. 1991) and responsible for the lobate aspect of many lava flow margins.

In summary, the morphologies of fluidized ejecta appear to vary with latitude, thus supporting a relationship of ejecta fluidization to some aspect of the distribution, abundance, and/or physical state of subsurface volatiles, probably water. *In-situ* investigations are needed, however, to resolve important controversies concerning the physical mechanism of ejecta fluidization and the cause of morphologic variations.

C. Periglacial Ice

The theoretical possibility that periglacial morphogenesis may have been important in the geologic evolution of Mars was considered likely even before the Mariner 9 mission (Wade and de Wys 1968). Periglacial interpretations based on Viking spacecraft observations drew widespread acceptance during the 1970s and 1980s (Carr and Schaber 1977; Carr et al. 1977; Lucchitta 1981; Squyres and Carr 1986). Pingos, polygons, striping, rock glaciers and a host of other classic periglacial features have been proposed. Squyres and Carr (1986) mapped the global distribution of several classes of hypothesized periglacial landforms and found a striking latitudinal control on their distribution. These features, including "lobate debris aprons" discussed below, occur primarily at latitudes greater than about $\pm 30°$, and are particularly concentrated between 30° and 50°. This distribution is rather consistent with the theoretical distribution of stable ground ice shown in Fig. 1 (Farmer and Doms 1979; Zent et al. 1986; Fanale et al. 1986).

Figure 4 shows a "lobate debris apron" on Mars in a mountainous region east of Hellas. This feature displays a sharp flow front and pronounced longitudinal ridges and furrows, indicating down-slope motion. The source of the debris flow is in a compound cirque-like headwall region on the south (pole-facing) side of the mountain; the north-facing slope of this mountain is heavily gullied and lacks cirque-like structures. Many similar debris aprons in this region tend to be associated with the south-facing slopes of major massifs. These landforms have been interpreted as rock glaciers (Squyres and Carr 1986).

Rock glaciers are bouldery accumulations of ice-cored or ice-cemented debris slowly flowing downslope glacier-style (Vitek and Giardino 1987). Terrestrial rock glaciers are usually associated with periglacial alpine environments in regions too dry to support true glaciers but cold enough to sustain perennial ground ice (Corte 1987). Rock glaciers produce flow features commonly associated with alpine glaciers, such as longitudinal medial moraine-like ridges, transverse compressional ridges, and abrupt flow fronts. The resemblance to true dirty ice glaciers is often quite remarkable, and in some regions there is a continuum between true glaciers and rock glaciers (Corte 1987; Gardner and Bajewsky 1987).

Active rock glaciers have a core or lenses of nearly pure ice beneath 30 to 300 cm of surficial ice-free blocky rubble (Hassinger and Mayewski 1983; Johnson and Lacasse 1988). Terrestrial rock glaciers frequently show preferred orientations, often developing only on one side of a mountain range. Important controlling factors include minimization of solar insolation (Ellis and Calkin 1979; Hassinger and Mayewski 1983), prevailing wind direction, susceptibility of local rock types to frost shattering (Johnson and Lacasse 1988), and patterns of vegetation growth. Rock glaciers become inactive piles of pure rocky debris, often maintaining flow morphologies, if periglacial conditions give way to more temporate conditions causing the ice core to melt.

Terrestrial rock glaciers are often associated with glacial cirques and small cirque glaciers (Morris 1987; Gardner and Bajewsky 1987; Johnson and Lacasse 1988; Hassinger and Mayewski 1983), giving rise to the hypothesis that rock glaciers, or at least some of them, have direct glacial origins; in essence, according to this hypothesis, rock glaciers are dying debris-rich glaciers. The other leading end-member hypothesis is that rock glaciers are purely periglacial landforms; they are fed by multiple rockfalls and rock avalanches, and mobilized by the creep of interstitial ice formed by refreezing of percolating snow melt (see reviews by Vitek and Giardino [1987], and by Corte [1987]).

There is no evidence that the debris apron in Fig. 4 has been impact cratered or has suffered desiccation or massive melting (as might be revealed by kettles), or otherwise is degraded. The pristine appearance of this debris flow is consistent either with continuing activity, or with extremely low rates of degradation. Terrestrial rock glaciers and glacial analogs suggest that this feature has a core or lenses of ice as shallow as one or two meters beneath a bouldery surface. This debris apron occurs at 42° S latitude on the shady side of the mountain, within the region where ice theoretically should be stable within a meter of the surface (Fig. 1). Most other debris aprons in this region south of 42° S similarly show few signs of degradation, whereas those north of here are often heavily cratered or pitted. Thus, it appears that in this region ground ice may still be present very close to the surface south of 42° S, particularly on the south sides of major massifs, while shallow ice was once present but no longer is north of this latitude.

Mars has many similar lobate debris aprons and related features widely distributed in the latitude belts from 30 to 50° N and S (Squyres and Carr 1986). One can only highly recommend such features for active *in-situ* exploration for ice by penetrators or rovers. Gamma-ray emissions from hydrogen (water) should be observable by the Mars Observer gamma-ray spectrometer if ice exists within a meter of the surface over an area on the order of 10^4 km^2. No individual lobate debris aprons are this large, but in several regions multiple coalesced debris aprons blanket even larger areas and would be suitable targets in remote exploration for ice.

D. Glacial Ice

Periglacial interpretations of Mars suggest the added possibility that glaciation also may have been an important process under certain conditions. The possible role of glaciation as seen in the geologic record of Mars has seemed more doubtful than periglacial activity to many observers, but there were early suggestions of possible glacial processes (Faul 1973; Carr et al. 1980; Howard 1981; Lucchitta 1981,1982). More recently, additional evidence has been presented favoring a previously unrecognized major role for glaciation in Martian geologic history (Kargel and Strom 1990,1992). Baker et al. (1991) integrated this hypothesis into a global geologic conceptual scheme as outlined in Sec. V.

The geomorphological characteristics of the proposed Martian glacial landforms are consistent with terrestrial glacial features. Moreover, the regional pattern of the landforms, in spatial association with one another, is so similar to that of terrestrial glacial associations that a similar cause is indicated.

Glacial landforms on Mars may yield important quantities of ice. For hydrological purposes, such landforms can be composed of unsorted, ice-deposited sediments (e.g., till) or sorted, water-deposited sediments. Because the postulated glaciation on Mars (Kargel and Strom 1990,1992) is a relict event, one needs to look for deposits where significant quantities of water may have once flowed but subsequently became trapped by less permeable confining areas, or where stagnant glacial ice may have been buried in regions of active sedimentation. Such areas could now contain important fossil ice reserves.

On Earth fossil Pleistocene ice remnants are absent in most subarctic localities because of perennially warm postglacial conditions. However, Arctic periglacial regions of former glacial cover often contain abundant subsurface fossil ice. The current periglacial environment at high and middle latitudes on Mars suggests that ancient glacial ice deposits might exist in those regions. However, besides what current conditions allow, a limiting factor is whether past conditions during deglaciation and the immediate postglacial period permitted the survival of ground ice anywhere on the planet. If not, then segregated ice would have completely melted and entered the ground water system where it may have refrozen as disseminated ice in the subsurface. If deglaciation occurred under periglacial conditions, then there is a strong likelihood that segregated ice would have survived to the present time. An evaluation of the ice potential of glacial sediments requires an understanding of the conditions during and following deglaciation, and therefore an understanding of the genesis of glacial landscapes.

Substantial progress has been made in regional glacial geomorphological interpretations, mapping, and chronostratigraphic correlations, although many results are preliminary (Kargel and Strom 1990,1992). Certain important constraints and probable relationships with possible resource implications have emerged. Several Martian surface features morphologically resemble terrestrial landforms associated with large-scale melting of glacial ice, including outwash plains, tunnel valleys (subglacially eroded channels), eskers (sinuous ridges of sediment deposited by sub- or englacial streams), kames (ice-contact lake deposits), and kettles (collapse pits formed by disintegration of buried glacial ice). Such landforms are common in deglaciated subarctic regions of former Pleistocene ice cover, and seem to imply that similar conditions of active continental glaciation followed warm conditions which favored massive and rapid melting of the ice sheet.

The analogous terrestrial glacial landscapes, dominated by glaciofluvial, glaciolacustrine, and ice disintegration features, are often regions where no fossil ice is preserved today. However, many recent glacial sediments deposited near the margins of active or stagnant glaciers contain cores or segre-

gated lenses of ice (Clayton 1964; Driscoll 1980). These ice cores ablate when perennial conditions develop that are unfavorable to the persistence of ice, but the ice cores may persist indefinitely as long as humid periglacial conditions persist. Similarly, sediments containing fossil glacial ice could currently exist on Mars at middle to high latitudes under periglacial conditions.

The presence of eskers, kames, outwash plains and similar features, while clearly documenting deposition by glacial melt water, are not necessarily indicative of conditions unfavorable to the persistence of relict ice cores in these deposits. This is because melt water features can form during seasonal surface runoff during the warmest summer afternoons or by sudden and sometimes catastrophic releases of subglacially-stored water resulting from subglacial geothermal melting. Ice cores and stagnant glacial ice may be preserved indefinitely beneath the penetration depth of diurnal or seasonal thermal wave. Thus, a test for conditions prohibiting relict ice is to look for indicators of deep subsurface melting or sublimation, rather than features merely indicating deposition by melt water. Deep subsurface melting may be revealed by extensive pitting of the surface.

Kettles are common through much of Canada, particularly Saskatchewan, and form most of Minnesota's "10,000 lakes." Kettles are most commonly 100 to 1500 m across and 3 to 30 m deep, but may attain diameters greater than 10 km and depths over 100 m. The dimensions of kettles depend on the sizes of buried ice blocks and the thickness of the overlying glacial sediment blanket. The larger ice blocks may create their own landform systems, including eskers, kames, outwash plains, and glaciotectonic structures such as ice-squeezed sedimentary ridges deformed around the periphery of the stranded ice mass (Kaszycki 1987).

Figure 5 shows one particularly striking terrain in the Argyre Basin of Mars dominated by several large depressions, interpreted as large kettles. Several depressions in this area have single or double inner terraces probably formed by deposition of sediment (kames) in ice-marginal lakes as the ice block melted; other depressions have sharp raised rims possibly formed by glaciotectonic squeezing of soft, wet sediment by the weight of the ice block. One large depression in this area (not in the image shown) has an associated channel, probably eroded by melt water. The large terraced deposits shown in Fig. 5, exterior to the large depressions, could be lacustrine sediments deposited in a proglacial lake in ice-free areas between wasting ice masses, then partially eroded by wave action; the multiple terraces probably reflect a fluctuating water depth. Terracing related to sudden variations in base level is a ubiquitous feature of terrestrial proglacial lacustrine and fluvial environments (Thwaites 1926; Grout et al. 1959; Sollid and Reite 1983; Kaszycki 1987). If this interpretation of Fig. 5 is correct, then it implies massive melting of glacial ice, further implying high temperatures for at least part of the year. This in turn indicates that there was a deficit between winter snow accumulation and summer melting, which is consistent with evidence based on other hypothesized glacial landscapes in Argyre (Kargel and Strom

Figure 5. Terrain on floor of Argyre basin dominated by large, irregular depressions, interpreted as kettle basins. Note also terracing interior and exterior to depressions. Scene width approximately 50 km.

1990,1992). However, these observations do not prohibit the possibility that parts of Argyre were deglaciated under periglacial conditions. This may have allowed remnants of glacial ice to persist under an insulating layer of sediment. For example, in Fig. 5 a local periglacial microclimate may have prevailed along the south-pole-facing walls of the large depressions.

Figure 6 shows another part of Argyre dominated by an intensely pitted landscape similar in appearance to the kettled ground moraine in Saskatchewan, Canada. These depressions range from less than 300 m to about 1500 m in diameter, suggesting thawing to a depth of several tens of meters or more. Thus even though this plains material may have been ice rich at one time, it probably lacks any zones of segregated ice today.

A glacial deposit lacking kettles does not necessarily indicate that the deposit presently contains or ever contained ice. However, if there are no kettles, and if permafrost conditions are currently prevalent in the region, then the deposit may warrant further exploration for its ice potential.

Figure 7 shows an example of another probable glacial landform in Ar-

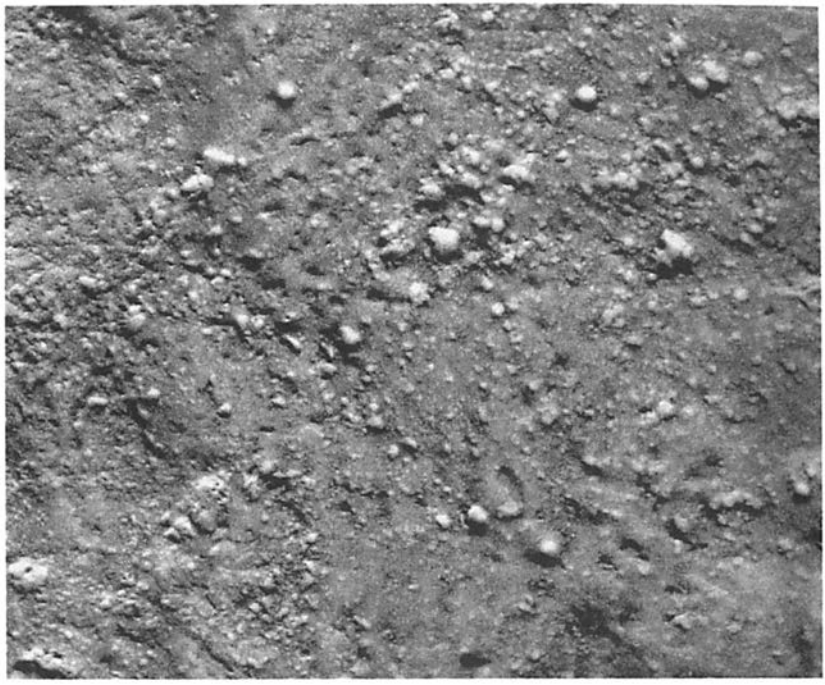


Figure 6. Terrain on floor of Argyre basin bearing large numbers of small irregular depressions, interpreted as kettles. Scene width approximately 50 km.

gyre, interpreted as an esker. Eskers often exhibit morphological characteristics resembling fluvial systems including meandering habits and anastomosing, tributary, and distributary patterns. The characteristics of the ridge system in Fig. 7 compare favorably with certain terrestrial eskers (Kargel and Strom 1992). Other associated glacial-like landforms are associated with this ridge system, supporting the esker hypothesis. Terrestrial eskers commonly are formed with cores and lenses of segregated ice, and we expect the same would be true of Martian eskers. Thus, as long as periglacial conditions prevailed during deglaciation, eskers such as the ones in Fig. 7 would be appropriate targets for exploration.

Most terrestrial eskers occur in subarctic regions now characterized by temperate climates. Therefore, any ice core which they may have previously had long ago melted. The melting of ice cores in eskers often results in the formation of kettles and longitudinal troughs running along the esker crest, producing a double- or multiple-ridged profile (Price 1973). Figure 8 shows a part of Argyre's esker system having double-ridged profiles, and according to Price's interpretation of similar terrestrial eskers, these eskers should no

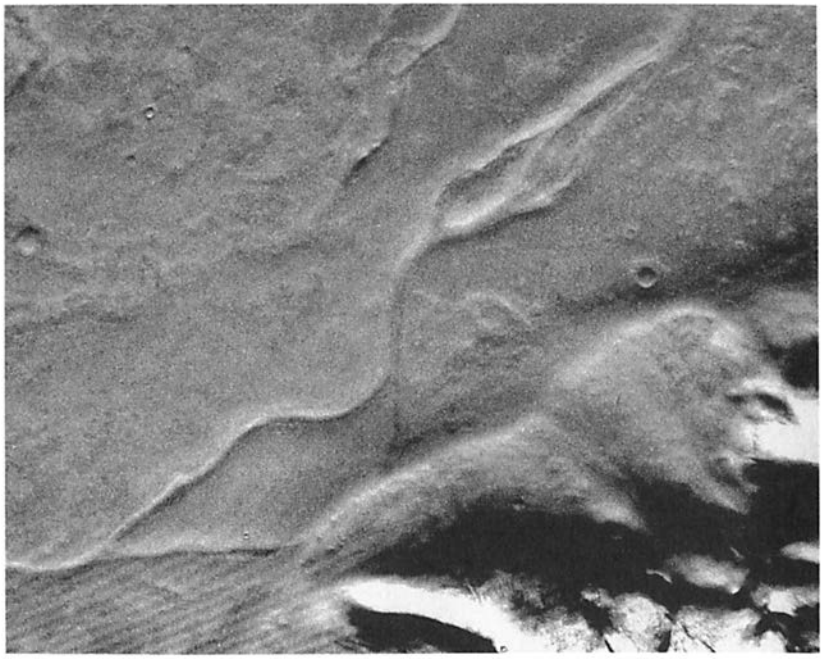


Figure 7. Terrain on floor of Argyre basin characterized by smooth plains deposits interpreted as proglacial lake sediments, and anastomosing ridges interpreted as eskers. Scene width approximately 50 km.

longer contain segregated ice.

In summary, Mars exhibits a diversity of probable glacial landforms, some of which may still contain ice. We recommend considering some of these features for active *in-situ* and remote exploration for ice. Evaluation of possible exploration targets will require reconstruction of the physical environment during and subsequent to deglaciation by a careful analysis of regional landform patterns and of individual landforms.

V. CONCEPTUAL SCHEMES

The extensive evidence of past water and ice processes on Mars requires an explanation. The formation of channels and valleys, the growth and dissipation of glaciers, and the emplacement and wastage of ground ice all indicate dynamic transfers of water on a planetary scale. On Earth such transfers occur through the hydrological cycle. Mars is clearly not now experiencing an Earth-like hydrological cycle of any significant magnitude, but the

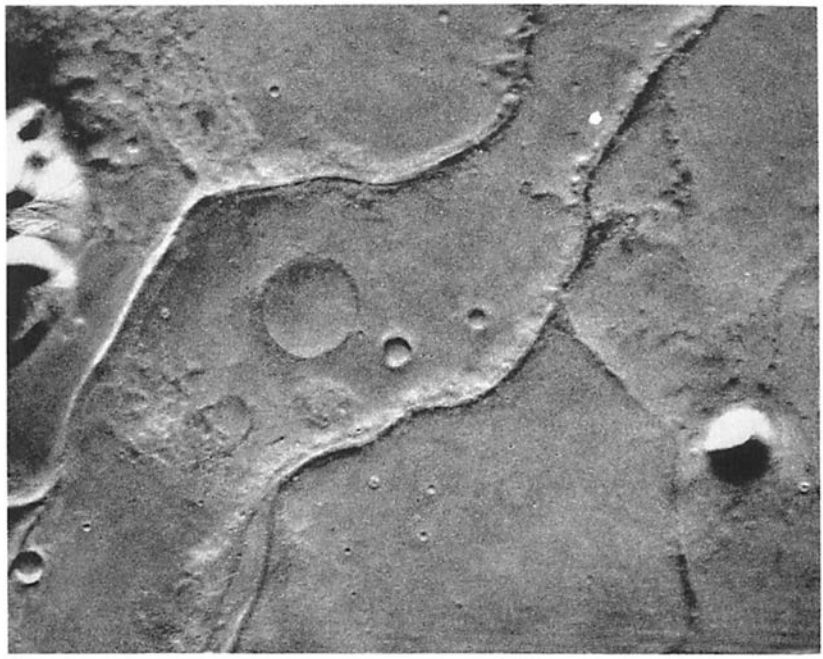


Figure 8. Terrain on floor of Argyre basin characterized by smooth plains and esker-like ridges. The ridges are often doublets in this area, possibly formed from melting an ice core.

geomorphological evidence is equally clear that present conditions are inappropriate to cause the abundant, relict aqueous landforms. We propose that the widespread geologic water/ice indicators described above are best conceived as part of an overall genetic scheme of planetary evolution, rather than as isolated, unrelated occurrences. While there is yet no scientific consensus for such a Martian scheme, it is clear that the real causative elements, if known, will be profoundly important in any resource development strategy. The plate-tectonic scheme of Earth evolution illustrates this significance for terrestrial mineral and petroleum exploration. The schemes described below are illustrative, and will certainly require confirmation by critical scientific testing before such concepts can be relied upon for resource utilization.

A. Ancient Ocean-Land-Atmosphere Cycling

Baker et al. (1991) inferred that the ponded water of the northern Martian plains, named Oceanus Borealis, was a key element of a cyclic causative system of Martian planetary water transfers. In its preliminary form, their

model involves the concentration of volcanism in the Tharsis area during post-Noachian Martian planetary history. Massive and rapid emplacement of magma beneath this domical uplift, perhaps analogous to terrestrial rift volcanism, drove out huge quantities of water from subsurface aquifers confined by ground ice. Large amounts of CO_2 were rapidly released by volcanic degassing. Driven by volcanic heat in a massive hydrothermal system, cataclysmic water flows emanated from the equatorial zone of fractures peripheral to the Tharsis Bulge. The huge water volume concentrated in the fractures, plus ponded water in the Valles Marineris, all became rapidly transferred to the Northern Plains. Subsequent evaporation of water and sublimation of the Oceanus Borealis ice cover should have occurred at atmospheric pressures elevated well above present conditions. Cataclysmic ocean formation would produce a transient greenhouse atmosphere of water vapor and carbon dioxide. Sources of CO_2 would include clathrate from the melted permafrost (Milton 1974), dissolved magmatically derived gas in the ground water, and adsorbed gas in the regolith inundated by Oceanus Borealis. As global warming occurred by the combined effects of these gases, other water frozen in permafrost would be released to the Oceanus Borealis. Possible release of adsorbed CO_2 in the previously cold regolith could have provided an additional feedback mechanism. The resulting maritime climate would have made precipitation possible. This maritime climate was probably relatively short-lived, similar to terrestrial glacial climates, and may have been modulated by cycles of planetary obliquity and orbital eccentricity.

The general process envisioned is a complex set of interactions between the Martian lithosphere and atmosphere operating over long time scales throughout the 4.5 Gyr of Martian history. The cyclic sequences of events evolved through time, such that the latest aqueous epochs were relatively small, yielding "lakes" rather than "oceans." The Amazonian paleolake of the Mare Elysium Basin (Scott and Chapman 1991) would be an example of such an event. The smaller magnitude of these events would be paralleled by smaller modifications to the Martian climate, a chief response perhaps being the formation of inactive polar layered deposits and small ice caps rather than massive glacial ice sheets.

B. Hydrothermal Cycling of Water

Hydrothermal systems are a common, natural consequence of volcanic, impact or tectonic (e.g., faulting, rifting) processes in permeable (permeability, $k > 10^{-14}$ cm^2; Norton1984), fluid-rich rock. On Earth, fossil hydrothermal systems are a major source of valuable ore deposits. Active systems provide a significant source of thermal energy for several areas. Examples include Wairaki geothermal field, New Zealand and Krafla geothermal field, Iceland. Salts, metals, and gases dissolved in hydrothermal fluids are important by-products of active regions. However, another important "resource" of hydrothermal regions is water. Water is the medium in which the thermal energy, salts, metals and gases are transported to the surface environment.

Because Earth has an abundant atmospheric supply, this water is generally not considered a resource. However on Mars, which today lacks such an abundant supply, active hydrothermal systems could provide a valuable source of near-surface water.

Hydrothermal systems are capable of circulating and focusing large quantities of mineral-rich fluids into narrow effluent zones for prolonged periods. Thermal perturbations in the subsurface induce density changes in the fluid flow field and result in the migration of surrounding fluids toward the heat source. Buoyancy forces then transport these waters to the surface and near surface environment, producing seeps and springs. The magmatic intrusion producing hydrothermal circulation cools by conductive processes until it solidifies. Upon solidification, the intrusion fractures and continues cooling predominantly by convective processes as ground water circulates through the intrusion. Hydrothermal circulation associated with volcano formation can last for several million years or more.

Similarly, many geologic features on Mars (e.g., volcanoes, impacts, tectonic features and the extensive lava flows and possible igneous sill intrusions associated with intercrater plains formation) would have produced hydrothermal activity if the features formed in a permeable medium (e.g., megaregolith or permeable basalt) which contained water or ice (Fig. 9). The high mineral content of hydrothermal waters could depress the freezing point of water in the Martian environment, thus enabling these fluids to flow longer distances before freezing. Partial release of warm hydrothermal water to the atmosphere may produce local microclimatic effects as it has in the Wairaki geothermal system, New Zealand (Elder 1981). Such an environment may further prolong surface flow. In any event, ice-covered rivers (Wallace and Sagan 1979; Carr 1983) of hydrothermal water would eventually form with warm basal water slowly cooling as it flowed away from the source area (Brakenridge et al. 1985).

In the absence of an ancient global atmospheric hydrological cycle capable of bringing sufficient precipitation to valley headwaters, hydrothermal systems might have played a major role in valley formation on Mars. There appears to be a correlation between the presence of valleys and the formation of geologic features capable of producing sizable hydrothermal activity (Gulick et al. 1991). Valleys have developed on several volcanoes near large impact craters, and in the intercrater plains regions. The degree of valley development in these regions appears to be proportional to the expected magnitude and duration of hydrothermal activity, with impact features having the least developed valleys and volcanoes and intercrater plains having the best developed (Gulick and Baker 1990*b*). However, surface lithology would have been a major control as to the type of valley formed (Gulick and Baker 1989*a*). If the surface was covered by a fairly erodible, relatively less permeable layer (e.g., ash or fine-grained aeolian material), then runoff valleys could have formed. However, if the surface was highly permeable (i.e., basalt flows), water would have quickly infiltrated and recharged the near-surface aquifers,

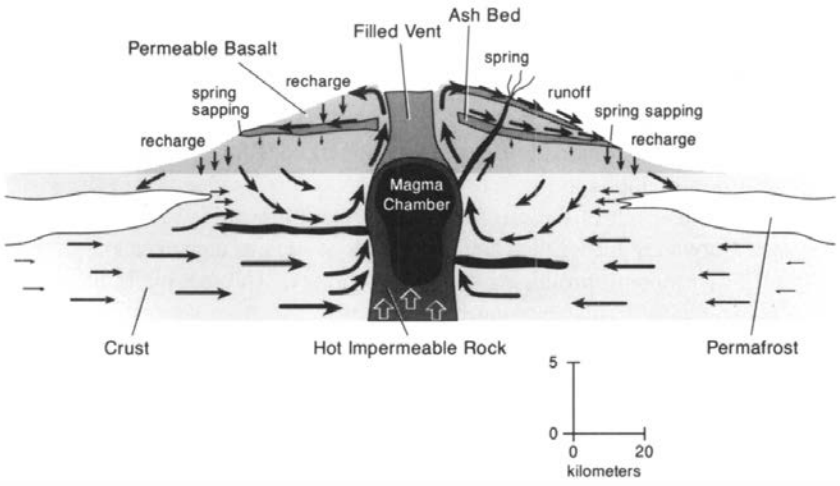


Figure 9. Idealized cross-section of an active hydrothermal system as might be produced by a magmatic intrusion associated with volcano formation on Mars. Ground ice above and near the intrusion would be melted. Heat transfer from the magma to the surrounding region would produce perturbations in the subsurface flow field resulting in a net flow of ground water toward the intrusion. Depending on the size of the intrusion, ground water underlying the permafrost within tens of kilometers or more would be mobilized. As the water nears the heat source, buoyancy forces (resulting from the decreased density of the warmer water) produce a net upward flow of water toward the surface. Water may reach the surface in the form of liquid or vapor or both. The time it takes for the water to re-enter the ground-water system to repeat the process depends on the surface and near-surface lithologic conditions. If the surface permeability is relatively low, water will flow on the surface and re-enter the subsurface at a greater distance from the source region than if the surface permeability were higher. The path which this hydrothermally derived water takes has important implications for fluvial valley formation. For example, on the left-hand flank, the surface is composed of permeable basalt so water quickly infiltrates and recharges a near-surface aquifer. The water again intersects the surface farther down the flank, forming a spring. At this site, sapping processes (i.e., erosion by ground-water outflow) may eventually form a valley. On the right-hand flank, the surface is mantled with ash. This ash, which is much less permeable than the underlying basalt, allows more water to flow at the surface. With continued surface flow, water will erace the surface and may eventually form a runoff valley. (Gulick and Baker 1989*b*; Gulick 1993).

thus providing an environment conducive to sapping, where these aquifers again intersected the surface via a seepage face along a cliff, fault, or lava flow front (Gulick and Baker 1989*b*). Preliminary numerical modeling results of Gulick et al. (1991) and Gulick (1992,1993) indicate that water discharges and volumes transported to the surface by magmatically induced hydrothermal systems are comparable to those required to form the Hawaiian fluvial valleys. Additional views on the role of hydrothermal activity in relation to igneous sill intrusions and impact cratering on Mars are provided by Wilhelms

and Baldwin (1989) and Brakenridge (1990), respectively.

Former hydrothermal systems associated with now extinct volcanic and impact-generated thermal anomalies on Mars would be important resource areas. Such regions could provide exploitable sources of hydrated salts, clays and other minerals, as well as hydrothermally altered minerals. Recent results from the ISM experiment of the Soviet Phobos mission (Bibring et al. 1989) tend to support this statement. They report that the slopes of the Tharsis volcanoes appear to be systematically enriched in hydrated minerals relative to the much drier surrounding plateau.

More speculative perhaps is the presence of extant hydrothermal systems. While there is no evidence of extant volcanism on Mars, in regions of most recent thermal activity, however, shallow zones (<1 km) of hydrothermal circulation might still exist in the subsurface. On Earth, hydrothermal systems can remain active well after the intrusion solidifies and volcanic activity ceases, although surface expression of such zones may have long since ceased. In fact, most terrestrial hydrothermal regions remain undetected unless there is physical or chemical evidence at the surface or they are discovered when digging shallow wells. However, if such regions still exist on Mars, then these areas could provide a valuable, accessible source of near-surface water and thermal energy.

There is some indication, although speculative in nature, that active hydrothermal systems might still exist on Mars. One such location is the Tharsis region. Mouginis-Mark (1990) identified small outflow channels at the base of the Olympus Mons escarpment, in northwestern Tharsis, and to the west of Ceraunius Fossae. These outflow channels are located on some of the youngest lava flow surfaces on Mars, indicating that substantial discharges of subsurface water continued into the relatively recent past. Because the Tharsis region is the youngest and by far the largest volcanic region on Mars, we propose that these discharges may reflect the last vestiges of the massive hydrothermal circulation systems which would have inevitably developed from the Tharsis thermal anomaly. Although there is no obvious surface manifestation at the present, hydrothermal water may still be circulating in the near surface environment in the Tharsis region.

Other potential locations for extant hydrothermal systems may be associated with the anomalous regions in the southern hemisphere of Mars. If the interpretation is correct that these regions indicate near-surface brines or ice melting, then the possibility that these reflect subsurface hydrothermal systems must be considered. Zent and Fanale (1986) addressed the persistence of ice/brine at the given locations in the present Martian environment. They conclude that if these deposits are not being replenished then the maximum lifetime of such deposits is on the order of 10^4 yr. However, if ice/brine is being replenished even at a very low rate, then such deposits can persist for 10^7 yr. If the deposits are being replenished, then we propose that the replenishment mechanism may be from subsurface hydrothermal circulation. The ice/brine deposits may be the surface manifestation of such systems.

VI. WATER-RESOURCE EXPLORATION

A strategy for finding useful water resources on Mars must allow for discovery. The relevant possibilities cannot all be conceived in advance of on-site exploration. For this reason, we believe that a practical approach of learning as one explores is called for. The following strategies apply to different potential resources.

A. Hydrated Minerals

Water may be present not only in the pore space of clays and other hydrated minerals; it also exists in the molecular structure of the minerals themselves. Hydrated clays can contain up to approximately 20% structurally bound water. This bound water could also be an important water source on Mars. Hydrated minerals are a predictable and detectable source of water and can be located anywhere on the planet. The principal drawback to the use of this water is the high energy cost associated with its extraction (Table I). Clays must be heated to temperatures of up to 500° C to release this bound water. However, the development of new techniques for extracting this water may make it a more feasible source to exploit. For example, the use of microwave energy, as recently proposed by Gwynne et al. (1991), converts more of the available energy into releasing the water. They point out that microwaves can be used without a pressurizing chamber, unlike conventional ovens which require a convection medium. Several regions of hydrated mineral deposits have already been identified in the Phobos mission data set of Mars. The spectral signature of hydrated minerals was detected on the slopes of the Tharsis volcanoes and regions within Valles Marineris (Bibring et al. 1989). Hydrated minerals should be especially concentrated in regions of former or extant hydrothermal systems. Likely geologic features such as volcanoes and impact craters, where vigorous hydrothermal activity probably occurred can be located using existing Viking images and the higher-resolution images of the upcoming Mars Observer mission. The actual presence of hydrated minerals at the surface can be detected spectroscopically from an orbiter (Bibring et al. 1989) or from Earth-based telescopes (see Singer et al. 1979 for summary).

Hydrated salts may contain over 50% more bound water and would be energetically more economical than phyllosilicates. Likely salts on the Martian surface include epsomite ($MgSO_4$-$7H_2O$), gypsum ($CaSO_4$-$2H_2O$), and hydrohalite ($NaCl$-$2H_2O$). These minerals would have formed in closed evaporative basins where ponding occurred. Likely locales include the floors of Argyre and Hellas, certain impact craters, and depressions within the northern plains and the southern highlands.

B. Hydrothermal Resources

Several methods used in the exploration of terrestrial hydrothermal systems might be useful in detecting extant hydrothermal systems on Mars. Gravity mapping may be helpful in detecting hydrothermal systems associated with

shallow intrusions (Rybach 1981). A negative anomaly, in some cases, may result from the steam fraction in high porosity reservoir rocks as well as to the lowered density caused by thermal expansion. However, positive gravity anomalies can also be associated with hydrothermal activity where several factors such as silicification by self-sealing may mask the negative anomalies.

Several others may be useful. Electrical resistivity mapping may aid in the location of hydrothermal regions (Nakumara and Sumi 1981) because of the inherent low salinity of volcanic rocks and the generally high salt content of hydrothermal water. However, the insertion of probes or shallow wells into the subsurface may be needed to utilize this technique. Remote sensing using imagery in the thermal infrared might be used to locate regions of anomalously high surface temperatures (Lumb 1981). Various geochemical methods might eventually be utilized to aid in identifying hydrothermally derived minerals. Near infrared spectral mapping of the surface, similar to that used in the Phobos mission (Bibring et al. 1989), could be helpful in locating extant or extinct hydrothermal deposits or hydrated minerals. Earth-based or spacecraft radar investigations might be useful in detecting possible outgassing regions of near-surface water or brines.

C. Ground Ice

Various geological, hydrological, and geophysical techniques are used in locating ground water on Earth. Some of these techniques would be useful in locating present ground-ice reserves on Mars. Hydrogeological techniques utilize aerial photographs, regional geologic maps and ground reconnaissance. Such methods are usually the first approach used in ground-water exploration on Earth and a similar approach might be used in ground-ice exploration on Mars. Geologic and geomorphic mapping is perhaps the most useful. In constructing maps, the permeability, porosity and lateral extensiveness of the geologic units should be noted as well as locating natural barriers (e.g., faults) to pre-existing ground-water flow. Permeable units (e.g., uncemented sands, permeable basalts) surrounded by deposits of low permeability (e.g., clays) may have trapped ground-water flow. Such areas are possible sites of present-day ground-ice reserves.

Initial small piloted missions to Mars probably would not depend on successful exploitation of Martian resources. However, the success of a continued buildup of a human presence on Mars will require local production of large quantities of fuel, water, and oxygen. Confirmation and mapping of exposed ice and hydrated minerals on the Martian surface can be accomplished by orbital visible and infrared spectroscopy, and orbital radar mapping can contribute to these studies and may detect subsurface brines. However, confirmation of concealed hydrous deposits will require surface studies by automated impact penetrators, rovers or soft landers equipped with drills capable of reaching several tens of meters or more, and geophysical methods. These same exploration tools could establish the mechanical properties of the overburden and hydrous deposits, and thereby impose appropriate engineering

constraints on the design of mining hardware and techniques.

The use of automated impact penetrators would be an economical way of collecting information about the nature and chemical composition of subsurface materials. Penetrators capable of reaching depths of 5 to 10 m would extend below the desiccated zone at high latitudes. A targeting precision ranging from 50 to 300 km would enable the largest Martian landforms such as shield volcanoes and polar layered deposits to be targeted, but would be insufficient to target most water-related landforms. Individual lobate debris aprons, a comparatively large class of suspected hydrous landforms, range from 5 to 50 km across; individual esker-like ridges and valleys are characteristically 1 to 2 km across, and any water-bearing valley floor sedimentary deposits would be even narrower. A penetrator targeting capability with km-scale precision would certainly require approach and terminal guidance control, and perhaps could be accomplished with a descent imaging device and simple terrain matching software onboard the penetrator probe. Descent imagery would also be useful for a geologic characterization of the immediate impact site.

VII. FUTURE MARS SCIENCE MISSIONS

Several upcoming Mars science missions will yield a wealth of new information regarding the nature, location, and fate of water on Mars. The next orbiter, Mars Observer, is scheduled to be launched in September, 1992 and will arrive at the red planet eleven months later. The Mars Observer will carry several scientific instruments which will yield direct information about water on Mars. The Gamma-Ray Spectrometer (GRS) will be used to map the elemental composition of the Martian surface with a spatial resolution up to a few hundred kilometers (Solar System Exploration Division 1991*a*). GRS will provide information about the present-day seasonal cycling and distribution of H_2O in the surface material. Such data will yield a better understanding of the present day distribution of water and whether areas which the geomorphic evidence indicates should still contain water have retained these volatile-rich reservoirs. One of the objectives of using the Mars Observer Thermal Emission Spectrometer (TES) is to determine and map the composition of the Martian surface materials (Solar System Exploration Division 1991*a*). TES will be able to distinguish between surfaces which are composed of rocks, unconsolidated material, or ice. The Mars Observer Camera's (MOC) narrow-angle fixed assembly has a selectable resolution from approximately 1.5 m per pixel to 11 m per pixel. Some images from MOC will yield a resolution which is more than 5 times better than the Viking Orbiter's highest resolution images (approximately 1.5 m per pixel in selected areas). Although MOC has no independent direct targeting capability, inevitably some landforms or regions suspected of containing significant quantities of ice will be sampled.

Mars-94/96 are the upcoming Mars missions of the former Soviet Union. In addition to rovers and penetrators, these missions will employ the use of

balloons. Experimental packages attached to the balloons will allow surface sampling of geochemical information while the balloon is on the ground at night. During the day, the camera will image these regions and atmospheric data will be gathered as the balloon floats above the surface. This mission will enable somewhat random areas of the Martian surface to be analyzed in detail. Mars Observer is scheduled to relay data from Mars-94 back to Earth.

Mars Environmental Survey mission (MESUR) is the next U. S. Mars surface mission scheduled to be launched at around the turn of the century. Four probes would be launched in the year 2001, while successive launches in 2003 and 2005 would bring the total number of Mars probes launched to 16 (Solar System Exploration Division 1991*b*). Each entry probe/lander would carry a descent and surface imager as well as various atmospheric, geochemical, and geophysical instruments. The addition of seismometers to these probes will enable detection of "marsquakes," which will indicate regions of geologic activity on the planet. The seismology would help locate any currently active volcanic centers, if they exist, and possibly any associated extant hydrothermal systems.

The success of future long-term human exploration or habitation on Mars will undoubtedly depend on the successful location and utilization of indigenous sources of water. Only with further geological, geophysical, and geochemical exploration of the red planet will the most accessible and economical of these reserves become apparent.

Acknowledgments. Our Mars studies have been supported in part by the NASA Planetary Geology and Geophysics Program. We thank B. Jakosky and B. Luchitta for helpful reviews. We also thank R. Strom for stimulating discussions relating to many of the ideas presented in this chapter.

REFERENCES

Baker, V. R. 1982. *The Channels of Mars* (Austin: Univ. of Texas Press).

Baker, V. R., Strom, R. G., Gulick, V. C., Kargel, J. S., Komatsu, G., and Kale, V. S. 1991. Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature* 352:589–594.

Barlow, N. G. 1988. Crater size-frequency distributions and a revised Martian relative chronology. *Icarus* 75:285–305.

Barlow, N. G., and Bradley, T. L. 1990. Martian impact craters correlations of ejecta and interior morphologies with diameter, latitude, and terrain. *Icarus* 87:156–179.

Bibring, J. P., Combes, M., Langevin, Y., Soufflot, A., Cara, C., Drossart, P., Encrenaz, Th., Erard, S., Forni, O., Gondet, B., Ksanfomality, L., Lellouch, E., Masson, Ph., Moroz, V., Rocard, F., Rosenqvist, J., and Sotin, C. 1989. Results from the ISM experiment. *Nature* 341:591–593.

Brakenridge, G. R. 1990. The origin of fluvial valleys and early geologic history, Aeolis Quadrangle, Mars. *J. Geophys. Res.* 95:17289–17308.

Brakenridge, G. R., Newsom, H. E., and Baker, V. R. 1985. Ancient hot springs on Mars origins and paleoclimatical significance of small Martian valleys. *Geology* 13:895–892.

Brass, G. W. 1980. Stability of brines on Mars. *Icarus* 42:20–28.

Bridges, N. T., and Barlow, N. G. 1989. Variation of Martian rampart crater ejecta lobateness in comparison to latitude, longitude, terrain, and crater diameter. *Lunar Planet. Sci.* XX:105–106 (abstract).

Carr, M. H. 1983. Stability of streams and lakes on Mars. *Icarus* 56:476–495.

Carr, M. H. 1987. Water on Mars. *Nature* 326:30–35.

Carr, M. H., and Schaber, G. G. 1977. Martian permafrost features. *J. Geophys. Res.* 82:4039–4055.

Carr, M. H., Crumpler, L. S., Cutts, J. A., Greeley, R., Guest, J. E., and Masursky, H. 1977. Martian impact craters and emplacement of ejecta by surface flow. *J. Geophys. Res.* 82:4055–4065.

Carr, M. H., Baum, W. A., Blasius, K. R., Briggs, G. A., Cutts, J. A., Duxbury, T. C., Greeley, R., Guest, J., Masursky, H., Smith, B. A., Soderblom, L. A., Veverka, J., and Wellman, J. B. 1980. In *Viking Orbiter Views of Mars*, ed. R. Spitzer, NASA SP-441.

Clark, R. N. 1978. Mars: Water-ice features in near infrared spectra of small areas. *Proc. Amer. Astron. Soc.* 10:567.

Clayton, L. 1964. Karst topography on stagnant glaciers. *J. Glaciology* 5:107–112.

Conant, J. B. 1951. *Science and Common Sense* (New Haven, Conn.: Yale Univ. Press).

Corte, A. E. 1987. Rock glacier taxonomy. In *Rock Glaciers*, eds. J. R. Giardino, J. F. Schroeder, Jr., and J. D. Vitek (Boston: Allen and Unwin), pp. 27–40.

Costard, F. M. 1989. The spatial distribution of volatiles in the Martian hydrolithosphere. *Earth, Moon, and Planets* 45:265–295.

Downs, G. S., Reichley, P. E., and Green, R. R. 1975. Radar measurements of Martian topography and surface properties the 1971 and 1973 oppositions. *Icarus* 26:273–312.

Driscoll, F. G., Jr. 1980. Wastage of the Klutlan ice-cored moraines, Yukon Territory, Canada. *Quaternary Res.* 14:31–49.

Elder, J. 1981. *Geothermal Systems* (New York: Academic Press).

Ellis, J. M., and Calkin, P. E. 1979. Nature and distribution of glaciers, neoglacial moraines, and rock glaciers, east-central Brooks Range, Alaska. *Arctic and Alpine Res.* 11:403–420.

Ernst, W. G. 1976. *Petrologic Phase Equilibria* (San Francisco: W. H. Freeman).

Fanale, F. P., Salvail, J. R., Zent, A. P., and Postawko, S. E. 1986. Global distribution and migration of subsurface ice on Mars. *Icarus* 67:1–18.

Farmer, C. B., and Doms, P. E. 1979. Global seasonal variation of water vapor on Mars and the implications for permafrost. *J. Geophys. Res.* 84:2881–2888.

Faul, H. 1973. The Cliff of Nix Olympica. Unpublished manuscript.

Freeze, R. A., and Cherry, J. A. 1979. *Groundwater* (Englewood Cliffs, N. J.: Prentice-Hall).

Fyfe, W. S., Turner, R. J., and Veerhoogen, J. 1958. *Metamorphic Reactions and Metamorphic Facies.* Geological Soc. of America Memoir 73 (Baltimore: Waverly Press).

Gardner, J. S., and Bajewsky, I. 1987. Hilda rock glacier stream discharge and sediment load characteristics, Sunwapta Pass area, Canadian Rocky Mountains. In *Rock Glaciers*, eds. J. R. Giardino, J. F. Schroeder, Jr., and J. D. Vitek (Boston: Allen and Unwin), pp. 161–174.

Grout, F. F., Sharp, R. P., and Schwartz, G. M. 1959. Glacial geology. In *The Geology of Cook County Minnesota*, Univ. of Minnesota Geological Survey Bull.

39 (Minneapolis: Univ. of Minnesota Press), pp. 59–74.
Gulick, V. C. 1992. Magmatic intrusions and hydrothermal sytems on Mars. In *Proc. of Martian Surface and Atmosphere Through Time (MSATT) Workshop*, eds. R. M. Haberle and B. M. Jakosky, LPI Tech. Rept. 92-02, pp. 50–51 (abstract).
Gulick, V. C., and Baker, V. R. 1989*a*. Fluvial valleys and martian paleoclimates. *Nature* 341:514–516.
Gulick, V. C., and Baker, V. R. 1989*b*. The role of hydrothermal circulation in the formation of fluvial valleys on Mars. *Lunar Planet. Sci.* XX:369–370 (abstract).
Gulick, V. C., and Baker, V. R. 1990*a*. Origin and evolution of valleys on martian volcanoes. *J. Geophys. Res.* 95:14325–14344.
Gulick, V. C., and Baker, V. R. 1990*b*. Valley development on Mars: A global perspective. *Lunar Planet. Sci.* XXI:443–444 (abstract).
Gulick, V. C., Marley, M. S., and Baker, V. R. 1991. Numerical modeling of hydrothermal systems on martian volcanoes. *Lunar Planet. Sci.* XXII:509–510 (abstract).
Gwynne, O., Cozzatti, J. P., Zent, A., and McKay, C. P. 1991. Getting water from water of hydration. Resources of Near-Earth Space: Proc. Second Annual Symp. UA/NASA SERC, Jan. 7–10, Tucson, Ariz., Abstract book, p. 30.
Hassinger, J. M., and Mayewski, P. A. 1983. Morphology and dynamics of the rock glaciers in southern Victoria Land, Antarctica. *Arctic and Alpine Res.* 15:351–368.
Howard, A. D. 1981. Etched plains and braided ridges of the south polar region of Mars: Features produced by basal melting of ground ice? In *Reports of Planetary Geology Program—1981*, NASA TM-84211, pp. 286–288.
Huguenin, R. L., and Clifford, S. M. 1980. Additional remote sensing evidence for oasis on Mars. In *Reports of Planetary Geology Program—1979–1980*, NASA TM-81776, pp. 153–155.
Huguenin, R. L., Clifford, S. M., Sullivan, C. A., and Miller, K. J. 1979. Remote sensing evidence for oases on Mars. In *Reports of Planetary Geology Program—1978–1979*, NASA TM-80339, pp. 208–214.
Jakosky, B. M. 1985. The seasonal cycle of water on Mars. *Space Sci. Rev.* 41:131–200.
Jakosky, B. M., and Farmer, C. B. 1982. The seasonal and global behavior of water in the Mars atmosphere complete global results of the Viking atmospheric water experiment. *J. Geophys. Res.* 89:2999–3019.
Johansen, L. A. 1979. The latitude dependence of Martian splosh cratering and its relationship to water. In *Reports of Planetary Geology Program—1978–1979*, NASA TM-80339, pp. 123–125.
Johnson, P. G., and Lacasse, D. 1988. Rock glaciers of the Dalton Range, Kluane Ranges, south-west Yukon Territory, Canada. *J. Glaciology* 34:327–332.
Kargel, J. S. 1986. Morphologic variations of Martian rampart crater ejecta and their dependencies and implications. *Lunar Planet. Sci.* XVII:410–411 (abstract).
Kargel, J. S. 1989. First and second-order equatorial symmetry of Martian rampart crater ejecta morphologies. Fourth Intl. Conf. on Mars, Jan. 10–13, Tucson, Ariz., Abstract book, p. 132.
Kargel, J. S., and Strom, R. G. 1990. Ancient glaciation on Mars. *Lunar Planet. Sci.* XXI:597–598 (abstract).
Kargel, J. S., and Strom, R. G. 1992. Ancient glaciation on Mars. *Geology* 20:3–7.
Kargel, J. S., Croft, S. K., Lunine, J. I., and Lewis, J. S. 1991. Rheological properties of ammonia-water liquids and crystal-liquid slurries planetological applications. *Icarus* 89:93–112.
Kaszycki, C. A. 1987. A model for glacial and proglacial sedimentation in the shield terrane of southern Ontario. *Canadian J. Earth Sci.* 24:2373–2391.

Kuzmin, R. O. 1980. Morphology of fresh Martian craters as an indicator of the depth of the upper boundary of the ice-bearing permafrost a photogeologic study. *Lunar Planet. Sci.* XI:585–586 (abstract).

Lucchitta, B. K. 1981, Mars and Earth comparisons of cold-climate features. *Icarus* 45:264–303.

Lucchitta, B. K. 1982. Ice sculpture in the Martian outflow channels. *J. Geophys. Res.* 84:9951–9973.

Lumb, J. T. 1981. Prospecting for geothermal resources. In *Geothermal Systems: Principles and Case Histories*, eds. L. Rybach and L. J. P. Muffler (New York: Wiley), pp. 77–103.

McCord, T. B., Clark, R., and Huguenin, R. L. 1978. Mars: Near-infrared spectral reflectance and compositional implications. *J. Geophys. Res.* 83:5433–5441.

McCord, T. B., Clark, R. N., and Singer, R. B. 1982. Mars: Near-infrared reflectance spectra of surface regions and compositional implications. *J. Geophys. Res.* 87:3021–3032,

Milton, D. J. 1974. Carbon dioxide hydrated and floods on Mars. *Science* 183:654–656.

Morris, S. E. 1987. Regional and topoclimatic implications of rock glacier stratigraphy. In *Rock Glaciers*, eds. J. R. Giardino, J. F. Schroeder, Jr., and J. D. Vitek (Boston: Allen and Unwin), pp. 107–126.

Mouginis-Mark, P. J. 1979. Martian fluidized crater morphology variations with crater size, latitude, altitude, and target material. *J. Geophys. Res.* 84:8011–8022.

Mouginis-Mark, P. J. 1987. Water or ice in the Martian regolith? Clues from rampart craters seen at very high resolution. *Icarus* 71:268–286.

Mouginis-Mark, P. J. 1990. Recent water release in the Tharsis region of Mars. *Icarus* 84:362–373.

Nakumara, H., and Sumi, K. 1981. Exploration and development in Takinoue, Japan. In *Geothermal Systems: Principles and Case Histories*, eds. L. Rybach and L. J. P. Muffler (New York: Wiley), pp. 247–272.

National Research Council. 1991. *Opportunities in the Hydrological Sciences* (Washington, D. C.: National Academy Press).

Norton, D. 1984. Theory of hydrothermal systems. *Ann. Rev. Earth and Planet. Sci.* 12:155–177.

Parker, T. J., Saunders, R. S., and Schneeberger, D. M. 1989. Transitional morphology in the west Deuteronilus Mensae region of Mars: Implications for modification of the lowland/upland boundary. *Icarus* 82:111–145.

Price, R. J. 1973. *Glacial and Fluvioglacial Landforms* (New York: Hafner).

Robinson, G. R., Jr., and Haas, J. L., Jr. 1983. Heat capacity, relative entropy, and calorimetric entropy of silicate minerals an empirical method of prediction. *Amer. Mineralogist* 68:641–553.

Roth, L. E., and Saunders, R. S. 1986. Mars seasonally variable radar reflectivity II. *Lunar Planet. Sci.* XVII:730–731 (abstract).

Roth, L. E., Saunders, R. S., and Schubert, G. 1985. Seasonally variable reflectivity. *Lunar Planet. Sci.* XVI:712–713 (abstract).

Rybach, L. 1981. Geothermal systems, conductive heat flow, geothermal anomalies. In *Geothermal Systems: Principles and Case Histories*, eds. L. Rybach and L. J. P. Juffler (New York: Wiley), pp. 3–31.

Schultz, P. H., and Gault, D. E. 1979. Atmospheric effects on martian ejecta emplacement. *J. Geophys. Res.* 84:7669–7687.

Schultz, P. H., and Gault, D. E. 1982. Impact ejecta dynamics in an atmosphere: Experimental results and extrapolations. In *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, eds. L. T. Silver and P. H. Schultz, Geological Soc. of America SP-90 (Boulder: Geological Soc. of America), pp.

153–174.
Schultz, P. H., and Gault, D. E. 1984. On the formation of contiguous ramparts around Martian impact craters. *Lunar Planet. Sci.* XV:732–733 (abstract).
Scott, D. H., and Chapman, M. G. 1991. Mars Elysium Basin: Geologic/volumetric analysis of a young lake and exobiologic implications. *Proc. Lunar Planet. Sci. Conf.* 21:669–678.
Singer, R. B., McCord, T. B., and Clark, R. N. 1979. Mars surface composition from reflectance spectroscopy: A summary. *J. Geophys. Res.* 84:8415–8426.
Sinton, W. M. 1967. On the composition of Martian surface material. *Icarus* 6:222–228.
Solar System Exploration Division. 1991*a*. *Mars Observer Participating Scientist Program*, NASA Research Announcement NRA-91-OSSA-8.
Solar System Exploration Division. 1991*b*. *Solar System Exploration Division (SSED) Strategic Plan in Volume II MFPE: Mission from Planet Earth*, pp. 18–24.
Sollid, J. L., and Reite, A. J. 1983. The last glaciation and deglaciation of central Norway. In *Glacial Deposits in North-West Europe*, ed. J. Ehlers (Rotterdam: A. A. Balkema), pp. 41–60.
Squyres, S. W. 1984. The history of water on Mars. *Ann. Rev. Earth Planet. Sci.* 12:83–106.
Squyres, S. W., and Carr, M. H. 1986. Geomorphic evidence for the distribution of ground ice on Mars. *Science* 231:249–252.
Strom, R. G., Croft, S. K., and Barlow, N. G. 1992. The martian impact cratering record. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 383–423.
Tanaka, K. L., and Chapman, M. G. 1992. Kasei Valles, Mars: Interpretation of canyon materials and flood sources. In *Proc. Lunar Planet. Sci. Conf.* 22:73–83.
Teller, J. T., and Clayton, L., eds. 1983. *Glacial Lake Agassiz*, Geological Assoc. of Canada Special Paper 26 (Toronto: Univ. of Toronto Press).
Thwaites, F. T. 1926. The origin and significance of pitted outwash. *J. Geology* 34:308–319.
Vitek, J. D., and Giardino, J. R. 1987. Rock glaciers: A review of the knowledge base. In *Rock Glaciers*, eds. J. R. Giardino, J. F. Schroeder, Jr., and J. D. Vitek (Boston: Allen and Unwin), pp. 1–26.
Von Englehardt, W., and Zimmermann, J. 1982. *Theory of Earth Science* (Cambridge: Cambridge Univ. Press).
Wade, F. A., and de Wys, J. N. 1968. Permafrost features on the Martian surface. *Icarus* 9:175–178.
Wallace, D., and Sagan, C. 1979. Evaporation of ice in planetary atmospheres ice-covered rivers on Mars. *Icarus* 39:385–400.
Wilhelms, D. E., and Baldwin, R. J. 1989. The role of igneous sills in shaping the Martian uplands. *Proc. Lunar Planet. Sci. Conf.* 19:355–356.
Zent, A. P., and Fanale, F. P. 1986. Possible Mars brines: Equilibrium and kinetic considerations. *Proc. Lunar Planet. Sci. Conf.*, 16, *J. Geophys. Res. Suppl.* 91:D439–D445.
Zent, A. P., Fanale, F. P., Salvail, J. R., and Postawko, S. E. 1986. Distribution and state of H_2O in the high-latitude subsurface of Mars. *Icarus* 67:19–36.
Zent, A. P., Fanale, F. P., and Roth, L. E. 1990. Possible Martian brines: Radar observations and models. *J. Geophys. Res.* 95:14531–14542.
Zisk, S. H., and Mouginis-Mark, P. J. 1980. Anomalous region on Mars implications for near-surface liquid water. *Nature* 288:735–738.
Zisk, S. H., and Mouginis-Mark, P. J. 1982. Alternate models for the Solis Lacus radar anomaly on Mars. In *Papers Presented to the Third International Colloquium on Mars* (Houston: Lunar and Planetary Inst.), pp. 294–296.

A CHEMICAL APPROACH TO CARBON DIOXIDE UTILIZATION ON MARS

ALOYSIUS F. HEPP
NASA Lewis Research Center

GEOFFREY A. LANDIS
Sverdrup Technology, Inc.

and

CLIFFORD P. KUBIAK
Purdue University

This chapter examines several novel proposals for CO_2 reduction by chemical, photochemical, and photoelectrochemical means. Photolytic reduction of CO_2 to CO and O_2 can occur under mild conditions using a series of recently synthesized trinuclear nickel catalysts. We examine potential uses of CO, a by-product of CO_2 reduction, and carbon, as reducing agents in metal oxide processing to produce structural or power materials; CO_2 produced in these reactions can then be recycled to generate O_2 and CO. Reduction of CO_2 with hydrogen produces methane through the Sabatier process. Further partial oxidation of methane can be carried out to produce acetylene, through the Sachsse process.The impact of such propellants on future missions to Mars is assessed.

I. INTRODUCTION

A. A Chemical Approach to In-Situ Resource Utilization

This chapter examines several novel proposals for CO_2 reduction through chemical, photochemical and photoelectrochemical means. The organization is as follows: the introduction discusses various technologies for carbon dioxide reduction. The discussion of CO_2 chemistry is followed by examples of proposed light-assisted devices. We then give examples of CO use for metal oxide reduction and as a raw material to produce other propellants. An alternative example of a chemical reduction sequence of CO_2 to produce acetylene (C_2H_2) is discussed along with a cursory mission analysis. During this discussion, we include other options or examples of earlier work whenever possible to provide as broad a background as possible to a worker new to the field.

One goal of our effort is to design photochemical and photoelectrochemical systems for the extraterrestrial manufacturing of products required to sustain a human presence on Mars from atmospheric carbon dioxide. The approach involves coupled catalytic cycles to simultaneously produce oxygen and carbon monoxide. Efficient metal catalysts and photocatalysts are to be used for both half-reactions: (i) the reduction of CO_2 to CO; and (ii) the evolution of O_2 . Oxygen can be used for life support or as an oxidizer in a rocket engine. Another goal of our work is to emphasize the numerous potential uses for CO. Carbon monoxide can be used as a rocket fuel, for the production of other propellants and numerous petroleum products, in CO fuel cells, or for the winning of elemental metals and semimetals from oxides produced on Mars (SiO_2 , Al_2O_3, Fe_2O_3) for structural and power materials. Finally, we discuss an alternative production sequence of an unsaturated hydrocarbon to point out the utility of a hydrogen-poor propellant as an option to consider for *in-situ* propellant production.

The goal of our chemical approach as illustrated throughout this chapter, is to point out the vast amount of chemistry and photochemistry that is known for CO_2 and CO with metals and metal compounds. Our hope is to reinvigorate the debate over the technologies to be used to exploit *in-situ* resources. With an enhanced arsenal of technologies available to combat the challenges of distance, time and hostile environment, the dream of human exploration of the solar system may become a reality.

B. Availability and Exploitation of Carbon Dioxide on Mars

1. Availability of Carbon Dioxide on Mars. Utilization of resources available *in-situ* is a critical enabling technology for a permanent human presence in space. A permanent presence on Mars (Kaplan 1988), for example, requires a tremendous infrastructure to sustain life under hostile conditions [low oxygen partial pressure (Meyer and McKay 1989), ultraviolet radiation (Oro and Holzer 1979), low temperatures (Jakosky and Haberle 1990), etc]. Consequently, there have been numerous studies on the exploitation of the most accessible of Martian resources, atmospheric carbon dioxide, for propulsion (see, e.g., Ash et al. 1978; Ash et al. 1989; French 1989; Frisbee 1987; Galecki 1988; Johnson and Leonard 1989; LeCompte and Stets 1991; Ramohalli and Sridhar 1991; Stancati et al. 1991).

Atmospheric CO_2 is abundant [the atmosphere of Mars consists of 95% carbon dioxide (Meyer and McKay 1989; see Table I)]. It is available at all points on the surface as far as we know—requires no precursor mission to verify. It is chemically simple requiring no precursor missions to verify composition or properties; and can be obtained by simple compression, with no requirements of mining or beneficiation equipment operation.

2. Technologies for Carbon Dioxide Reduction. The splitting of CO_2 to

CO and O_2 is thermally demanding:

$$2CO_2 \longrightarrow 2CO + O_2;\ \Delta H^\circ = 566\ kJ\,mol^{-1},\quad \Delta G^\circ = 514\ kJ\,mol^{-1} \tag{1}$$

and can be effected at temperatures in excess of 1000 K over zirconia (Richter 1981) at low efficiencies. At elevated temperatures (500–1200°C), many transition metals can deoxygenate CO_2 (Sneeden 1982*b*):

$$Fe + CO_2 \longrightarrow FeO + CO. \tag{2}$$

However, this reaction is complicated by the formation of higher oxides (and carbonates) by further reactions (Sneeden 1982*b*):

$$3FeO + CO_2 \longrightarrow Fe_3O_4 + CO. \tag{3}$$

TABLE I

Composition of Martian Lower Atmosphere[a]

Gas	Volume	Fraction
Carbon Dioxide	(CO_2)	95.32
Nitrogen	(N_2)	2.7
Argon	(Ar)	1.6
Oxygen	(O_2)	0.13
Carbon Monoxide	(CO)	0.07
Water Vapor	(H_2O)	0.03

[a] Source: (Owen et al. 1977).

Carbon dioxide can be heated in the presence of carbon itself to form CO at temperatures of 800 to 1000°C over Pt or Fe (Sneeden 1982*b*):

$$Coal + CO_2 \longrightarrow 2CO. \tag{4}$$

On the appropriate metal or oxide catalyst at elevated temperatures, a reverse water gas shift reaction followed by a separate water electrolysis step can produce CO and O_2 from CO_2 (Sneeden 1982*b*):

$$CO + H_2O \leftrightarrow CO_2 + H_2 \tag{5}$$

$$H_2O \longrightarrow H_2 + 1/2O_2. \tag{6}$$

Other options include the Sabatier process to form methane (see discussion below under propellant production on Mars); electrolysis in aqueous or nonaqueous media (reaction [7] and [8], respectively) (Cotton and Wilkinson 1988) and photosynthesis (reaction [9]):

$$CO_2 + 2e^- + 2H^+ \longrightarrow CO + H_2O \tag{7}$$

$$2CO_2 + 2e^- \longrightarrow C_2O_4^{2-} \tag{8}$$

$$6CO_2 + 6H_2O \longrightarrow C_6H_{12}O_6 + 6O_2. \tag{9}$$

There are numerous reviews that discuss the chemistry and technology of reducing CO_2 (see, e.g., Palmer and van Eldik 1983; Sneeden 1982*b*); reactions (1) through (10) are examples of standard thermal CO_2 reduction chemistry. After a brief discussion of Sabatier technology (reaction [10]) below, we focus on recently developed photolytic CO_2 reduction chemistry, the focus of this approach is low-mass, mechanically simple devices for reducing CO_2 .

C. Propellant Production on Mars

1. General Considerations and Background. As outlined by numerous research groups (see, e.g., Ash et al. 1978,1989; Baker and Zubrin 1990; French 1989; Frisbee 1987; Galecki 1988; Johnson and Leonard 1989; LeCompte and Stets 1991; Landis and Linne 1991; Ramohalli and Sridhar 1991; Zubrin et al. 1991), tremendous advantages in chemical propulsion missions can be achieved by using Martian resources for propulsion.

Major issues for propellant production systems include specific impulse (I_{sp}), safety, storage issues, capital requirements, manufacturability, reliability, and import mass leverage (which, depends on local water availability for practical extraction). We will limit our discussion to an examination of I_{sp}, reliability, and mass leverage issues. We have addressed storage (Hepp et al. 1991*a*) and performance issues (Landis and Linne 1991) in greater depth elsewhere; other issues related to propellant production have been addressed in detail by others (see, e.g., Ash et al. 1978,1989; Baker and Zubrin 1990; French 1989; Frisbee 1987; LeCompte and Stets 1991; Ramohalli and Sridhar 1991; Zubrin et al. 1991).

A review of the atmosphere of Mars (Table I) provides many possible options for fuels and oxidizers (Clark 1989; LeCompte and Stets 1991). The fuel options include C-free fuels (H_2, SiH_4, N_2H_4, etc.), H-free fuels (CO, $(CN)_2$) etc., or C,H-containing fuels (CH_4 , CH_3OH, C_2H_2 , C_6H_6, etc.); the oxidizer options include materials other than O_2 (H_2O_2 , NO_x, etc.). We will focus on C-containing fuels oxidized by oxygen. For example, Mars-derived carbon monoxide can be used directly as a fuel, at a specific impulse of $\sim$300 s. For higher specific impulses it is necessary to synthesize hydrocarbon fuels, as discussed in several following sections.

The reduction of CO_2 to hydrocarbons can be accomplished with hydrogen. This results in producing a large mass of rocket fuel on Mars from a small amount of hydrogen brought from Earth. For ΔV return missions such as manned Mars missions, most of the initial mass required to be placed in low Earth orbit is rocket fuel. Of this, a large fraction comprises the fuel for the return trip and the fuel in low Earth orbit required to boost the return fuel to Mars. A mission where the return propellant need not be shipped to Mars, would greatly reduce the required mission mass. Thus, production of rocket

propellant from available resources is an extremely high-leverage approach to reducing mission mass. An excellent example of Mars-derived resource utilization is the use of the Sabatier process on Mars proposed by Zubrin and Baker (Baker and Zubrin 1990; Zubrin et al. 1991).

2. *The Zubrin/Baker Option for Methane Production.* Synthesis of methane/oxygen fuels from indigenous Martian materials has been discussed by several other workers (Ash et al. 1978,1989; French 1989). A Mars mission involving processing of methane fuel is detailed by Zubrin and Baker (Baker and Zubrin 1990; Zubrin et al. 1991). Their proposal was for an unmanned preliminary mission to bring to Mars: (1) the return spacecraft; (2) a quantity of liquid hydrogen; and (3) an atmospheric processing module, followed two years later by a manned mission. The processing module processes the hydrogen along with atmospheric carbon dioxide into methane and oxygen by the reaction:

$$4H_2 + CO_2 \longrightarrow 2H_2O + CH_4. \tag{10}$$

This is a highly exothermic reaction, the rate of which will be limited primarily by the ability to remove the heat produced. The evolved water is recycled to hydrogen and oxygen by electrolysis. Additional oxygen is produced by reduction of carbon dioxide as discussed in the previous section (see also Ash et al. 1989; Baker and Zubrin 1990; Zubrin et al. 1991; Clapp and Scardera 1989).

The purpose of this chemistry is to produce a large amount of return fuel from a small amount of hydrogen. The required hydrogen is about 5% of the mass of the fuel produced. Another advantage is the ease of storage. Hydrogen brought from Earth is converted into methane and water by reaction (10) within the two days of arrival on Mars, thus eliminating the significant difficulties of long-term cryogenic storage of hydrogen. The Baker/Zubrin proposal envisioned completion of propellant manufacture, resulting in a fully fueled return vehicle on Mars before the manned crew is launched, and contained several safeguards to ensure that the manned crew would reach the fueled return vehicle despite any credible worst-case.

II. PHOTOLYTIC CARBON DIOXIDE REDUCTION

A. Background

Many studies focus on obtaining oxygen from CO_2 and various uses for oxygen including life support and propulsion; discussion of carbon monoxide, the co-product from CO_2 reduction revolves around its use as a fuel, being oxidized back to CO_2 (see, e.g., Galecki 1988; Frisbee 1987; Linne et al. 1990). These studies often focus on the initial placement of systems using Earth-derived technologies with large infrastructures to handle power and reliability requirements (Ash et al. 1989; Frisbee 1987). It may prove essential in the initial phase of the exploration and colonization of Mars to employ novel, lightweight photochemically based systems to produce oxygen and

reduced carbon species. These mechanically simple systems could serve as low-capacity carbon dioxide separation and conversion devices. It is expected that even after they are supplanted by more sophisticated systems (Ash et al. 1989; Frisbee 1987) in the future, photochemically based systems will be useful backups in case of a malfunction of the "second generation" systems.

The solar power density on the surface of Mars is ~250 W m^{-2}, approximately one-fourth that on Earth (Appelbaum and Flood 1990), and more than adequate to power the direct photochemical or photoelectrochemical reduction of CO_2 . We propose the development of energy self-sufficient photochemical systems for the production of oxygen and carbon monoxide from carbon dioxide on the surface of Mars.

The splitting of CO_2 to CO and O_2 represents the reverse reaction of a low-temperature CO/O_2 fuel cell (Wu and Kubiak 1983):

$$CO + 1/2O_2 \longrightarrow CO_2; \Delta E = 1.33V. \tag{11}$$

Our approach is to drive the endoenergenic reverse reaction by photochemical and photoelectrochemical electron/hole separation.

The reduction of CO_2 to CO using metal complexes has been studied extensively (Darensbourg and Kudaroski 1983; Kubiak and Ratliff 1991). The reduction of CO_2 can be accomplished by thermal reduction of CO_2 and O-atom transfer to another substrate (Wu et al. 1987; Bryan et al. 1987):

$$^*CO_2 + Sub \longrightarrow Sub = O + ^*CO \tag{12}$$

multiple bond metathesis with another unsaturated species (DeLaet et al. 1987):

$$^*CO_2 + C \equiv E \longrightarrow O = ^*C = E + CO \tag{13}$$

photochemical activation (DeLaet et al. 1988); or electrocatalysis with nickel clusters (Ratliff et al. 1992) as shown in Fig. 1.

Of several purely photochemical schemes for CO_2 reduction to CO examined to date, the reduction by nickel cluster electrocatalysts (Ratliff and Kubiak 1991) holds great promise for a photochemical device. The proven ability of nickel cluster compounds to electrocatalyze the reduction of CO_2 to CO very near the expected thermodynamic potential will be coupled with photocatalytic and photoelectrochemical systems for oxygen evolution to accomplish the overall splitting of CO_2 to CO and O_2 .

B. Photochemical Carbon Dioxide Reduction

A proposed $Pt(Ni_3)_2$ molecular assembly for the photochemical splitting of CO_2 to CO and O_2 is presented in Fig. 2. In the initial step, photoexcitation of the $Pt(OH)_2$ chromophore induces the elimination of oxygen and reduction of Pt(II) to Pt(0). The Pt(0) center is capable of electron transfer to the pendant Ni_3^+ clusters. Each reduced Ni_3 cluster thus formed is known to effect the reduction of CO_2 to CO. After discharge of both the O_2 and CO, the

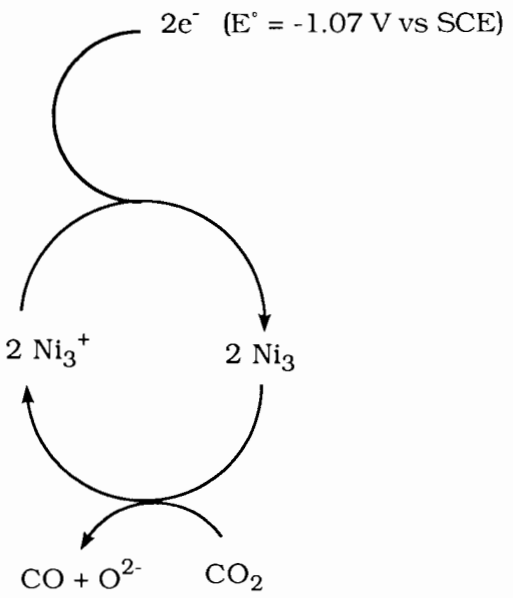


Figure 1. Electrocatalytic splitting of carbon dioxide using nickel cluster catalyst. See text for discussion.

hydroxylated catalyst can be restored by combination of OH^- with the Pt(II) center. We note the system is catalytic in H_2O. No exogenous H_2O or aqueous solvent is required; only one mole of H_2O per mole of $Pt(Ni_3)_2$ catalyst is required. Essentially "dry" devices for operation in a Martian environment can be constructed by immobilization of the $Pt(Ni_3)_2(H_2O)$ in an ionic resin or polymer such as Nafion.

A necessary component of the photochemical system is the splitting of water. The separation of water into H_2 and O_2 is a problem of both tremendous fundamental interest and significant technical difficulty. Aqueous solutions of ruthenium bipyridine oxo complexes have been found by Meyer et al. to catalyze the splitting of water to oxygen (Gilbert et al. 1985). The requirement of an aqueous solvent for the ruthenium catalysts, however, renders them awkward for use on Mars ($T \approx -50°C$). Platinum electrodes exhibit the lowest known overpotentials for oxygen evolution and are generally regarded to be the oxygen electrode materials of choice. Recent molecular photochemical studies on platinum hydroxide and related alkoxide complexes suggest that the photochemical elimination of O_2 can be achieved via intermediary peroxides (Ni and Kubiak 1992). These photochemical reductive eliminations result in the net two-electron reduction of platinum(II) to platinum(0). We propose to couple these photoredox systems and others like them with the known electrocatalytic reduction of CO_2 to CO by nickel clusters (Ratliff et al. 1992).

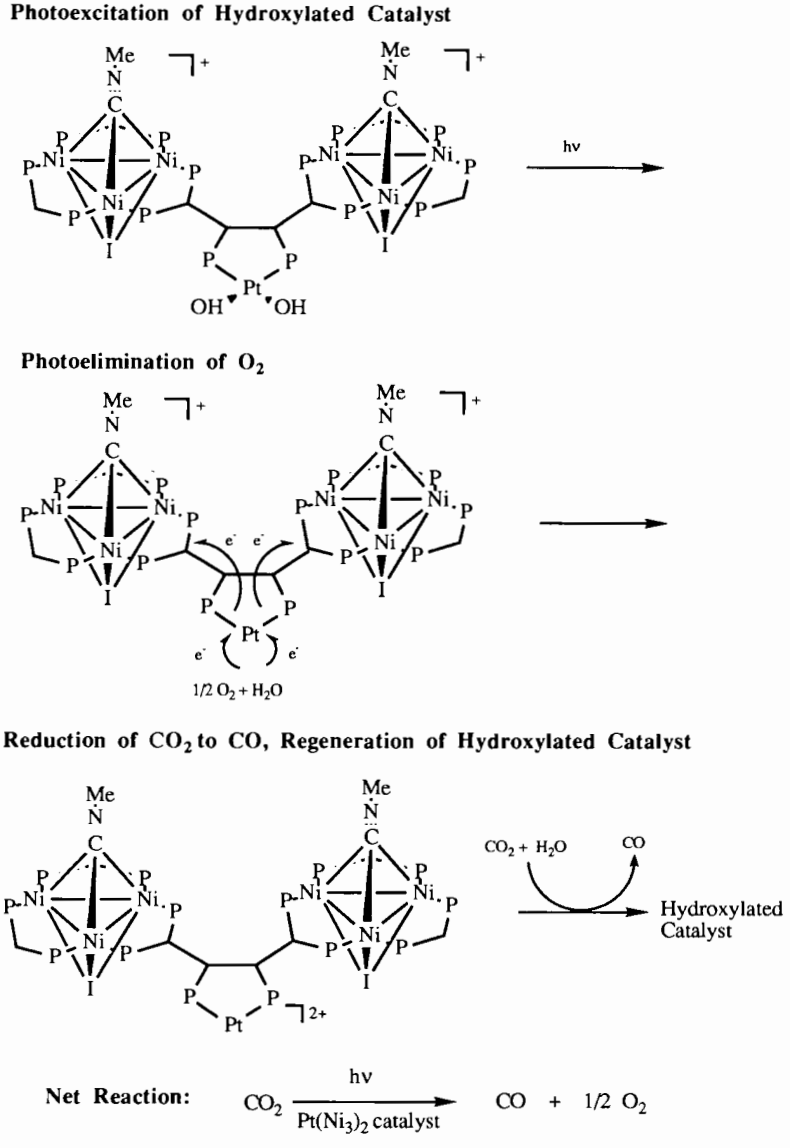


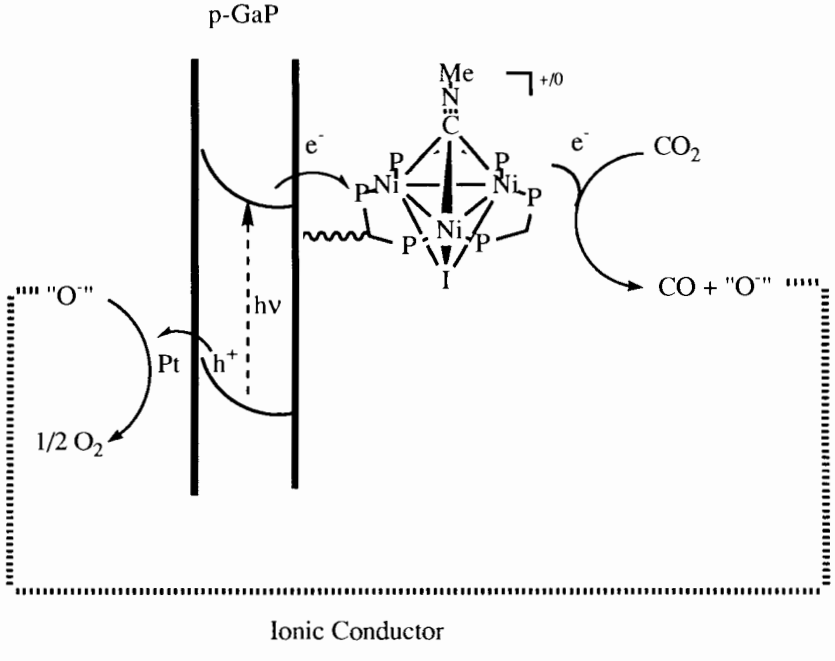
Figure 2. A photochemical system for splitting carbon dioxide. Note that the reaction is also catalytic in water. The reaction produces as much water as is consumed. Added or exogenous water is not required.

The potentials for platinum(0) reoxidation to platinum(II) match well with the potentials for the electrocatalytic reduction of CO_2 to CO by nickel clusters (Ratliff et al. 1992).

C. Photoelectrochemical Carbon Dioxide Reduction

Another approach is to prepare solid-state semiconductor photoelectrochem-

ical devices for the splitting of CO_2 to CO and O_2 . In particular, p-GaP ($E_g = 2.26$ eV) has nearly ideal band placement for the photocathodic reduction of Ni_3 cluster electrocatalysts for CO production (Ratliff et al. 1992) coupled with anodic oxygen evolution. Irradiation with light above the bandgap of p-GaP should result in the photoreduction of Ni_3 cluster electrocatalysts sites on the surface. These electrocatalytic sites are known to reduce CO_2 to CO. The photogenerated holes (h^+) can migrate to the platinized dark side of the device for the reduction of "O^-" ions (which may take the form of O^{2-}, OH^-, HCO^{2-}, or CO_3^{2-} depending on the ionic conductor employed and conditions). A proposed p-GaP device for the splitting of CO_2 is presented in Fig. 3.



Net Reaction: $CO_2 \xrightarrow[\text{p-GaP/Ni}_3\text{/Pt}]{h\nu} CO + 1/2\ O_2$

Figure 3. A semiconductor photoelectrochemical system for splitting carbon dioxide. See text for discussion regarding energy requirement of absorbed photon.

Functioning of such a direct photoelectrochemical reaction on the surface of Mars requires the presence of ultraviolet photons of energy greater than 2.26 eV. Although the solar spectrum has not been measured at the surface of Mars,

the Martian atmosphere, with no ozone layer, apparently allows penetration of solar ultraviolet to the surface. The integrated space (AM0) solar spectrum, i.e., total number of photons with energy greater than the photon energy listed, is known for the distance of Earth from the Sun (Thekaekara et al. 1969). On Mars, the portion of the solar spectrum with $h\nu > 2.26$ eV represents an electrolysis current equivalent to roughly 3 mA per cm^2 of solar-exposed surface.

Other issues surrounding the use of photolytic reactors for CO_2 reduction include the relative merits of bringing photons into large-area hermetic photoreactors vs use of vacuum photovoltaic converters with power conveyed to reactors of more optimized dimensions; and the projected effects of photon to electrical, or chemical, efficiencies; and diurnal thermal cycles on operations. While these are important operational issues, they are outside of the scope of this chapter, the reader is directed to other sources for relevant discussion (Appelbaum and Flood 1990; Landis and Appelbaum 1990; Landis and Appelbaum 1991).

III. CARBON MONOXIDE UTILIZATION ON MARS

A. Chemistry of Carbon Monoxide

Carbon monoxide is a tasteless, odorless, and toxic substance with a melting point of −205°C (68 K) and a boiling point of −190°C (83 K) (Cotton and Wilkinson 1988); it is a minor component of the Martian atmosphere (see Table I).

As discussed above, carbon monoxide as a resource on Mars, derived from CO_2 reduction, is most often discussed as a fuel (Ash et al. 1978,1989; Johnson and Leonard 1989; Galecki 1988; Frisbee 1987). However, exothermic reaction with O_2 to produce CO_2 is only one potential use for CO. Alternatives take advantage of the reducing chemistry of CO or CO as a source of carbon (Cotton and Wilkinson 1988; Martin and Webb 1978; Cochran and Fitzgerald 1981; Hughes et al. 1982; Ichikawa et al. 1986; Sneeden 1982*a*) as outlined below.

B. Carbon Monoxide Reduction of Metal Oxides

Typical examples of CO reduction chemistry are shown in reactions (14) through (18) for the potential production of materials useful for structural and power systems. The described chemistry implies that methods of isolating individual oxides (Al_2O_3, Fe_2O_3, Fe_3O_4, SiO_2 ; see Table II) are readily available for the Martian soils. The subjects of mining, separation and beneficiation of extraterrestrial material are important but outside the scope our discussion; information on this subject can be obtained from other sources (see, e.g., Hepp et al. 1991*b*; Steurer 1982; Waldron et al. 1979).

$$Al_2O_3 + 3C \longrightarrow 2Al + 3CO \tag{14}$$

$$Fe_2O_3 + 3CO \longrightarrow 2Fe + 3CO_2 \tag{15}$$

$$SiO_2 + 4CO \longrightarrow SiC + 3CO_2 \quad (16)$$

$$CO + SO_2 + H_2O \longrightarrow C + H_2SO_4 \quad (17)$$

$$2Fe_3O_4 + CO \longrightarrow C + 3Fe_2O_3 \quad (18)$$

TABLE II
Approximate Elemental Composition and Chemistry of Viking 1 Lander Site[a]

Compounds	Percent	Elements[b]	Percent
SiO_2	44.7	O	42.6
Fe_2O_3	18.2	Si	20.9
MgO	8.3	Fe	12.7
Not determined[c]	8.2	Not determined[d]	7.5
SO_3	7.7	Mg	5.0
Al_2O_3	5.7	Ca	4.0
CaO	5.6	S	3.1
TiO_2	0.9	Al	3.0
Cl	0.7	Cl	0.7
		Ti	0.5

[a] Source: (Toulmin et al. 1977).
[b] Calculated from the stoichiometry of oxides determined in Martian soils.
[c] Elements that were too light to be detected ($Z \leq 11$) or present below limits of detectability.
[d] Mostly C, N, or O (from nonmetal oxide sources).

Reaction (14) describes a patented process (Cochran and Fitzgerald 1981) for producing aluminum using carbon and electricity. For each kg of aluminum produced, 3.45 kg of carbon and 32.7 MJ of electrical energy are required. The carbon could be obtained from processes such as reactions (17) or (18) or by electrolysis. The carbon requirement stated for the reaction (14) is more than 5 times as large as the stoichiometry indicates. If the excess carbon is used to provide reaction heat, it may be more reasonable to use direct electrical energy for this purpose. Electrical energy could be obtained from solar cells.

Reaction (15) describes a low-temperature reduction of iron oxide that produces iron and carbon dioxide. This system has been studied using mixtures of CO and H_2 as a reductant (Hughes et al. 1982). Reaction 15 could also apply to a high-temperature reactor operating at blast furnace temperatures.

Reaction (16) summarizes a two-step process for reducing SiO_2 to SiC at temperatures of 1100°C under a CO atmosphere (Ichikawa et al. 1986). Reactions (17) (25°C) and (18) (200°C) describe relatively low-temperature methods for producing carbon. The sulfuric acid and iron oxide can then be recycled to produce O_2 at elevated temperatures (Martin and Webb 1978) or processed further to produce other materials (i.e., [15] to produce iron). Recycling the products from (17) and (18) results in a process for disproportionating CO into carbon and CO_2 .

TABLE III

Thermodynamic Values for CO/H_2 Reactions[a]

Product	Reaction	ΔH (kJ mol^{-1})
Methane	$CO + 3\ H_2 \longrightarrow CH_4 + H_2O$	−206
	$2\ CO + 2\ H_2 \longrightarrow CH_4 + CO_2$	−248
Alkanes	$CO + 2\ H_2 \longrightarrow (-CH_2-) + H_2O$	−165
	$2\ CO + H_2 \longrightarrow (-CH_2-) + CO_2$	−207
	$3\ CO + 2\ H_2O \longrightarrow (-CH_2-) + 2\ CO_2$	−249
Methanol	$CO + 2\ H_2 \longrightarrow CH_3OH$	−90.8
Alcohols	$n\ CO + 2n\ H_2 \longrightarrow C_nH_{2n+1}OH + (n-1)\ H_2O$	−124.8

[a] Adapted from Sneeden (1982*a*).

While the practicality of this chemistry has yet to be established, it is important to explore all aspects of CO chemistry to produce as many critical materials from *in-situ* resources as possible with a minimum of expended energy. Both C and CO are discussed as reducing agents for reactive elements (Al and Si) and Fe, direct electrochemical production of metals is another option for a planet without free reductants. The reader is directed to other sources for a discussion of these and Earth-derived processing options (see, e.g., Carroll 1983; Hepp et al. 1991*b*; Stancati et al. 1991; Steurer 1982; Waldron et al. 1979).

Carbon can also be used to produce acetylene. A standard terrestrial method for producing acetylene uses calcium carbide in a two-step process starting with carbon and calcium oxide:

$$CaO + 3C \longrightarrow CaC_2 + CO \text{ (electric arc)} \quad (19)$$

$$CaC_2 + 2H_2O \longrightarrow C_2H_2 + Ca(OH)_2. \quad (20)$$

While calcium is available on Mars (Table II), it is likely as a silicate or carbonate and which would have to be separated and refined before use. In order to use hydrogen most efficiently, $Ca(OH)_2$ would have to be recycled back to the oxide. Manufacture of acetylene by this process requires electrical power for reaction (19) of about 10 to 11 kW-hr kg^{-1} of acetylene produced (Copenhave and Bigelow 1949). It has the advantage of producing high-purity acetylene with few by-products. Alternative, organic-based production methods are discussed below.

C. Synthesis of Rocket Fuels from Carbon Monoxide

Once carbon monoxide is produced on Mars, further chemistry by use of Fischer-Tropsch type reactions can be used to produce alcohols and higher hydrocarbons, as shown in Table III. This synthetic route can be used to produce almost any organic compound of interest, including hydrocarbon fuels, polymers, and processing feedstock for further use.

Alcohol fuels are a special case. The energy content of alcohols is lower than that of the corresponding hydrocarbons, and thus they have a lower specific impulse. The advantage of alcohols is the great ease of storage. With melting points of −97°C and −115°C, respectively, methanol and ethanol are liquid over nearly the entire Mars temperature range. Once CO has been produced by the reactions discussed previously, methanol can be produced with high selectivity by catalytic hydrogenation (Sneeden 1982*a*):

$$CO + 2H_2 \longrightarrow CH_3OH. \tag{21}$$

This reaction is currently done on a production scale, with a world production on the order of 15 Mton yr^{-1}. Alternatively, methanol can be produced directly from carbon dioxide (Sneeden 1982*a*):

$$CO_2 + 3H_2 \longrightarrow CH_3OH + H_2O. \tag{22}$$

This reaction is less exothermic than the CO reaction, but removal of water from methanol requires an additional step. Ethanol (and higher alcohols) can be produced by Fischer-Tropsch chemistry (Sneeden 1982*a*), although with lower selectivity.

IV. ACETYLENE AS A MARS-DERIVED ROCKET FUEL

A. Hydrocarbon Rocket Fuels and Propellant Mass Leverage

If carbon dioxide is the only resource assumed to be from Mars, then production of hydrocarbon fuels requires hydrogen brought from Earth. It is desirable to maximize the total impulse of fuel produced and minimize the amount of hydrogen required. Thus, it is optimal to utilize a fuel with a minimum hydrogen content. We can define the propellant mass leverage as the mass of propellant produced divided by the mass of Earth-derived components (in this case, hydrogen). The propellant mass leverage of several fuels (for stoichiometric combustion) is shown in Table IV.

It is important to note that mass leverage is not the only factor to be considered in a figure of merit; the specific impulse (I_{sp}) is also quite important, because it determines the amount of fuel required. Hydrogen/oxygen can produce an I_{sp} of up to 470; the hydrocarbon fuels about 375, the alcohols slightly less, and carbon monoxide about 280. Carbon monoxide contains no Earth-derived hydrogen, and so has a hydrogen leverage of infinity, but has low I_{sp}.

Because stoichiometric hydrogen/oxygen fuel is nearly 90% oxygen by mass, a mass leverage of a factor of 9 can be achieved simply by using oxygen derived from carbon dioxide reduction as the oxidizer for Earth-derived hydrogen. A difficulty of liquid hydrogen as a fuel is that it is extremely difficult to store. An additional increase by a factor of 2 in propellant leverage is obtained by use of Mars-derived methane, as shown in reaction (10) above. Use of higher carbon hydrocarbons will gain additional leverage.

B. Production of Acetylene on Mars

Additional gains in propellant leverage can be obtained by synthesis of higher hydrocarbons. Ethane and ethylene produce only modest improvements over methane. Of hydrocarbon fuels, the minimum hydrogen content fuel is acetylene (C_2H_2 , H-C≡C-H). Despite the higher exhaust molecular weight, acetylene has a theoretical vacuum specific impulse slightly better than that of methane. Depending on chamber pressure and area ratio, the specific impulse can be up to about 415 s. This is because the higher exhaust molecular weight is offset by the energy content of the triple bond, resulting in a high combustion temperature. We assume 400 s here (Landis and Linne 1991).

TABLE IV

In-situ Propellant Mass Leverage[a]

Fuel	Reaction Stoichiometry		Leverage	I_{sp}
Hydrogen (Baseline)	H_2	+ 1/2 O_2 (Earth)	1	470
Hydrogen	H_2	+ 1/2 O_2 (Mars)	9	470
Methane	CH_4	+ 2 O_2	20	375
Ethane	C_2H_6	+ 7/2 O_2	24	375
Ethylene	C_2H_4	+ 3 O_2	31	375
Acetylene	C_2H_2	+ 5/2 O_2	53	415
Methanol	CH_3OH	+ 3/2 O_2	20	350
Ethanol	C_2H_5OH	+ 3 O_2	24	350
Carbon monoxide	CO	+ 1/2 O_2	∞	280

[a] See discussion in text. Production of oxygen on Mars is assumed.

Acetylene is a gas at room temperature with a boiling point of −80°C (assuming appropriate pressure to maintain liquid phase), making it even easier to store than methane, which boils at −165°C. Acetylene is thermodynamically unstable with respect to formation of hydrogen and carbon and hence is a potential explosion hazard. Solutions to the hazard include storage as a liquid and/or dilution with another material such as CH_4, N_2 or CO_2 (Copenhave and Bigelow 1949; Duncan et al. 1973). Acetylene can be produced by thermal or electric arc pyrolysis of methane at around 1250°C (Duncan et al. 1973):

$$2CH_4 + 184,000\ \text{kJ} \longrightarrow C_2H_2 + 3H_2. \tag{23}$$

A standard production sequence for acetylene is the partial oxidation of methane, the Sachsse process, where combustion of the methane with oxygen provides the energy required for pyrolysis (Duncan et al. 1973; Morrison and Boyd 1973):

$$6CH_4 + O_2 \longrightarrow 2C_2H_2 + 2CO + 10H_2. \tag{24}$$

This is an industrial reaction sequence which is well developed. The hydrogen can be recycled to methane and reused in reaction (10).

Use of acetylene instead of methane decreases the requirement for hydrogen by another factor of 4 at no reduction in I_{sp}. The high flame temperature of

the oxygen-acetylene flame will require some development of rocket engine technology. An alternative possibility is to reduce the flame temperature by adding an additional component into the fuel mixture.

One attractive possibility is using as fuel a mixture of C_2H_2 and CO. Because CO is formed as a by-product of the reactions used to produce acetylene, no additional chemical technology would be necessary. Burned with oxygen, CO produces a theoretical I_{sp} of ~300 s (Clapp and Scardera 1989), depending on the assumed combustion. One potential problem with this mixture may be the large difference in boiling and melting temperatures of acetylene and carbon monoxide (Copenhave and Bigelow 1949). Another alternative is to further process the C_2H_2 into aromatic hydrocarbons such as benzene (C_6H_6). Benzene and related aromatics can be prepared by catalytic trimerization of acetylenes using inorganic and organonickel catalysts (Jolly and Wilke 1975; Jolly 1982). This issue warrants further study and is not pursued further here (see, e.g., Landis and Linne 1991) and does not have a major impact on the analysis that follows.

A mixture of fuels will produce a composite specific impulse equal to the rms average of the individual specific impulses, weighted by mass. Thus, an equal mass mixture of acetylene and CO would produce a theoretical I_{sp} of about 350 s. This represents some penalty in I_{sp} over the 400 s of acetylene/oxygen alone, but the mixture has only half the requirement for hydrogen brought from Earth, and a much lower flame temperature. The leverage of hydrogen is very large: $< 1\%$ of the fuel mass is Earth-derived hydrogen.

With such high leverage of hydrogen, it becomes possible to consider use of Mars sources of hydrogen. Water is believed to be present in the form of permafrost beneath the surface and in the form of water ice in the polar caps (Duke 1986). Use of such a resource, however, would require both precursor missions to locate the resource, and mining and refining equipment to dig out and purify the water. The Viking Orbiter mapped the water vapor content of the Martian atmosphere (Meyer and McKay 1989) and as a result we now know the atmosphere to be nearly saturated with water vapor, about 0.03% composition by volume, varying with location and season. This results in an amount of precipitable water between 1 and 100 μm (Duke 1986; Kaplan 1988). Water can be precipitated out of the Martian atmosphere by either of two relatively simple mechanical processes: adiabatic expansion (Jones 1990), or isothermal compression (Meyer et al. 1990). Water could be produced from the atmosphere at a rate on the order of one kilogram per 10^6 m^3 of atmosphere processed. This water could then be electrolyzed to produce the hydrogen required for fuel production, and oxygen. The amount of oxygen produced would be sufficient to eliminate the need for reaction (1) above, assuming a stoichiometric fuel ratio. In this case, no reactants need be brought from Earth at all for fuel production on Mars.

C. Effect of Acetylene Use on Mission Mass

ΔV from the surface of Mars to a Mars-Earth transfer orbit is about 5.7 km s^{-1} on the average. At an I_{sp} of 400, a maximum of 23% of the rocket mass can be injected into the transfer orbit. For the CO/acetylene mixture I_{sp} of 350, this is reduced to 18%. While a large fraction of this may be tank weight, it still is likely that the reduction in mass fraction due to lower I_{sp} is more than compensated for by the decrease in requirement of Earth-derived hydrogen.

One figure of merit for savings is initial mass in low Earth orbit. Clearly, mission mass savings will depend on the details of the mission architecture, including such details as use of aerobraking and aerodynamic decelerators, whether a Venus swingby is used, whether a habitat is placed in high or low Mars orbit, etc.

A rough figure of merit can be calculated from the required orbital ΔV. Average ΔV needed to go from low Earth orbit to trans-Mars injection (Hohmann transfer) is 4.21 km s^{-1} (the actual value will depend on the mission year, because the orbit of Mars is significantly eccentric). The average return ΔV from the surface of Mars to trans-Earth injection is 5.68 km s^{-1}. The mass ratio, or total fueled vehicle mass over final mass, is exponential in the mission ΔV:

$$M_i/M_f = \exp(\Delta V/gI_{sp}). \tag{25}$$

Under the most optimistic assumptions, assuming aerobraking at Mars and Earth arrival at no cost in added mass, an I_{sp} of 450 s (LH_2/LOX), and no allowances for fuel tank mass and engines, every ton of mass injected from Mars to Earth requires 2.6 tons of fuel on Mars. Shipping this fuel to Mars would require an additional 4.3 tons of fuel in low Earth orbit. Manufacturing return fuel on Mars will thus reduce the initial mass in low Earth orbit by nearly a factor of 7. More pessimistic assumptions adding weight for tanks, aerobrake mass, etc, will increase the advantage of Mars-manufactured propellant even further. Lower values of I_{sp}, as would be required for space-storable propellants, will also increase this factor, while habitats or vehicle mass left in Mars orbit or left behind on the surface will decrease the factor. In any case, however, manufacturing fuel from *in-situ* resources on Mars results in huge savings in mass.

V. CONCLUSIONS

We propose a novel chemical approach to *in-situ* material utilization (ISMU) for manned Mars missions. Carbon dioxide reduction can be carried out not only using purely chemical means but also by taking advantage of another accessible Martian resource, ultraviolet photons. The proposed light-assisted carbon dioxide reduction chemistry has the potential to produce lightweight, mechanically simple devices for oxygen production. It is anticipated that

the invention of new devices and systems studies will result in an enhanced understanding of photochemical applications for ISMU.

A parallel study of carbon monoxide as a reductant in processing native oxides on Mars is underway (Hepp et al. 1991*a*). Novel processes involving carbon monoxide in metal and semimetal oxide reduction may enhance the value of the products of carbon dioxide reduction and conserve hydrogen for other uses.

Relatively simple and well-understood chemical reactions can be used to produce hydrocarbon rocket fuels on Mars from hydrogen. Use of such a process allows an amount of fuel to be produced on Mars which is nearly 100 times the mass of hydrogen brought from Earth. If such a process produces the return propellant for a manned Mars mission the required mission mass in low Earth orbit is reduced significantly over a system using all Earth-derived propellants.

A simple processing sequence such as manufacturing oxygen on Mars would reduce fuel requirements on Mars by a factor of 5 if CH_4 propellant is brought entirely from Earth. The simplest and technologically best established processing sequence is the Sabatier process (reaction [10]) followed by water electrolysis (reaction [6]). This produces CH_4 /O_2 with a 12:1 mass leverage. The mass leverage can be increased to 20:1 by the addition of an O_2 production process.

A further decrease in the requirement for Earth-derived hydrogen is found if the carbon monoxide produced as a by-product of acetylene production is also used as a fuel component. Propellant brought from Earth could be entirely eliminated if a convenient source of hydrogen on Mars such as atmospheric water could be used.

If only one single idea is to be emphasized, it is that the carbon dioxide atmosphere of Mars is a significant, abundant resource for manufacturing critical materials on Mars. There are many possible chemical sequences to utilize the CO_2 and reduced by-products CO and carbon. The processes discussed for making hydrocarbons and alcohols from CO_2 and reducing metal oxides with CO and carbon are thermodynamically feasible but may not be practical. There are likely to be other sequences that are more useful.

At this stage in the definition of manned Mars missions, it is important to explore the full range of the known chemistries of simple carbon compounds. A more thorough exploitation of easy-to-obtain resources will enhance the potential for *in-situ* resource utilization. Further progress in simplifying *in-situ* manufacturing technology will provide for lighter, less-expensive missions and increase the likelihood of manned planetary exploration in our lifetime.

REFERENCES

Appelbaum, J., and Flood, D. J. 1990. *Solar Radiation on Mars, Update 1990*, NASA TM-103623.

Ash, R. L., Dowler, W. L., and Varsi, G. 1978. Feasibility of rocket propellant production on Mars. *Acta Astronaut.* 5:705–724.

Ash, R. L., Werne, J. A., and Haywood, M. B. 1989. Design of a Mars oxygen processor. In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 479–487.

Baker, D., and Zubrin, R. 1990. Mars direct: Combining near-term technologies to achieve a two-launch manned Mars mission. *J. Brit. Interplanet. Soc.* 43:519–525.

Bryan, J. C., Geib, S. J., Rheingold, A. L., and Mayer, J. M. 1987. Oxidative addition of carbon dioxide, epoxides, and related molecules to WCl_2 $(PMePh_2)_4$ yielding tungsten (IV) oxo, imido, and sulfido complexes. Crystal and molecular structure of $W(O)Cl_2(CO)(PMePh_2)_2$. *J. Amer. Chem. Soc.* 109:2826–2828.

Carroll, W. F., ed. 1983. *Research on the Use of Space Resources*, JPL Publ. 83-36.

Clapp, W. M., and Scardera, M. P. 1989. Applications of in-situ carbon monoxide-oxygen propellant production at Mars. In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 513–537.

Clark, B. C. 1989. Survival and prosperity using regolith resources on Mars. *J. Brit. Interplanet. Soc.* 42:161–166.

Cochran, C. N., and Fitzgerald, N. M. 1981. Energy Efficient Production of Aluminum by Carbothermic Reduction of Alumina. U. S. Patent 4,299,619.

Copenhave, J. W., and Bigelow, M. H. 1949. *Acetylene and Carbon Monoxide Chemistry* (New York: Reinhold), pp. 1–6; 311–329.

Cotton, F. A., and Wilkinson, G. 1988. *Advanced Inorganic Chemistry*, 5th ed. (New York: Wiley), pp. 243–244; 1226–1239.

Darensbourg, D. J., and Kudaroski, R. A. 1983. The activation of carbon dioxide by metal complexes. *Adv. Organomet. Chem.* 22:129–168.

DeLaet, D. L., del Rosario, R., Fanwick, P. E., and Kubiak, C. P. 1987. CO_2 chemistry and electrochemistry of a binuclear cradle complex of Ni(0), Ni_2 $(\mu\text{-CNMe})(CNMe)_2(PPh_2CH_2PPh_2)_2$. *J. Amer. Chem. Soc.* 109:754–758.

DeLaet, D. L., Lemke, F. R., Gao, J., and Kubiak, C. P. 1988. Photochemical activation of CO_2. Transient absorbance kinetic studies of the addition of CO_2 to a metal to bridging ligand charge transfer state of a binuclear Ni(0) complex. *J. Amer. Chem. Soc.* 110:6904–6906.

Duke, M. B. 1986. Mars resources. In *Manned Mars Missions, Working Group Papers*, vol. 2, NASA M-002, pp. 519–531.

Duncan, D. A., Detz, C. M., and Sargent, H. B. 1973. Acetylene. In *Kirk-Othmer Encyclopedia of Chemical Technology*, vol. 1, 3rd ed. (New York: Wiley), pp. 192–237.

French, J. R. 1989. Rocket propellants from martian resources. *J. Brit. Interplanet. Soc.* 42:167–170.

Frisbee, R. H. 1987. Mass and power estimates for Mars in-situ propellant production systems. Paper presented at the 23rd AIAA/ASME/SAE/ASEE Joint Propulsion Conf., San Diego, Ca. AIAA Paper 87-1900.

Galecki, D. L. 1988. In-situ propellant advantages for fast transfer to Mars. Paper presented at the 24th AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conf., Boston, Mass. AIAA Paper 88-2901.

Gilbert, J. A., Eggleston, D. S., Murphy, W. R., Jr., Geselowitz, D. A., Gersten, S. W., Hodgson, D. J., and Meyer, T. J. 1985. Structure and redox properties of the

water-oxidation catalyst $[(bpy)_2(OH_2)RuORu(OH_2)(bpy)_2]^{4+}$. *J. Amer. Chem. Soc.* 107:3855–3864.

Hepp, A. F., Landis, G. A., and Linne, D. L. 1991*a*. Material processing with hydrogen and carbon monoxide on Mars. In *Space Manufacturing 8, Proc. of the Tenth Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 150–158.

Hepp, A. F., Linne, D. L., Landis, G. A., Groth, M. F., and Colvin, J. E. 1991*b*. Production and use of metals and oxygen for lunar propulsion. Paper presented at the AIAA/NASA/OAI Conf. on Advanced Space Exploration Technologies, Cleveland, Oh. AIAA Paper 91-3481 and references therein.

Hughes, R., Kam, E. K. T., and Magadam-Zadeh, H. 1982. The reduction of iron ores by hydrogen and carbon monoxide and their mixtures. *Thermochimica Acta* 59:361–377.

Ichikawa, H., Takada, K., and Saito, M. 1986. Manufacture of Silicon Carbide Whiskers. Japanese Patent 61,291,496.

Jakosky, B. M., and Haberle, R. M. 1990. Year-to-year instability of the Mars south polar cap. *J. Geophys. Res.* 95:1359–1365.

Johnson, S. W., and Leonard, R. S. 1989. Manned Mars missions and extra-terrestrial resource engineering test and evaluation. In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 455–468.

Jolly, P. W. 1982. Nickel catalyzed oligomerization of alkynes and related reactions. In *Comprehensive Organometallic Chemistry*, vol. 8, eds. G. Wilkinson, F. G. A. Stone and E. W. Abel (Oxford: Pergamon Press), pp. 649–670.

Jolly, P. W., and Wilke, G. 1975. *The Organic Chemistry of Nickel*, vol. 2. (New York: Academic Press), pp. 94–116.

Jones, D. 1990. Isentropic processing for in-situ propellant production on Mars. Presented at Case for Mars IV, Boulder, Co., June.

Kaplan, D. 1988. *Environment of Mars 1988*, NASA TM-100470.

Kubiak, C. P., and Ratliff, K. S. 1991. Approaches to the chemical, electrochemical, and photochemical activation of carbon dioxide by transition metal complexes. *Israeli J. Chem.* 31:3–15.

Landis, G. A., and Appelbaum, J. 1990. Photovoltaic power system operation on Mars. Presented at Case for Mars IV, Boulder, Co., June.

Landis, G. A., and Appelbaum, J. 1991. Design considerations for Mars PV power systems. In *Proc. of the 21st IEEE Photovoltaic Specialists Conf.*, vol. 2 (New York: Inst. of Electrical and Electronic Engineers), pp. 1263–1270.

Landis, G. A., and Linne, D. L. 1991. Acetylene fuel from atmospheric CO_2 on Mars. *J. Spacecraft and Rockets*, in press.

LeCompte, M. A., and Stets, J. P. 1991. Propellant manufacturing on Mars: Issues and implications. In *Space Manufacturing 8, Proc. of the Tenth Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 369–379.

Linne, D. L., Roncace, J., and Groth, M. F. 1990. Mars in situ propellants: Carbon monoxide and oxygen ignition experiments. Paper presented at the 26th AIAA/ASME/SAE/ASEE Joint Propulsion Conf., Orlando, Fl. AIAA Paper 90-1894.

Martin, L. R., and Webb, H. M. 1978. Transportable fuels via a carbon resource recovery technique. Paper presented at the AIAA/ASERC Conf. on Solar Energy: Technology Status, Phoenix, Ariz. AIAA Paper 78-1780.

Meyer, T. R., and McKay, C. P. 1989. The resources of Mars for human settlement. *J. Brit. Interplanet. Soc.* 42:147–160.

Meyer, T. R., Brown, L., and Tarantino, J. 1990. Atmospheric water on Mars: Energy extraction estimates. Presented at Case for Mars IV, Boulder, Co., June.

Morrison, R. T., and Boyd, R. N. 1973. *Organic Chemistry*, 3rd ed. (Boston: Allyn and Bacon).

Ni, J., and Kubiak, C. P. 1992. The activation of carbon-oxygen bonds. *Advances in Chemistry Series*, no. 230, eds. W. R. Moser and D. W. Slocum (Washington, D. C.: American Chemical Soc.), pp. 515–528.

Oro, J., and Holzer, G. 1979. The photolytic degradation and oxidation of organic compounds under simulated martian conditions. *J. Molec. Evol.* 14:153–160.

Owen, T. C., Biemann, K., Rushneck, D. R., Biller, J. E., Howarth, D. W., and Lefleur, A. L. 1977. The composition of the atmosphere at the surface of Mars. *J. Geophys. Res.* 82:4635–4639.

Palmer, D. A., and van Eldik, R. 1983. The chemistry of metal carbonate and carbon dioxide complexes. *Chem. Rev.* 83:651–731.

Ramohalli, K. N. R., and Sridhar, K. R. 1991. Extraterrestrial materials processing and related transport phenomena. Paper presented at the 29th Aerospace Sciences Meeting, Reno, Nev. AIAA Paper 91-0309.

Ratliff, K. S., Leutz, R. E., and Kubiak, C. P. 1992. Carbon dioxide chemistry of the trinuclear complex $[Ni_3(\mu_3\text{-CNMe})(\mu_3\text{-I})(dppm)_3]I$. *Organometallics* 11:1986–1988.

Richter, R. 1981. Basic investigation into the production of oxygen in a solid electrolyte process. Paper presented at the 16th AIAA Thermophysics Conf., Palo Alto, Ca. AIAA Paper 81-1175.

Sneeden, R. P. A. 1982*a*. Organic synthesis where carbon monoxide is the unique source of carbon. In *Comprehensive Organometallic Chemistry*, vol. 8, eds. G. Wilkinson, F. G. A. Stone and E. W. Abel (Oxford: Pergamon Press), pp. 19–100.

Sneeden, R. P. A. 1982*b*. Reactions of carbon dioxide. In *Comprehensive Organometallic Chemistry*, vol. 8, eds. G. Wilkinson, F. G. A. Stone and E. W. Abel (Oxford: Pergamon Press), pp. 225–283.

Stancati, M. L., Jacobs, M. K., Cole, K. J., and Collins, J. T. 1991. In Situ Propellant Production: Alternatives for Mars Exploration. Report No. SAIC-91/1052 for NASA Contract NAS3-25809.

Steurer, W. H. 1982. *Extraterrestrial Materials Processing*, JPL Publ. 82-41.

Thekaekara, M. P., Kruger, R., and Duncan, C. H. 1969. Solar irradiance measurements from a research aircraft. *Applied Optics* 8:1731–1732.

Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christian, R. P., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.

Waldron, R. D., Erstfeld, T. E., and Criswell, D. R. 1979. Overview of methods for extraterrestrial materials processing. In *Space Manufacturing 3, Proc. of the Fourth Princeton/AIAA Conf.: Space Manufacturing Facilities*, eds. J. Grey and C. Krop (Washington, D. C.: AIAA), pp. 113–127.

Wu, J., and Kubiak, C. P. 1983. Electrocatalytic oxidation of carbon monoxide in a CO/O_2 fuel cell. *J. Amer. Chem. Soc.* 105:7456–7457.

Wu, J., Fanwick, P. E., and Kubiak, C. P. 1987. Carbon dioxide chemistry of a binuclear Ir(0) complex: Oxygen atom transfer to a coordinated aryl isocyanide ligand and the structure of $[Ir_2(\mu\text{-CO})(\mu\text{-H})(C(O)NHR)_2(CNR)_2\ (Me_2PCH_2PMe_2)_2]Cl$ (R=2,6-$Me_2(C_6H_3)$). *Organometallics* 6:1805–1807.

Zubrin, R. M., Baker, D. A., and Gwynne, O. 1991. Mars direct: A simple, robust, and cost effective architecture for the space exploration initiative. Paper presented at the 29th Aerospace Sciences Meeting, Reno, Nev. AIAA Paper 91-0326.

UTILIZING MARTIAN RESOURCES FOR LIFE SUPPORT

CHRISTOPHER P. McKAY
NASA Ames Research Center

THOMAS R. MEYER
Boulder Center for Science and Policy

PENELOPE J. BOSTON
Complex Systems Research

MARK NELSON and TABER MACCALLUM
Space Biospheres Ventures

and

OWEN GWYNNE
SETI Institute

The human exploration of Mars can benefit from the availability of resources on the Martian surface that can be used to provide life support consumables. The key compounds O_2, buffer gas (Ar/N_2), and H_2O are available on Mars and can be extracted from the atmosphere. Water may also be available from soil water and ground ice. The soil could be used as radiation shielding and could provide useful plant growth medium. Fairly autonomous processes can be designed to extract and stockpile Martian consumables. The utilization of O_2, obtained from Martian atmospheric CO_2 will probably be practical even on initial human missions. The ability to utilize these materials in support of a human exploration effort allows missions that are more robust and economical than would otherwise be possible.

I. INTRODUCTION

All the resources needed to support life are available in some form on the surface of Mars and these resources can be used to extend human exploration of the Martian surface. Current plans for the human exploration of Mars rely on the utilization of Martian resources to a varying degree depending on the mission design. Key requirements are the production of life support consumables and chemical fuels. Providing air and water for life support is, of course, a general requirement of all human missions. The basic needs of

human beings for life support can be fairly well defined at this time and are unlikely to change. The quantities required will scale with the crew size and mission duration, and may be higher in field operations where the level of physical effort, leakage, and the impracticallity of recycling may be inherently greater than at a base. By contrast, the need for on-site fuel production depends to a much larger degree on the scenarios and technology assumed in terms of the interplanetary spacecraft, Mars ascent/descent vehicles and surface mobility vehicles.

In the early phases of human exploration it is probable that all of the resources needed for mission success will be transported from Earth—particularly those needed for the return flight home. However, on-site production of life support consumables could allow for extended mission operations (such as EVAs) that could increase the scientific return of the initial missions greatly. Mars is a complex and diverse planet with billions of years of geological and possibly biological history to explore. Furthermore, the scale of the natural features on Mars range from meters to thousands of kilometers (e.g., Valles Marineris is over 3000 km long). Thus, any methods that augment the range and capability of the initial human explorations are certain to enhance the scientific return of the mission greatly (Stoker et al. 1992). Furthermore, the assessment and utilization of resources to provide life support at the Mars base will provide the foundation for determining if Mars has the inventories of water, nitrogen and carbon dioxide necessary for the fabrication of a habitable state for the entire planet—a possible long-term goal of human settlement (McKay et al. 1991).

In this chapter we consider how water, air, and food for human exploration of Mars can be supplied with indigenous resources in the initial missions. In the first sections of the chapter we review the availability of resources on Mars and the likely requirements for human exploration. We then consider systems for the production of water, air and food in a variety of applications ranging from individual life support units in space suits to full scale greenhouses. We put particular emphasis on the production of food in greenhouse for two reasons: (1) this topic has not yet received a level of study on par with existing studies for providing water and air, and (2) the ongoing experiment at Biosphere 2 provides for the first time a relevant study in closed ecosystems involving humans. We also briefly consider areas related to life support such as power and transportation on the surface of Mars.

II. RESOURCES ON MARS FOR LIFE SUPPORT

A. Sources of Water on Mars

It is clear that Mars was at one time a very wet planet. There is extensive geologic evidence that Mars had a lot of water freely flowing on its surface in the past (see, e.g., Carr 1981,1987). The two main forms of evidence are the valley networks and the runoff channels. The valley networks are dendritic drainage systems some of which appear to be quite old, ~3.8 Gyr. Current

climatic conditions cannot support liquid water, hence the runoff channels are evidence that the conditions on Mars have changed considerably over time. The outflow channels, apparently caused by rapid large-scale flows of water, indicate that Mars' inventory of water is large.

There is no direct measurement of the amount of water on Mars and estimates vary considerably. Table I compares the amount of CO_2, N_2 and H_2O in Mars' atmosphere today, the amount that Mars should have if it formed from the same material that the Earth did (Earth scaling), and the range of scientific estimates of Mars volatile inventory (McKay and Stoker 1989). Of the values listed in Table I, perhaps the most interesting are those based upon geologic evidence for fluvial erosion; Carr (1986) has estimated that Mars had an initial endowment of water that would correspond to a layer over 1 km thick covering the planet. If Mars had this much water, then models of planetary formation suggest that it also had large amounts of CO_2 and N_2, much larger than that which is observed in the present atmosphere.

TABLE I
Estimates of the Amount of Key Compounds on Mars

	CO_2 (hPa)	N_2 (hPa)	H_2 (meters)
Amount in Mars' present atmosphere	~ 10	0.2	$\sim 7 \times 10^{-6}$
Original endowment, scaling based on Earth	27,000	300	1,200
Original endowment, range of current estimates	200–20,000	2–300	6–1,000

It is probable that Mars still retains most of its volatiles. Although there has been non-negligible loss to space, the current loss rate of volatiles is insufficient to deplete these initial estimated reservoirs. If the current rate of atmospheric escape has remained constant for water (as 2H and O) at 6×10^7 cm^{-2} s^{-1} (Liu and Donahue 1976; McElroy et al. 1977) and N_2 at 5.6×10^5 cm^{-2} s^{-1} (Fox and Dalgarno 1983); the total loss over the past 4.5 Gyr would be 2.5 m of water and 1.4 hPa (1 hPa = 10^2 Pa = 1 mbar) of N_2. While escape rates, particularly of N_2 may have been higher in the past (as much as 30 hPa of N_2 is estimated to have escaped; McElroy et al. 1977), the current view is that the bulk of the Martian water is tied up in subsurface reservoirs, probably as permafrost (Squyres and Carr 1986).

In terms of accessibility for utilization by human explorers, we divide the sources of water on Mars into four categories: (1) the atmosphere; (2) the polar caps; (3) a deep subsurface permafrost or aquifer layer; (4) the soil. All of these sources have been suggested as resources that could be utilized

to support a human presence on Mars.

Atmospheric Water. Throughout most of the year, the Martian atmosphere contains as much water as it can hold with respect to nighttime temperatures. Water frost was occasionally observed at Viking Lander 2 site at 48° latitude (Hart and Jakosky 1986). However, at the low temperatures, even saturated conditions imply a very small amount of total water. Figure 1 in the Chapter by Jakosky and Zent shows the column amount of water determined from the Viking Orbiter as a function of latitude over the Martian year. Values range from 1 to 90 precipitable microns (Jakosky and Farmer 1982)—the equivalent thickness of the layer if the water condensed into liquid. These values are 10,000 times smaller than the typical value on the Earth.

Atmospheric moisture on Mars is found preferentially in the northern latitudes above 60°, and at low elevations in both hemispheres. Northern hemisphere spring is the wettest season. The annual average water vapor in the Martian atmosphere as a function of latitude and longitude (Jakosky and Farmer 1982) shows a strong correlation with surface topography with lower elevations having much more water. The lowest place in the southern hemisphere, Hellas Basin (−45° S, 290° W) has the highest annual averaged water content in the southern hemisphere and at its peak has more than twice the atmospheric water than any other place on the planet (Jakosky and Farmer 1982). Lower elevations on Mars tend to be wetter both in the northern and southern hemisphere. The bottom of Hellas Basin is more than 4 km below the planetary reference level and thus, the pressure there is ~44% higher than the planetary average—reaching as high as 15 hPa.

Although on the planetary scale (see Table I), the total amount of water in the Martian atmosphere is small, ~1.3 km^3 (~360×10^9 gallons), it is huge compared to the requirements of a human research base. For comparison, this amount of water could supply the needs of McMurdo Station, the main US research base in Antarctica for 33,000 yr—assuming peak summer usage with 1000 people on station and no recycling (McKay 1985). Clearly, there is enough Martian atmospheric water to supply the needs of a research base. The main problem with extracting water from the Martian atmosphere is the low concentration. Table II shows the fraction of water at saturation in the Martian atmosphere over a range of frost points (the temperature at which the atmosphere is saturated). Pollack et al. (1977) determined a frost point temperature at Viking 2 of 195 K by monitoring atmospheric fog formation at night through its effect on the view of the Martian moon Phobos. The frost point was determined to be 191 K at Viking 1 and 196 K at Viking 2 by early morning plateaus in the daily temperature curves (Hess et al. 1977). These frost points correspond to a partial pressure of water vapor over ice of between 0.04 and 0.09 Pa. At the time of these measurements the total vertical column abundances of water vapor (Farmer et al. 1977) were equivalent to a layer of liquid 6 to 18 μm thick at the Viking sites. If this column amount of water was distributed with a scale height equal to that of the atmosphere, it would represent a surface pressure of 0.02 to 0.07 Pa of water. For a nominal partial

pressure of water of 0.1 Pa, in order to extract 1 kg of water the amount of Martian air that would need to be processed is 19,500 kg or 10^6 m^3—assuming 100% extraction efficiency.

TABLE II
Maximum Water Concentration in the Martian Atmosphere vs Frost Point Temperature[a]

Frost Temperature (°C)	Atmospheric Water	Partial Pressure of Water
−40°	1.6%	12.9 pa
−50°	0.49	3.94
−60°	0.135	1.08
−70°	323 ppm	0.26
−80°	66.7 ppm	0.053
−90°	11.7 ppm	0.009

[a] Table adapted from Meyer and McKay (1984).

Extraction of atmospheric water, even for long-term or large-scale utilization, would not present problems of depletion of a local resource because the atmospheric water would be resupplied by exchange with the polar caps and regolith. One clear advantage of the use of atmospheric water during the exploration phase is that this source is certainly the best characterized at the present time and will be even better understood after the Mars Observer mission.

Polar Caps. The Martian polar caps are composed of two components; the CO_2 seasonal cap that forms in winter and the permanent cap, composed predominately of water ice, that persists throughout the year (see, e.g., Clifford 1987). The permanent caps are the potential sources of water. The permanent north polar cap is very roughly circular and centered about the pole. It has a diameter of about 1000 km, extending down to 80° N, with occasional ice deposits as far south as 75° N. The permanent south polar cap is much smaller, only about 350 km in diameter and is centered at 86° S, 30° W. The thickness of the polar ice deposits is uncertain. Estimates based upon the topology of surface features on the caps suggest that the northern cap has a thickness of 4 to 6 km at its center near the north pole. Similarly the south polar deposits are believed to reach a maximum thickness of 1 to 2 km. The wintertime temperatures at the polar caps tend to be close to the CO_2 frost points, ~150 K, and can reach upwards of 200 K in the summer.

Deep Subsurface Water. Mars almost certainly has a permafrost that is rich in water. Geomorphological and topographical features such as craters with rims that look like mud slides suggest the fluid-like flow of soil material and are thought to be the result of significant ice present in the soil. Such features tend to be found poleward of 40° in both hemispheres (Squyres and Carr 1986), indicating permafrost in these locations. The geomorphology can be used to crudely estimate the fraction of ice in these soils, and values of 5 to 10% have been suggested as possible. It was previously thought that this

near-surface water-rich permafrost does not occur near the equatorial regions. This is illustrated in Fig. 4 of the Chapter by Jakosky and Zent. However, recent modeling work by Paige (1992) suggests that, for possible ranges of surface conditions, ground ice could be stable at the equator. It is important to note that, unlike the strongly bound water of hydration observed in the soils at the Viking Lander sites, this water would be water ice. This ground ice does not necessarily exist right at the surface (note that it was not found within the top few centimeters at the northern Viking 2 site at 48° N) but would exist at depths below the limit of summertime defrosting—meters. At depths of a kilometer or more, temperatures reach levels high enough that melting would occur. It is possible that liquid water reservoirs exist at these depths. If the water contained salts, the resulting solution would be liquid at colder temperatures, resulting in liquid existing closer to the surface.

Soil Water. The Viking Landers found that water was released from the soil upon heating to high temperatures (Biemann et al. 1977). The samples were obtained from the near surface and not kept cold before analysis—thus, any easily-bound or adsorbed water may have been lost. Upon onset of heating, most samples lost little H_2O at 200°C, but all evolved significant and approximately equal amounts at temperatures of 350°C and 500°C. The total amount of water released at 350°C was ~0.3% by weight and further heating to 500°C resulted in a net release of ~1%. Because these data are from the Viking GCMS (Gas Chromatograph Mass Spectrometer) which was not designed to detect water, these results are highly uncertain (Biemann et al. 1977) but the soil water may be an available resource and needs further evaluation. Nevertheless, the soil water content is believed to be higher than the GCMS results and may be as much as 15%. Stoker et al. (see their Chapter) suggest that salts could contain up to 10% water of hydration and clays, if present, could also contain large quantities of interlayer absorbed water.

The soil at the two Viking landing sites had very similar elemental composition which could suggest that a uniform mantle of æolian dust covers most of the Martian surface. Likewise, the similarity of the water analysis at the two Viking sites would suggest that similar amounts of water would be contained in the dust in other locations. The high temperatures required to release this water indicates that it is fairly tightly bound and it is therefore probably not affected by the range of temperatures occurring on the surface of Mars. Thus, it is probable that windblown soil anywhere on the planet will have similar properties (see, e.g., the Chapter by Stoker et al.).

The possibility of fairly stable brine solutions at or just below the Martian surface has been suggested (Zisk and Mouginis-Mark 1980) as a possible explanation of strong Earth-based radar reflectivities that are seasonally variable and indicate unusual smoothness near Solis Lacus and Noachis-Hellespontus. These same areas were suggested by Huguenin et al. (1979) to be possible sites of near-surface water activity. The stability of near-surface brine solutions has been discussed by Zent and Fanale (1986), who conclude that certain brines may exist for periods of 10 Myr under current conditions.

B. Atmospheric Sources of Breathable Air

The Martian atmosphere is composed of predominantly carbon dioxide with small, but potentially useful percentages of nitrogen, argon and water vapor. The values for the atmospheric gases on Mars are listed in Table III and are in terms of the volume fraction of the atmospheric components, except for the entry for dust which is the "clear" sky value expressed as a mass fraction. From these gases alone, it is possible to prepare water and breathable air.

TABLE III
Composition of Mars (Owen et al. 1977) and Earth Air (Verniani 1966)

Gas	Concentration	
	Mars	Earth
CO_2	95.3%	0.03%
N_2	2.7	78.08
Ar	1.6	0.93
O_2	0.13	20.9
H_2O	0.03*	~2*
CO	0.07	0.12 ppm
Ne	2.5 ppm	18 ppm
Kr	0.3 ppm	1.1 ppm
Xe	0.08 ppm	0.087 ppm
O_3	0.03 ppm*	40 ppb
Other		<7 ppm
Dust	~10 ppm/m*	—

* Variable.

The total pressure of the Martian atmosphere averages ~8 hPa and varies from 6 to 10 hPa with season as the major constituent, CO_2, condenses to form the seasonal polar caps.

C. The Soil of Mars

The soil is an important potential resource on Mars from the life support point of view. Its most important and immediate use may be to provide mass for radiation shielding and material for use in construction, including greenhouses. In addition, it could be a substrate for plant growth, and a source of water, industrial process materials, and a variety of other useful compounds (Clark 1979; Chapter by Stoker et al.).

Our understanding of the soil composition and structure on Mars is based primarily on the direct measurement of the elemental composition and soil properties at the two Viking landing sites. In addition, analysis of the SNC meteorites (thought to have originated on Mars) and groundbased spectral observations provide information for current models of the Martian soil (Chapter by Stoker et al.). Based upon photogeological analysis of Orbiter imaging

the likely distribution of many geologic units has been identified. Units of volcanic (presumably basaltic), sedimentary, and windblown origin have been identified (Carr 1981). The elemental abundance of the loose debris at the Viking landing site was determined by the Viking X-Ray Fluorescence Experiment for elements with atomic number 12 (Mg) and above. Thus, there was no direct detection of such important elements as H, C, N, or O. The results are shown in Table IV (Clark et al. 1977). Most abundant were silicon (21% by mass) and iron (13%), both assumed to be in the form of oxides SiO_2 and Fe_2O_3, respectively. Magnesium, aluminum, sulfur, calcium, and titanium were also detected at the one to a few percent level. Sulfur is about 10 to 100 times more abundant than in terrestrial rocks and soils. Phosphorus, while not listed because its signal is masked by S and Si, is nevertheless thought to be present (Toulmin et al. 1976) especially as it is found in the SNC meteorites at concentrations of ~0.3% by weight (Chapter by Stoker et al.). Thus, all the major elements of life (C, H, N, O, P, S) are present on the surface of Mars, and most of the trace elements required, such as Fe, Mg, Al, have been directly detected as well. Stoker et al. (in their Chapter) do suggest that potassium (0.08% by weight on Mars compared with ~0.83% on Earth) may be in short supply.

TABLE IV

Elemental Composition of the Viking 1 Lander Site

Element	Percent by Mass
Mg	5.0 ± 2.5
Al	3.0 ± 0.9
Si	20.9 ± 2.5
S	3.1 ± 0.5
Cl	0.7 ± 0.3
K	< 0.25
Ca	4.0 ± 0.8
Ti	0.5 ± 0.2
Fe	12.7 ± 2.0
L*	50.1 ± 4.3
X**	8.4 ± 7.8
Rb	< 30 ppm
Sr	60 ± 30 ppm
Y	70 ± 30 ppm
Zr	< 30 ppm

* L is the sum of all elements not directly determined.

** If the detected elements are all present as their common oxides (Cl expected) then X is the sum of components not directly detected, including H_2O, NaO, CO_2, and NO_x.

The mineralogical state of the elements was not directly detected, but based upon chemical considerations, the mineralogy of most of the elements

has been inferred (Toulmin et al. 1977; Clark et al. 1982). These results are listed in Table IV in the Chapter by Stoker et al. The fine-grained debris at the Viking landing sites probably consists of palagonite (a weathered basalt) or iron-rich clays, such as montmorillonite, with minor amounts of kieserite ($KSO_4 \cdot H_2O$), calcite ($CaCO_3$), and rutile (TiO_2) (Chapter by Stoker et al.).

The Martian soil also contains significant quantities of adsorbed volatiles. Trace amounts of O_2 (70–790 nanomoles cm^{-3}) and other gases were released upon humidification (Oyama and Berdahl 1977,1979). The soil is an improbable source of O_2 because to supply 1 kg of oxygen by humidification would require over 40 m^{-3}, about 60 metric tons, of soil (assuming 790 nmoles O_2 cm^{-3}). In addition, when heating to 500°C, soil samples released from 50 to 500 ppm CO_2 (Biemann et al. 1977) which is not like to be a useful resource given the CO_2 in the atmosphere.

The absence of organic material and the release of O_2 upon wetting in the Martian soil have led to the suggestion that the Martian surface layer contains oxidizing compounds such as H_2O_2 as well as other peroxides and superoxides (Klein 1978,1979). It is hypothesized that these could be produced by ultraviolet radiation (190–300 nm) reaching the Martian surface in the absence of a significant ozone layer. If all of the oxygen (up to 790 nanomoles cm^{-3}) released from the soil was from the decomposition of H_2O_2, then the concentration of this oxidant in the soil would be ~30 ppm. The pH of the soil is not known.

Models for the formation of Mars and its evolution suggest that there may be huge deposits of carbonate materials (as yet undetected). These carbonates may be locally concentrated in ancient dry lake basins and valleys on the Martian surface (McKay and Nedell 1988). Based on the Viking elemental analyses (Toulmin et al. 1977; Clark et al. 1982), the most likely carbonates to occur on Mars are $CaCO_3$ (calcite), $MgCO_3$ (magnesite), $CaMg(CO_3)_2$ (dolomite), $FeCO_3$ (siderite), and possibly $MnCO_3$ (rhodochrosite). Based upon estimates of the initial endowment of N_2 on Mars (McKay and Stoker 1989), there may also be large quantities of nitrates associated with the putative carbonate deposits. In addition to volatiles bound into the soil in mineral phases (such as carbonates and nitrates), there may be significant amounts of CO_2 and H_2O adsorbed in a porous regolith layer. There may be as much as 4000 kg m^{-2} adsorbed CO_2, which is equivalent to a surface pressure of over 100 hPa (Fanale and Cannon 1979).

III. REQUIREMENTS FOR LIFE SUPPORT

A. Human Needs

Rooted in their basic physiology, humans share a requirement for food, O_2, and a source of potable water. While these requirements vary among individuals and with the situation, they are fairly well understood and are almost certainly not subject to change over the next many decades over which human exploration of Mars will commence. As centuries of recorded history have

shown, humans who disregard any of these essential requirements for long do so at their peril. Thus, we can be reasonably certain of the need for life support consumables at a human base on Mars and begin plans for how to provide these materials.

A representative list of life support inputs and outputs for a typical human being appears in Table V. Not shown on this list are possible demands for water in a Mars habitat module that are not related to life support, such as washing. These needs for water depend sensitively on the design of the habitat, the technology incorporated in such items as clothing, dishes and toilets, as well as the level of fastidiousness of the crew. Obviously, these demands are highly uncertain; MacElroy et al. (1992) estimate the demand for washing water to be 20 kg per person-day. The life support requirements listed in Table V are similar to those used in a previous analysis by Meyer and McKay (1984) except that they included wash water (at 1.2 kg/person-day) in their analysis (see Figs. 1 and 2 of Meyer and McKay 1984).

TABLE V

Average Requirement for Human Life Support[a]

Inputs (kg/person-day)		Outputs (kg/person-day)	
Oxygen	0.83	Carbon dioxide	1.0
Food		Sold waste	
dry weight	0.668	dry weight	0.1
water content	0.552	water content	0.2
Drinking water	1.85	Urine	1.5
Water added to food	0.72	Perspiration	1.82
Total Mass	4.62		4.62

[a] Adapted from MacElroy et al. (1992).

In addition to the consumables listed in Table V, breathable air must contain an inert buffer gas to prevent oxygen toxicity and spontaneous combustion hazards. On Earth, of course, the buffer gas is N_2. CO_2 becomes toxic to the blood buffer system at levels of about 10 hPa (Parker and West 1973). The only candidates for buffer gas on Mars appear to be N_2 and Ar, the only inert gases with any appreciable concentration in the atmosphere (see Table III). Nitrogen and Argon are both narcotic (well known to deep sea divers) but only above certain levels, ~300 kPa for Ar and ~400 kPa for N_2 (Parker and West 1973). These are well above the partial pressures they would have within the Mars habitat, 100 kPa (Earth normal) or less. Although the buffer gas is not consumed directly it will have to be resupplied against loss by leakage and airlock operations. Estimating this requirement is difficult but based upon early space station designs Meyer and McKay (1984) suggested a value of 0.56 kg/person-day.

It is important to note that humans are a net source of water as well as CO_2. Fundamentally this is a result of the metabolism of food to provide energy. This is best illustrated by considering the oxidation of organic material (represented elementally by CH_2O):

$$CH_2O + O_2 \longrightarrow CO_2 + H_2O + 8.7 \times 10^{-19}\ \text{J} \qquad (1)$$

From the values in Table V it is apparent that the average human represents a net source of 0.4 kg/person-day. Thus, given efficient washing methods and recycling of wash and human waste water it is entirely possible that a human habitat supplied with food would not have any requirement for water whatsoever. Working with the figures in Table V, we can determine the level of recycling of human waste water (3.32 kg/person-day) required to supply the potable water demands (2.57 kg/person-day). If only 77% of the human waste water were returned as potable water, then the human life support requirements for water *per se*, would be zero. Even if the food is completely dehydrated (water demand now 3.122 kg/person-day), the potable water requirements can still be met by recycling, now at a level of 94%.

In the initial phase of human exploration of Mars, it is likely that food will be transported from Earth and that levels of recycling in the range of 90 to 95% can be achievable (Meyer and McKay 1984; MacElroy et al. 1992). Thus, water will probably not be required as a life support consumable during this phase—although it may have uses other than life support.

Of course, the above analysis hides the fact that water is being transported to the habitat in the form of wet or dry organic matter—food. If food is produced locally on Mars, then the water that is locked up in the chemical structure of the food must be provided to the plants. When a greenhouse is included in the mass budget of the life support system, then water is required to make up for inevitable losses and nonrecyclable waste material. Meyer and McKay (1984) have considered in some detail a possible life support system for a Mars habitat with and without a greenhouse included. They estimate that with a greenhouse the requirement for water is about 0.23 kg/person-day to make up for leakage and non-recyclable waste.

B. Plant Growth and Greenhouses

In this section we consider the requirements for the long term operation of a greenhouse as part of a Mars settlement. There are many questions that have not yet been answered regarding how best to grow food on the surface of Mars.

Mars has a rotation rate similar to that of the Earth and we assume that its surface gravity (0.38 g) would be adequate for long-term biological adaptation. Because Mars is 1.52 times farther from the Sun than is Earth, the amount of sunlight incident on the planet is only 43% of the terrestrial value. Even during dust storms the amount of sunlight incident on the Martian surface (Meyer and McKay 1989; Chapter by Haberle et al.) is greatly in excess of

the minimum light levels needed for photosynthesis, so light will not be a limiting factor, *per se*.

However, a major uncertainty is whether crops can grow in greenhouses on the surface exposed to the harsh radiation environment. Mars' thin atmosphere (0.2 m equivalent thickness in the zenith compared to 10 m on the Earth) and negligible magnetic field provide little protection against cosmic radiation and solar flare protons. Humans must be protected from this radiation, particularly during times of strong solar flares, but 5 m of Mars dirt covering the habitat section of the base would provide shielding equivalent to the column mass of the Earth's atmosphere. However, growing food requires large areas that may be impractical to cover with dirt. Furthermore, if the greenhouses are buried, light could become limiting and may have to be provided by artificial lighting or perhaps piped in with mirrors or fiber optics.

From an engineering point of view, surface greenhouses seem more desirable than those that are buried; they are simpler to build and maintain, can operate on natural lighting, and they can readily be expanded to accommodate base development. Engineering issues with surface greenhouses include making the shell strong enough to withstand the pressure differential, repairing any leaks, and keeping the greenhouse warm. The feasibility of surface greenhouses may depend on the development of crop strains that are tolerant of the Martian radiation environment. Radiation events that destroy the crops may be rare (perhaps once per solar cycle, 11 yr) and it may not be impractical to accommodate an occasional severe loss of crops given an ample reserve. Obviously, this method cannot be applied to radiation shielding requirements for humans or for plants used as a source of future seeds (Boston 1985). For surface greenhouses relying on natural light, strong transparent plastics must be developed that can withstand the rigorous thermal regime and radiation—particularly the ultraviolet—on Mars. These are important areas for further work if a Mars base is to become self sustaining. No design studies of Martian greenhouses that include all the relevant aspects have been published.

Regardless of how the greenhouses are built, the plants will require CO_2, nutrient fertilizer, and water. The CO_2 can be obtained directly from the Martian air, only slightly modified because most plants show improved growth under high CO_2 concentrations. The CO_2 level on Earth today ($\sim$0.35 hPa) is low enough that certain plants (particularly C-3 plants) are often limited in their ability to obtain CO_2 for photosynthesis and $\sim$0.15 hPa is a practical lower limit for CO_2 (McKay et al. 1991). There is no clear upper limit on CO_2 for plants (Rogers et al. 1983) although concentrations over 1 hPa can have adverse effects on certain species (Hicklenton and Jolliffe 1980; Huber et al. 1984). Seckback et al. (1970) reported on photosynthetic algae that thrive under pure CO_2. While some O_2 is required by most plants for aerobic mitochondrial respiration, most higher plants actually prefer oxygen levels well below the current value. The requirements vary depending on the plant but, in general, net primary production increases as O_2 is reduced from the ambient level of 210 hPa until a point is reached where the O_2 concentration

is low enough (~20 hPa) to cause metabolic complications (Salisbury and Ross 1985). It may be possible to adapt plants to accept even lower O_2 levels (~1 hPa) because the mitochondrial enzyme that requires O_2 has such a strong affinity for O_2 that it can function with O_2 levels of 0.1 hPa (Salisbury and Ross 1985). Nitrogen is required for all organisms and the amount of atmospheric nitrogen must be high enough to provide for biological nitrogen fixation. Bacteria can fix nitrogen at levels of 10 hPa and less (Klingler et al. 1989).

As discussed above, food production in a greenhouse requires that water be provided as a life support resource, whereas without a greenhouse, water recycling is probably sufficient to provide life support water given the net production of water by metabolism.

All the inorganic nutrients necessary for plant growth can probably be obtained from the Martian soil (Chapter by Stoker et al.), and the waste products from the habitat can supply the organic fertilizer. Using the soil as the plant growth medium may be more advantageous than hydroponics on Mars because the soil is ubiquitous and may provide a simpler, more reliable method. Without samples of the Martian soil it is difficult to determine how much initial soil preparation will be necessary, however, it is possible that it will need to be washed of salts, oxides, and toxins before being adjusted for pH and fertilized.

The average temperature within the greenhouse must be warm enough for liquid water and we assume that it must be $>10^{\circ}C$. Plants and anaerobic microbes can tolerate total pressures much lower than Earth normal (~1000 hPa). In addition to a few hPa of N_2 and O_2 and 0.15 hPa of CO_2, the atmosphere need only support the vapor pressure of water at greenhouse temperatures (6.1 hPa at $0^{\circ}C$). Thus, pressures as low as 10 hPa may be permissible for an atmosphere suitable for plants and anaerobic microorganisms. The important advantage of operating surface greenhouses at low pressure is that this reduces the structural strength required to maintain the pressure differential against the extremely low ambient pressure. The disadvantage is the necessity for humans to work in a space suit and the need to couple common systems with the high pressure habitat. Preliminary work on low-pressure greenhouses (Boston 1981; Schwartzkopf and Mancinelli 1991; Daunicht and Brinkjans 1992; André and Massimino 1992) supports this possibility but more long term studies with diverse plants are needed.

IV. PROVIDING WATER AND BREATHABLE AIR

Breathable air requires two components: oxygen and a suitably inert buffer gas. On Mars the most likely candidate buffer gas is a mixture of Ar and N_2, which together make up about 5% of the Martian atmosphere. Oxygen and water can also be obtained from the atmosphere. Table VI lists the amount of Mars air that must be processed to obtain these materials.

TABLE VI

Mass Flow Rates and Energy Requirements for Resource Extraction

Resource	Source	Amount to be processed	Energy required
1 kg Ar/N_2	Atmosphere	1,700 m^3	9.4 kW-hr[a] ideal = 0.1 kW-hr
1 kg H_2O	Atmosphere	10^6 m^3 (T_{frost}=196 K)	103 kW-hr[a] ideal = 0.3 kW-hr
	Hydrated soil	100 kg	10 kW-hr[b]
1 kg O_2	Atmospheric CO_2	70 m^3	12 kW-hr[c] ideal = 2.5 kW-hr
	Soil oxidant	60 tons (O_2 = 790 nmoles cm^{-3})	?

[a] Meyer and McKay (1984).
[b] Chapter by Stoker et al.
[c] See text.

Obtaining breathable air and water from the Martian atmosphere is appealing for many reasons. This resource is ubiquitous and fairly well described with present knowledge. Processing gases and liquids is much simpler and more reliable than handling solids. Furthermore the concentration of CO_2 and Ar/N_2 in the atmosphere are virtually constant with day and season. However, water vapor varies considerably.

It is useful to consider the idealized work to obtain the needed life support resource from the atmospheric reservoirs. These values are essentially the entropy of mixing for extracting H_2O and Ar/N_2 and the chemical binding energy for obtaining O_2 from CO_2. As shown in Table VI these values are quite modest and provide a measure of the possibilities that real engineering systems must attempt to reach.

The ideal separation work, (taken to be the entropy of mixing) for water in the Martian atmosphere can be determined from the following equation (Barron 1966)

$$W = RT\left[f \ln \frac{1}{f} + (1-f) \ln \frac{1}{1-f}\right] \tag{2}$$

where W is the ideal separation work per mole of Martian air, f is the water vapor mixing ratio, and R (8317 J $kmole^{-1}$ K^{-1}) is the gas constant. For values of $f << 1$ this reduces to

$$\simeq RTf[1 - \ln f] \tag{3}$$

The work expressed as per kg of water separated is then

$$W/\text{kg} - \text{H}_2\text{O} \simeq \frac{RT}{m}[1 - \ln f] \tag{4}$$

where $m = 18$ is the molecular weight of water. For a frost point temperature of 196 K—a typical value at the Viking Landing sites as discussed above—$f \simeq 10^{-4}$ and we find that the entropy of separation of water in the Martian atmosphere is about 10^6 J (= 0.3 kW-hr) per kg of water. Note that the theoretical energy of separation expressed per unit product depends only on the logarithm of the mixing ratio and is therefore not sensitive to this value. The energy required to change this water vapor from a gas at 8 hPa at 196 K to a liquid at 1000 hPa and 15°C is small by comparison. Thus, the theoretical minimal work required to extract water from the Martian atmosphere is quite modest, 0.3 kW-hr per kg of water. The ideal work required to overcome the entropy of mixing to extract the Ar and N_2 from the CO_2 (computed using Equation 2, with $f = 0.05$, $T = 215$ K) is 0.1 kW-hr per kg of buffer gas produced. The chemical energy required to separate CO_2 into CO and O_2 (this is less than required to produce C and O_2) is ~2.5 kW-hr per kg of O_2 produced.

Turning now to realistic engineering designs for obtaining these resources, Meyer and McKay (1984) have done a detailed study of the extraction of buffer gas from the Martian atmosphere and their design is still the best available for providing this important life support resource.

The main problem with providing buffer gas on Mars is in removing the unwanted CO_2. Fortunately CO_2 condenses at a much higher temperature than either Ar or N_2 and so this provides a practical method for separating these gases. The properties of CO_2 dictate the operational pressures and temperatures of the condensation system. In order to liquefy the CO_2, thereby simplifying the separation, it is necessary to compress the Martian air to above the triple point pressure of CO_2, 5.1 atm. The compression is the main energy requirement in the process. At this pressure most of the CO_2 condenses. The remaining CO_2 is then condensed to solid form by cooling via a Joule-Thompson expansion from 5 atm to a working pressure of 2 atm. Meyer and McKay (1984) estimated that the amount of work required to produce 1 kg of buffer gas by this method was 9.4 kW-hr. This is over a factor of 100 larger than the theoretical work because large amounts of Martian air must be compressed to extract the needed buffer gas. Meyer and McKay (1984) also considered the possibility of separating the Ar and N_2 if there was some requirement for this individual gases. This problem is similar to the separation of O_2 and N_2 from liquid air.

The same basic compression, cooling and condensation process was used by Meyer and McKay (1984) in their design for extraction of water from the Martian atmosphere. Here the amounts of Martian air that must be processed are truly enormous (see Table VI). However, since water condenses at a higher temperature than CO_2 it is not necessary to liquefy the CO_2. Thus, the bulk

of the Martian air that is compressed remains gaseous and, after the trace levels of water are removed by condensation, this compressed gas can be used to power regenerators as it expands back to the ambient pressure. This regenerates some of the energy expended in the initial compressors. However, because of the wide dynamic range of the compression scheme (1,500 to 1) the ratio of the size of the compressor stages varies considerably and the small stages at the high pressure end become a source of inefficiency. The total energy required to produce 1 kg of water was calculated to be 103 kW-hr. Clapp (1985) optimized the process suggested by Meyer and McKay (1984) and obtained a energy requirement of 69 kW-hr/kg water.

In these designs the main problem is the compression of the low pressure Martian air to a high operational systems pressure (5–10 atms). Table VII (reproduced from Meyer and McKay 1984) shows one scheme for achieving a 12 atm pressure on Mars. The process has been divided into four stages to improve the overall efficiency and to reduce the pressure rise in each stage to practical levels (~6×). Note that the efficiencies span the range for very large compressors (0.8; Kerrebrock 1977) to very small compressors (0.2; Ash et al. 1982).

TABLE VII

Calculation of Work to Isentropically Compress Mars Air From 0.008 to 12 Atmospheres in 4 Cycles[a]

Cycle	1	2	3	4
P inlet (atm)	0.008	0.05	0.3	2
P outlet	0.05	0.3	2	12
T inlet (K)	253	253	253	253
T outlet	373	371	378	366
Ideal work (kJ/kg)	106	106	110	101
Assumed efficiency	0.8	0.6	0.4	0.2
Real work	132	177	275	505

Total ideal work = 432 kJ/kg = 0.118 kW-hr/kg of Mars air processed
Total real work = 1089 kJ/kg = 0.303 kW-hr/kg of Mars air processed
Overall efficiency = 0.38

[a] From Meyer and McKay 1984.

Because the buffer gas and water are minor gases in the Martian atmosphere it is possible that direct compression and cooling of the entire Martian gas mixture—basically the calculation presented by Meyer and McKay (1984)—are not the optimal way to separate these gases. It is possible that separation methods that directly select the target molecules are more efficient. There has been some preliminary work along these lines employing molecular sieves to extract the buffer gas (Jones et al. 1985). However, there is not a good estimate of the energy required to achieve the separation. For water there may be a variety of ways to preferentially select this molecule

from the Martian atmosphere. This could include adsorption by molecular sieve substances, solution with hygroscopic solids like P_2O_5, hydration of salts such as gypsum ($CaSO_4$), and possibly even hydration of the natural Martian soil. In each case, the hydrated material would have to be processed (presumably by heating) and the collected water harvested. The now anhydrous material would then be re-exposed to the Martian environment where it would passively collect a new batch of water. This concept is appealing in its simple and direct approach to extracting water but the kinetics and energetics of the system have not yet been sufficiently analyzed to allow a meaningful comparison to the compression/cooling methods.

Providing O_2 on Mars is much more straightforward than obtaining buffer gas or water since O_2 is to be found in the major constituent of the atmosphere, CO_2. Since the initial suggestion of Ash et al. (1978) there has been considerable effort directed toward understanding O_2 production from the Martian atmosphere. This work has resulted in a prototype system now operating at the University of Arizona (K. Ramohalli, personal communication). This system is based on the zirconia membrane separation process in which CO_2 at high temperatures (1200°C) thermally decomposes to CO and O_2. The O_2 is then preferentially transported across the zirconia membrane (Ash et al. 1978,1982). The currently operating unit produces 130 liters of O_2 per day useing 48 watts of electrical and 1,800 watts thermal power—over 250 kW-hr kg^{-1} (Ramohalli and Sridhar 1992). As the unit is improved, particularly in terms of insolation and packaging the thermal heat load should drop significantly. Frisbee (1986) estimated the power required for mature O_2 production systems based on this approach to be about 12 kW-hr kg^{-1}.

V. PROVIDING FOOD

The most suitable organisms for food production in the Martian facilities primarily include higher plants because of our heavy cultural reliance on them and the antiquity of our experience with them. Single-cell protein production using algae and bacteria can provide some nutritional needs (see, e.g., Kihlberg 1972; Kharatyan 1978). Microbial processing of plant materials using many traditionally based techniques can also add to the palatability and nutritional value of foodstuffs. There may also be a place for aquaculture of such densely growing fish species as *Tilapia* and edible crustacean and shellfish. For Biosphere 2, methods of combining the raising of *Tilapia* fish with rice production have been developed. This system provides the high-protein fish at little incremental cost in required space, while supplying some of the fish's diet from the rice plants and associated water ferns (Leigh et al. 1987). The large grazing animal species commonly consumed on Earth may be inappropriate for several reasons. First, they are relatively inefficient in converting plant material to nutrients ranging from 1:20 to 1:6 depending upon species. Secondly, they require large amounts of space (with the possible exception of fowl) and lastly they share many of our pathogens with us thus

presenting a distinct risk of acting as hosts for potential disease organisms. However, raising of animals, if not an initial part of a Mars food system, should be considered as a later evolutionary phase. Including animals has an advantage in that they can consume many of the otherwise inedible portions of crops and provide appetizing food products with a high energy content. In Biosphere 2 to increase animal efficiencies and lessen fodder requirements pygmy varieties of pig and milk goat are used along with chickens.

For the growth of plants, sufficient and appropriate illumination, monitoring and control of the gas environment specifically tailored to the optimum production of the plants, and monitoring and control of plant nutrients, soil and pathogenic microbes and removal of organic volatiles are all key elements. The plant growing chambers can be well-integrated with the human environments, but the capability to isolate them, should some problem arise, is imperative. On the other hand it may prove advantageous to separate the greenhouses from the habitat if the pressure in the greenhouse is significantly below 1 atm.

Because of the wealth of experience in greenhouse, hydroponic, aeroponic, and conventional growing techniques that we possess, selecting the particular technique most appropriate for various food species should be possible with large scale simulations. Selecting the most promising types for early missions will primarily be a matter of taste (psychological benefit), ability to grow densely in an intercropped way, overall caloric and nutritional value, and minimal inedible fraction. Experiments to assess the response of candidate species to novel Martian conditions like 0.38 g and the illumination and radiation regimes likely to be received are imperative.

Hydroponics (the growth of plants in an inert substrate supplied by liquid nutrient) and its variant aeroponics (such nutrient is sprayed on exposed hanging roots) are appropriate for many species. However, some plants do not respond well to these growing conditions or undesirable changes occur in the seed stocks after repeated generations are produced this way (Boston 1985). For such plants, fairly conventional but highly compact soil agriculture is still an option. Soil based agriculture offers advantages in that there is less need for chemical sources of nutrients, greater buffering properties, more direct recycling of waste products back to the soil through composting techniques and lastly the presence of the soil microbial community will assist in the metabolism of toxic gases. The beneficial properties of soil in maintaining atmospheric composition is established (Bohn and Bohn 1986) and is used in the Biosphere 2 project as discussed below (Frye and Hodges 1990; Nelson et al. 1992*a*) and they could be used in space applications.

As discussed above, future missions which enlighten us as to the exact composition of Martian soil will provide more foundation for the development of suitable agricultural soil from the indigenous Martian material. At this point, one can only speculate that raw materials processed to a greater or lesser degree are available to construct appropriate soil.

The plants ultimately best suited to grow in Martian greenhouses can

be developed on Mars after Martian installations reach a certain degree of maturity. Eventually, the seed stocks and cultures of life support organisms will be maintained and developed at the Mars facilities. Of course, initial missions will need to bring the ancestors of these organisms with them from Earth.

Clearly, the amount of human labor in production of food materials must be kept to a minimum. Based on experience in an intensive indoor agriculture system at North Carolina State University, a one person-day is required to tend 40 m^2 of growing area (D. Raper, personal communication). That area is approximately equivalent to the amount of food material required for 2 people. Thus, one person could support two people. This seems reasonable for life on Earth, but in the Mars situation, this human labor burden must be reduced by automation of growing and tending tasks whenever feasible.

Microbial fermentation and other common industrial methods of large–scale production of biomass is a mature technology (see, e.g., Steele and Stowers 1991). The science of food processing has reached a high degree of refinement. Adaptation of these techniques to the kitchens of Mars will be simply a matter of doing it rather than inventing or developing new technologies. Traditional production of microbially altered foods like tempeh, tofu, fermented beverages, raised grain products like bread, and non-milk dairy products like soy cheese are also mature technologies which simply await adaptation to the particulars of a given mission scenario and design. Another application of microbes in the Martian food chain is as animal feed for aquaculture species of higher animals.

The same general technologies employed to produce single-cell protein can be applied to the production of chemicals and other substances. This technology is a common industry here on Earth with considerable effort spent trying to find or develop microbes which will produce useful products. One recent example is the conversion of solar energy to hydrogen peroxide by a cyanobacterium species (Roncel et al. 1989). The authors of that paper hope to make this an alternative to abiotic solar energy technology. Hydrogen peroxide is a potentially very useful substance on Mars (Clark 1985).

Aquaculture is another area of growing terrestrial experience. The benefits of aquaculture include the fact that because terrestrial aquatic organisms are already adapted to live in a buoyant medium, they are less likely to be adversely affected by the low gravity of Mars than land animals. Secondly, they are more different from humans biologically, especially non-fish species and share less in the way of potential pathogens with humans. Thirdly, the organisms are small, numerous and relatively fast growing compared to large ungulates. Lastly, the soiled water from the organisms can be used as fertilizing material for plant growth, and many aquaculture specimens could be fed on vat-grown or pond-grown microbes. The components are relatively easy to integrate into the overall life support system with minimal waste.

A. Biosphere 2

As discussed in the previous section there are many issues and design choices that must be resolved in adapting terrestrial agriculture to life support needs on Mars. Given the complexity of even artificial carefully controlled ecosystems, it is clear that the most fruitful approach is to construct large closed systems designed to support humans and experiment with these systems. Fortunately such experimentation has begun at the Biosphere 2 project.

The Biosphere 2 project has taken as its goal the development of bioregenerative life support systems needed for both near and long-term space applications including those focused on Mars. The construction of the facilities have been completed and long-term experiments involving humans sealed within the system have begun. The human habitat and the six other ecosystems—rain forest, savannah, desert, marsh, ocean, and intensive agriculture—have a combined volume of some 200,000 m^3 and a projected area of 13,000 m^2. The essential statistics of the various ecosystems within the Biosphere are detailed in Nelson et al. (1992*b*). Some of the features of the Biosphere 2 relevant to a Mars life support system are briefly described below.

Biosphere 2 maintains air quality within the human habitat by passing the air through soil bed reactors consisting of the main agricultural soil area. In the soil toxic gases are metabolized by the soil microbiota (Nelson et al. 1992*a*). The air volume of the entire facility can be pumped through the soil bed reactor in less than a day. Within a biosphere there is a considerable reservoir of biological carbon and potential for microbial soil activity releasing this carbon as CO_2 into the closed atmosphere as it cycles through the soil bed reactors. Maintaining the CO_2 levels within a closed biosphere below the limits for humans while at same time scavenging toxins will be a difficult task requiring monitoring and control of the atmospheric processors. Experience with Biosphere 2 will provide the first real tests of such systems. The Biosphere 2 project has advanced systems that recycle nutrients from human, animal and plant residue and return them to the soil instead of using chemical fertilizers. Crop production is based on integrated pest management techniques such as beneficial insects, small crop plots, intercropping and crop rotation and alternation of the cultivator of a crop used to keep insect and disease problems to a minimum. A total of 150 species of tropical and warm climate food crops are grown throughout the year, with about 50 crops in cultivation at any one time. These include papaya, figs, bananas, guavas, strawberries, a wide range of vegetables, grains, animal fodder crops, sweet potatoes, and even coffee trees. Most waste processing (including human solid and liquid wastes and kitchen graywater) is achieved by the microorganisms and plants in the marsh ecosystem. The Biosphere 2 contains a variety of monitoring and recording systems to allow the system state to be accurately characterized. This will produce valuable information that can be used to refine future biospheric systems.

The detailed study of the operation of Biosphere 2 should be an impor-

tant prelude to the creation of potentially long-term and evolutionary human habitations on Mars. It may provide valuable base-line data on the stability and sustainability of agricultural and ecosystems units of small size and of the health and performance of people in created biospheric systems.

TABLE VIII
Uses of Martian Resources for Life Support

Location	Uses
Precursor missions	Test O_2 production Accumulate water
Personal space suit and/or short range rover	O_2 production
Initial base	O_2 production Buffer gas (Ar/N_2) production
Long range rover or remote field camp	O_2 production Buffer gas (Ar/N_2) from base
Operational base with greenhouses	O_2 production Buffer gas (Ar/N_2) production Water production Food production

VI. STRATEGIES FOR PROVIDING RESOURCES ON MARS

As discussed above and shown in Table VI a good deal of power is needed to produce useful resources from the naturally occurring compounds on Mars. Power will almost certainly be the limiting factor in developing a base on Mars or in extending the range of human or robotic exploration. For the production of life support consumables the power requirements are much lower than for fuel production but the requirements for reliability and safe operations are more severe. The power systems considered for Mars are: solar, nuclear, and wind. However, in the case of nuclear power reactors the problem of removal of the spent fuel and other radioactive wastes is not solved at present and may preclude the use of these systems on the surface of Mars. Of these systems, solar certainly is the most reliable—even during dust storms (Meyer and McKay 1989; Chapter by Haberle et al.)—while at the same time being safe to maintain. Furthermore, failure modes in solar systems tend to be gradual and not catastrophic as they can be in nuclear or wind systems. Thus, it may be the case that critical habitat power systems such as the production of life support consumables will be operated on solar power. From Table VI, it is possible to estimate that the power required to provide for a 8 person crew

on an initial base on Mars utilizing oxygen and buffer gas production would be about 60 kW-hr. (see also Table 1 of Meyer and McKay 1984). This power level would require a solar array about 10 m on a side, based on the systems design for a solar power system of 450 kW-hr day^{-1} of Haberle et al. (see their Chapter).

The use of Martian resources to provide life support will depend on the size and type of system on Mars, ranging from personal space suits to a fully operational base. In Table VIII we have suggested what types of resources may be practical to use in a variety of systems. Clearly, the first step is testing on Earth and on precursor missions.

We suggest that the first and easiest system for use on Mars is O_2 production primarily because of the low mass flow rates and power required (see Table VI). We feel that such a system could be deployed as portable and field-maintainable units for use on short range rovers and possibly even on personal space suits. For larger installations buffer gas production becomes the next practical step. As discussed in detail above, water production is probably not necessary in these systems because of the water produced by human metabolism and assuming a reasonable efficiency of recycling. Eventually, in a operational base with greenhouse food production water extraction will be necessary.

In summary, the basic conclusions that derive from our analysis of production of life support consumables on Mars are: (1) Life support needs are well defined and unlikely to change. (2) Initial baseline missions, designed to succeed without use of Martian resources can be augmented by the successful production of life support consumables. (3) Oxygen production from CO_2 in the Martian atmosphere will probably be the first use of extraterrestrial resources for life support. (4) Ar/N_2 in the Martian atmosphere provides a useful source of buffer gas. (5) Input of water as a life support consumable is only needed if food is produced on-site (n.b. water may be needed for other uses). (6) Utilization of Martian resources will be limited by available power; critical life support functions should rely on solar power. (7) Biosphere 2 provides the first test of food production within a complete full scale integrated life support system large enough to be able to support humans.

Finally, we note that the use of Martian resources to provide water and breathable air and to grow food crops (under other than Earth normal conditions) is the first step toward the concept of making Mars a self-sufficient second home for life. These initial efforts naturally lead to reviving the Martian atmosphere and hydrosphere on a planetary scale and allowing the whole planet to participate in the life support function.

REFERENCES

Andre, M., and Massimino, D. 1992. Growth of plants at reduced pressures: Experiments in wheat technological advantages and constraints. *Adv. Space Res.*

12(5):97–106.
Ash, R. L., Dowler, W. L., and Varsi, G. 1978. Feasibility of rocket propellant production on Mars. *Acta Astronaut.* 5:705–724.
Ash, R. L., Richter, R., Dowler, W. L., Hapson, J. A., and Uphoff, C. W. 1982. Autonomous oxygen production for a Mars return vehicle. Paper presented at the 33rd Congress of the International Astronautical Federation, Sept., Paris. IAF publication #IAF-82-210.
Barron, R. F. 1966. *Cryogenic Systems* (New York: McGraw Hill).
Biemann, K., Oro, J., Toulmin, P., III, Orgel, L. E., Nier, A. O., Anderson, D. M., Simmonds, P. G., Flory, D., Diaz, A. V., Rushneck, D. R., Biller, J. E., and LaFleur, A. L. 1977. The search for organic substances and inorganic volatile compounds in the surface of Mars. *J. Geophys. Res.* 82:4641–4658.
Bohn, H. L., and Bohn, R. K. 1986. Soil bed scrubbing of fugitive gas releases. *J. Environ. Sci. Health* A21:561–569.
Boston, P. J. 1981. Low pressure greenhouses and plants for a manned research station on Mars. *J. Brit. Interplanet. Soc.* 34:189–192.
Boston, P. J. 1985. Critical life support issues for a Mars base. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 287–331.
Carr, M. H. 1981. *The Surface of Mars* (New Haven, Conn.: Yale Univ. Press).
Carr, M. H. 1986. Mars: A water-rich planet. *Icarus* 68:187–216.
Carr, M. H. 1987. Water on Mars. *Nature* 326:30–35.
Clapp, M. 1985. Water supply for a manned Mars base. In *Case For Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 557–566.
Clark, B. C. 1979. The Viking results—The case for man on Mars. *Adv. Astronaut. Sci.* 38:263–278.
Clark, B. C. 1985. The H-atom resource on Mars. In *Case For Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 527–535.
Clark, B. C, Baird, A. K., Rose, H. J., Jr., Toulmin, P., III, Christan, R. P., Kelliher, W. C., Castro, A. J., Rowe, C. D., Keil, K., and Huss, G. R. 1977. The Viking X-Ray fluorescence experiment: Analytical methods and early results. *J. Geophys. Res.* 82:4577–4594.
Clark, B. C., Baird, A. K., Weldon, R. J., Tsusaki, D. M., Schnabel, L., and Candelaria, M. P. 1982. Chemical composition of martian fines. *J. Geophys. Res.* 87:10059–10067.
Clifford, S. M. 1987. Polar basal melting on Mars. *J. Geophys. Res.* 92:9135–9152.
Daunicht, H.-J., and Brinkjans, H.-J. 1992. Gas exchange and growth of plants under reduced air pressure. *Adv. Space Res.* 12(5):107–114.
Fanale, F. P., and Cannon, W. A. 1979. Mars: CO_2 adsorption and capillary condensation on clays—significance for volatile storage and atmospheric history. *J. Geophys. Res.* 84:8404–8414.
Farmer, C. B., Davies, D. W., Holland, A. L., LaPorte, D. D., and Doms, P. E. 1977. Mars: Water vapor observations from the Viking orbiters. *J. Geophys. Res.* 82:4225–4248.
Fox, J. L., and Dalgarno, A. 1983. Nitrogen escape from Mars. *J. Geophys. Res.* 88:9027–9032.
Frisbee, R. H. 1986. Mass and Power Estimates for Martian In-Situ Propellant Production Systems. JPL Rept. D-3648.
Frye, R., and Hodges, C. 1990. Soil bed reactor work of the Environmental Research Laboratory of the University of Arizona in support of the research and development of Biosphere 2. In *Biological Life Support Technologies*, eds. M. Nelson and G. Soffen, NASA CP-3094 and Synergetic Press (Oracle, Ariz.), pp. 33–40.
Hart, H., and Jakosky, B. M. 1986. Composition and stability of the condensate observed at the Viking Lander 2 site on Mars. *Icarus* 66:134–142.

Hess, S. L., Henry, R. M., Leovy, C. B., Ryan, J. A., and Tillman, J. E. 1977. Meteorological results from the surface of Mars: Viking 1 and 2. *J. Geophys. Res.* 82:4559–4574.

Hicklenton, P. R., and Jolliffe, P. A. 1980. Alterations in the physiology of CO_2 exchange in tomato plants grown in CO_2-enriched atmosphere. *Canadian J. Botany* 58:2181–2189.

Huber, S. C., Rogers, H. H., and Mowry, F. L. 1984. Effects of water stress on photosynthesis and carbon dioxide partitioning in soybean glycine-max cultivar bragg plants grown in the field at different CO_2 levels. *Plant Physiol.* 76:244–249.

Huguenin, R. L., Miller, K. J., and Harwood, W. S. 1979. Frost-weathering on Mars: Experimental evidence for peroxide formation. *J. Molec. Evol.* 14:103–132.

Jakosky, B. M., and Farmer, C. B. 1982. The seasonal and global behavior of water vapor in the Mars atmosphere: Complete global results of the Viking atmospheric water detector experiment. *J. Geophys. Res.* 87:2999–3019.

Jones, D., Webb, C. F., LaPointe, M. R., Hart, H. M., and Larson, A. 1985. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 537–666.

Kerrebrock, J. L. 1977. *Aircraft Engines and Gas Turbines* (Cambridge: MIT Press).

Kihlberg, R. 1972. The microbe as a source of food. *Ann. Rev. Microbiol.* 26:427–466.

Kharatyan, S. G. 1978. Microbes as food for humans. *Ann. Rev. Microbiol.* 32:301–327.

Klein, H. P. 1978. The Viking biological experiments on Mars. *Icarus* 34:666–674.

Klein, H. P. 1979. The Viking mission and the search for life on Mars. *Rev. Geophys. Space Phys.* 17:1655–1662.

Klingler, J. M., Mancinelli, R. L., and White, M. R. 1989. Biological nitrogen fixation under primordial martian partial pressures of dinitrogen. *Adv. Space Res.* 9(6):173–176.

Leigh, L., Fitzsimmons, K., Norem, M., and Stumpf, D. 1987. An introduction to the intensive agriculture biome of Biosphere II. In *Space Manufacturing 6, Proc. of the Eighth Princeton/AIAA/SSI Conf.: Nonterrestrial Resources, Biosciences, and Space Engineering*, eds. B. Faughnan and G. Maryniak (Washington, D. C.: AIAA), pp. 76–81.

Liu, S., and Donahue, T. M. 1976. The regulation of hydrogen and oxygen escape from Mars. *Icarus* 28:231–246.

MacElroy, R. D., Kliss, M., and Straight, C. 1992. Life support systems for Mars transit. *Adv. Space Res.* 12(5):159–166.

McElroy, M. B., Kong, T. Y., and Yung, Y. L. 1977. Photochemistry and evolution of Mars' atmosphere: A Viking perspective. *J. Geophys. Res.* 82:4379–4388.

McKay, C. P. 1985. Antarctica: Lessons for a Mars exploration program. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 79–87.

McKay, C. P., and Nedell, S. S. 1988. Are there carbonate deposits in Valles Marineris, Mars? *Icarus* 73:142–148.

McKay, C. P., and Stoker, C. R. 1989. The early environment and its evolution on Mars: Implications for life. *Rev. Geophys. Space Phys.* 27:189–214.

McKay, C. P., Toon, O. B., and Kasting, J. F. 1991. Making Mars habitable. *Nature* 352:489–496.

Meyer, T. R., and McKay, C. P. 1984. The Atmosphere of Mars—Resources for the exploration and settlement of Mars. In *Case for Mars*, ed. P. J. Boston (San Diego: Univelt), pp. 209–232.

Meyer, T. R., and McKay, C. P. 1989. The resources of Mars for human settlement. *J. British Interplanet. Soc.* 42:147–160.

Nelson, M., Allen, J. P., and Dempster, W. F. 1992*a*. Biosphere 2: A prototype project for a permanent and evolving life system for a Mars base. *Adv. Space*

Res. 12(5):211–217.

Nelson, M., Leigh, L., Alling, A., MacCullum, T., Allen, J., and Alvarez-Romo, N. 1992*b*. Biosphere 2 test module: A ground-based sunlight-driven prototype of a closed ecological life support system. *Adv. Space Res.* 12(5):151–156.

Owen, T., Biemann, K., Rushnek, D. R., Biller, J. E., Howarth, D. W., and LaFleur, A. L. 1977. The composition of the atmosphere at the surface of Mars. *J. Geophys. Res.* 82:4635–4639.

Oyama, V. I., and Berdahl, B. J. 1977. The Viking gas exchange experiment results from Chryse and Utopia surface samples. *J. Geophys. Res.* 82:4669–4676.

Oyama, V. I., and Berdahl, B. J. 1979. A model of martian surface chemistry. *J. Molec. Evol.* 14:199–210.

Paige, D. 1992. The thermal stability of near-surface ground ice on Mars. *Nature* 356:43–45

Parker, J. F., Jr., and West, V. R., eds. 1973. *Bioastronautics Data Book*, NASA SP-3006.

Pollack, J. B., Colburn, D., Kahn, R., Hunter, J., Van Camp, W., Carlston, C. E., and Wolf, M. R. 1977. Properties of aerosols in the martian atmosphere, as inferred from Viking Lander imaging data. *J. Geophys. Res.* 82:4479–4496.

Ramohalli, K., and Sridhar, K. R. 1992. Space Processing of Indigenous Resources: Oxygen Production, An Overview. Paper presented at the AIAA 30th Aerospace Sciences Meeting, Jan., Reno, Nev. AIAA paper 92-0476.

Rogers, H. H., Thomas, J. F., and Bingham, G. E. 1983. Response of agronomic and forest species to elevated atmospheric carbon dioxide. *Science* 220:428–429.

Roncel, M., Navarro, J. A., and De la Rosa, M. A. 1989. Coupling of solar energy to hydrogen peroxide production in the cyanobacterium *Anacystis nidunlanas*. *Appl. Environ. Microbio.* 55:483–487.

Salisbury, F. B., and Ross, C. W. 1985. *Plant Physiology* (Belmont, Ca.: Wadsworth).

Schwartzkopf, S. H., and Mancinelli, R. L. 1991. Germination and growth of wheat in simulated Martian atmospheres. *Acta Astron.* 25:245–247.

Seckbach, J., Baker, F. A., and Shugarman, P. M. 1970. Algae thrive under pure CO_2. *Nature* 227:744–745.

Squyres, S. W., and Carr, M. H. 1986. Geomorphic evidence for the distribution of ground ice on Mars. *Science* 231:249–252.

Steele, D. B., and Stowers, M. D. 1991. Techniques for selection of industrially important microorganisms. *Ann. Rev. Microbiol.* 45:89–106.

Stoker, C. R., McKay, C. P., Haberle, R. M., and Andersen, D. T. 1992. Science strategy for humans exploration of Mars. *Adv. Space Res.* 12(4):79–90.

Toulmin, P., III, Clark, B. C., Baird, A. K., Keil, K., and Rose, H. J., Jr. 1976. Preliminary results from the Viking X-ray fluorescence experiment: The first sample from Chryse Planitia, Mars. *Science* 194:81–84.

Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christan, R. P., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretations of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.

Verniani, F. 1966. The total mass of the Earth's atmosphere. *J. Geophys. Res.* 71:385–391.

Zent, A. P., and Fanale, F. P. 1986. Possible Mars brines: Equilibrium and kinetic considerations. *Proc. Lunar Planet. Sci.* 16, *J. Geophys. Res. Suppl.* 91:D439–D445.

Zisk, S. H., and Mouginis-Mark, P. J. 1980. Anomalous region on Mars: Implications for near-surface liquid water. *Nature* 288:735–738.

ATMOSPHERIC EFFECTS ON THE UTILITY OF SOLAR POWER ON MARS

ROBERT M. HABERLE, CHRISTOPHER P. McKAY, and
JAMES B. POLLACK
NASA Ames Research Center

OWEN E. GWYNNE
SETI Institute

DAVID H. ATKINSON
University of Idaho

JOSEPH APPELBAUM
Tel Aviv University

GEOFFREY A. LANDIS
Sverdrup Technology

RICHARD W. ZUREK
Jet Propulsion Laboratory

and

DENNIS J. FLOOD
NASA Lewis Research Center

Solar power is likely to play an important role in the coming exploration of Mars. As a resource it is cheap, in great supply, easy to convert into useful work, and lacks dangerous failure modes that can jeopardize the safety of future human outposts. Yet Mars is farther from the Sun and it has an atmosphere. In particular, dust storms are a major concern because they can literally envelop the entire planet, and can last for several months. Engineering studies of solar-powered systems designed to operate on the surface of Mars must carefully consider these aspects of the Martian environment. This chapter summarizes what is known about the Martian atmosphere as it relates to solar radiation at the surface. Thus, the composition of the atmosphere is reviewed and the gases and aerosols that affect the solar beam are identified and characterized, including dust particles and water ice clouds. A methodology is presented that enables the calculation of solar radiation at the surface. It is shown that although the presence of dust reduces the available energy, there still remains an appreciable diffuse component that can be utilized by solar collectors. Water-ice clouds have the same effect, but

are of less concern than dust particles because of their generally lower optical depths, and greater scattering ability. It is estimated that the power requirements of a human outpost located at the Viking Lander 1 site (23°N) could be met with solar arrays about 30 by 30 m square, and about 1000 kg in mass. A system somewhat less than twice that size and mass could meet the power requirements for *in-situ* propellant production as envisioned in the "Mars Direct" scenario as proposed by Baker and Zubrin (1990). Thus, solar power is a viable energy source for future missions to Mars.

I. INTRODUCTION

Solar energy is likely to be an important power source for surface-based operations on Mars. As a resource it is economical, in great supply, and the photovoltaic cells needed to convert it to useful work offer many advantages over conventional systems. For example, they are robust, require minimal maintenance, gracefully degrade, and can sustain considerable damage before loosing their utility. These characteristics are particularly important for human outposts where safety is the major design requirement.

However, the design of solar-powered systems for Mars must consider the attenuating effects of its atmosphere. The Martian atmosphere, though thin in comparison with Earth's, does contain gases and aerosols that can attenuate the solar beam. Dust storms, for example, are known to produce visible optical depths of 5 or more—enough to completely block the Sun from view. However, because dust particles scatter as well as absorb solar radiation, there will be a diffuse component which can be significant (Appelbaum and Flood 1989*a*, *b*; Geels et al. 1989; Landis and Appelbaum 1990).

There are three goals for this chapter. The first is to review what is known about the Martian atmosphere as it relates to the deposition of solar energy at the surface. This includes a review of the abundance, distribution, properties, and variability of the principal optically active constituents: gases, dust particles, and water ice clouds. The second goal is to determine the effect of the atmosphere on the transmission of sunlight to the surface, and how that transmission varies with season, latitude, and time of day. The main concern is the effect of dust. Where possible, the spectral dependence of atmospheric transmissivity will also be discussed. Finally, the third goal is to assess the implications of the availability of sunlight at the Martian surface for the design of solar-powered surface systems. This will include a review of the state-of-the-art of solar cell technology, a power system analysis of a hypothetical human outpost, and a resource-processing application.

II. THE MARTIAN ATMOSPHERE

A. Gases

The principal gaseous constituents of the Martian atmosphere and their relative abundances by volume are listed in Table I. Of these, water vapor, ozone, and dust are variable. Together, these gases exert a mean pressure at the surface of

6.1 mbar which fluctuates seasonally by about 25% due to the condensation and sublimation of CO_2 in the polar regions. It is worth noting that ozone, which can absorb ultraviolet radiation, tends to be anti-correlated with water vapor. Water vapor in the lower atmosphere is photolyzed into H and OH which then participate in a catalytic cycle that destroys ozone (see, e.g., McElroy and Donahue 1972; Parkinson and Hunten 1972). Thus, regions of low water abundance can support large ozone concentrations. Ozone abundances as high as 60 μm atmospheres (1 μm atm = 2.69×10^{15} molecules cm^{-2}) have been observed near the edge of the north polar cap during winter (Barth 1985). Typical global-mean abundances, however, are less than several μm atm.

TABLE I[a]

Composition of the Martian Atmosphere	
CO_2	95.3%
N_2	2.7
Ar	1.6
O_2	0.13
H_2O	0.03[b]
CO	0.07
Ne	2.5 ppm
Kr	0.3 ppm
Xe	0.08 ppm
O_3	0.03 ppm[b]
Dust	10 ppm/m[b]

[a] Table from Owen et al. 1977.
[b] Variable.

B. Dust Particles

Background Levels. Dust particles may always be present in the Martian atmosphere. During the Viking mission, the solar optical depth was measured at the two Viking Lander sites for 1 1/3 Mars years (Fig. 1). At no time did the optical depth fall below 0.18. During the relatively clear periods, the optical depth was typically 0.2 to 0.5 . It is not clear how representative these values are because the Viking Lander sites are located in low-lying plains and may therefore see more dust in the overlying column. Furthermore, storm activity is suppressed in some years so that even lower values are possible. To be conservative, optical depths of 0.1 to 1.0 should be considered typical for those of the background dust haze.

The spatial distribution of the background haze is also uncertain. Modeling studies suggest that dust particles tend to become uniformly mixed with height (Haberle et al. 1982; Murphy et al. 1990) which is consistent with twilight observations at the Viking Lander sites (Kahn et al. 1981). Light profiles of the Martian limb indicate that a typical depth of the background

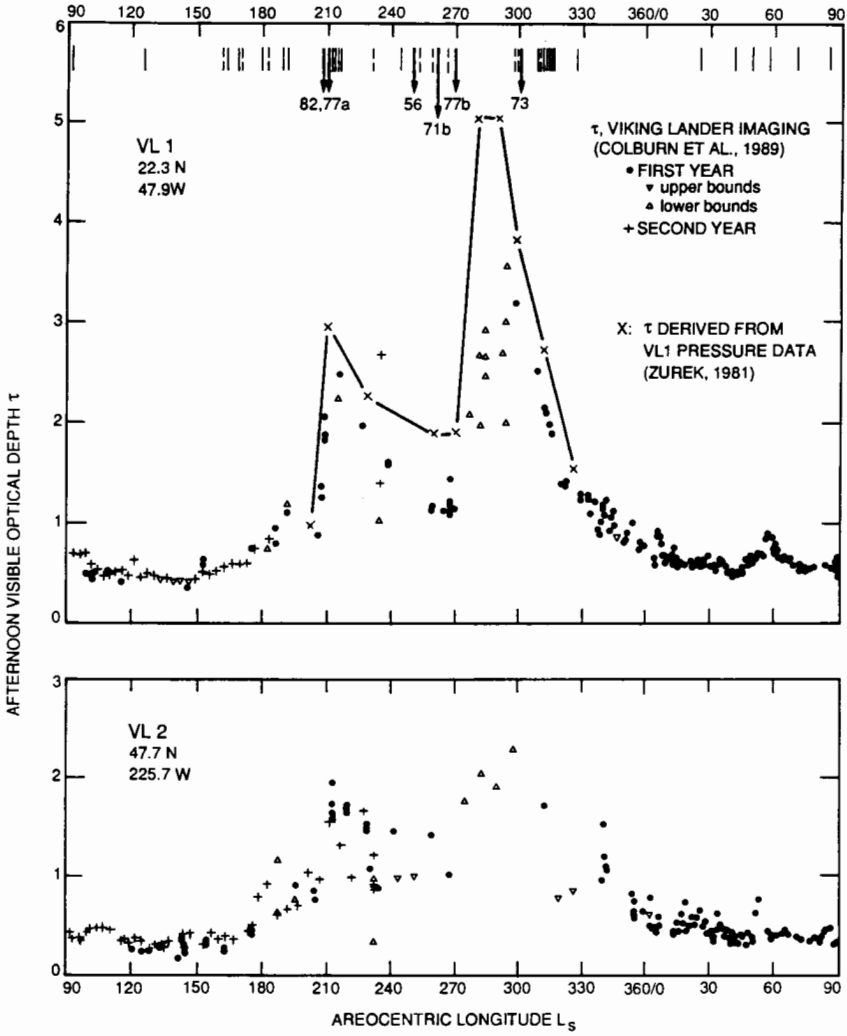


Figure 1. Visible optical depths derived from the Sun-diode measurements made at the two Viking Lander sites for more than 1 Mars year (from Colburn et al. 1989). The seasonal date on the abscissa covers one Mars year, as denoted by areocentric longitude, L_s. During the two planet-encircling dust storms in 1977, opacities were so large that only upper bounds could be estimated. Line indicates the opacity of a global haze inferred from Viking Lander 1 surface pressure data (Zurek 1981). Initial appearance or detection of all observed regional dust clouds, hazes or obscurations (vertical lines) and of planet-encircling (arrows) dust storms (Martin and Zurek 1993) are indicated at the top of the upper panel. Figure taken from Kahn et al. (1992).

haze is ~30 km, but that it can vary in latitude and season and may extend as high as 50 km (Jaquin et al. 1986).

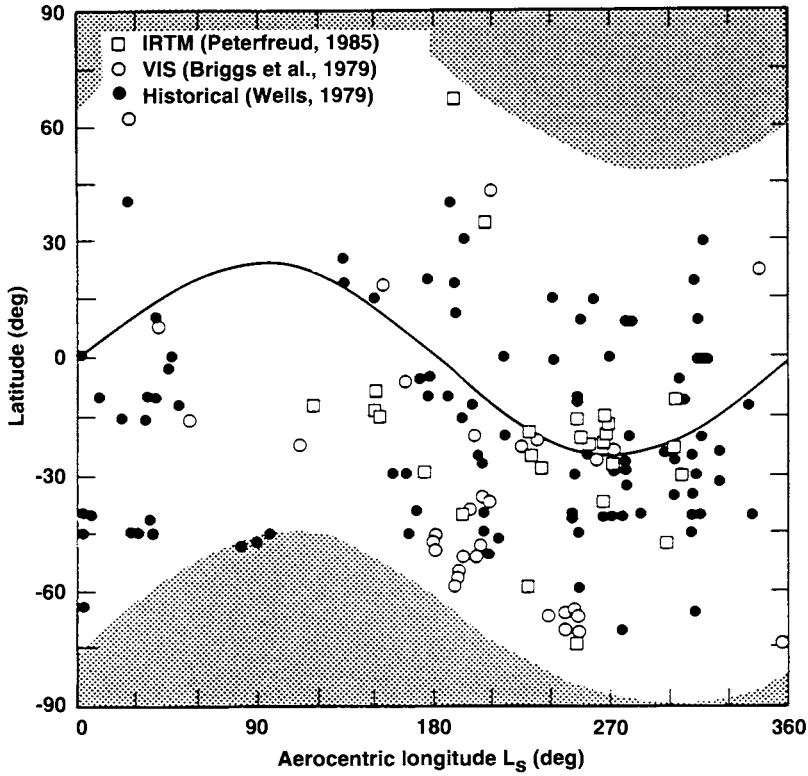


Figure 2. Latitude and time-of-year occurrence of local dust storms detected by the Viking Orbiters, from either infrared or visible imaging observations, and those observed from Earth, as compiled by Wells (1979). Taken from Peterfreund (1985).

Dust Storms. Dust particles are lifted into the Martian atmosphere by storms which range in size from local ($<10^5$ km^2) to global ($\sim 10^8$ km^2). Local storms are more frequent and may develop anywhere on the planet at any time of year. Our understanding of the frequency and distribution of local storms comes from the telescopic and Viking records which are limited in spatial and temporal coverage. The results of these observations are shown in Fig. 2. Although not pronounced, there is a tendency for these storms to cluster about the sub-solar point, and along the edge of the retreating south polar cap during spring. Based on the Viking results, approximately 100 local dust storms may occur each Martian year (Peterfreund 1985). Optical depths produced by these storms can be quite substantial ranging from 1 to 6 (Hunt 1979). Even higher values can be expected in localized regions. Local storms typically last a few days.

Occasionally, local storms can expand into regional (confined to a hemisphere and a region), or planet-encircling (covering both hemispheres) storms. Planet-encircling storms can be truly global in extent. Six such global dust storms have been observed (Table II). Of these, the 1971 storm was the longest, lasting for several months and completely obscuring the surface from view. In general the bigger the storm, the longer it lasts. Initial decay times for the two global dust storms observed by Viking in 1977 (Fig. 1) were 50 to 75 days (Pollack et al. 1977). A complete review of the characteristics of these great dust storms can be found in Zurek (1982).

TABLE II

Well-Documented Planet-Encircling Dust Storms

Year	L_s	Initial Location	Observation
1956	250°	30°S 31°E	Telescopic
1971	260°	29°S 38°E	Telescopic/Mariner 9
1973	300°	24°S 88°W	Telescopic
1977	204°	30°S 70°W	Viking
1977	268°	?	Viking
1982	200°	?	Viking

Optical depths during planet-encircling storms are likely to be highly variable in space and time. At the Viking Lander sites, only lower limits on the peak opacity could be estimated because the Sun was obscured by dust. Theoretical estimates have been made using the amplitude of daily pressure variations as a proxy for opacity (Zurek 1981). Such variations are sensitive to atmospheric heating and, hence, the global dust loading. Values as high as 6 have been determined using this method.

Unlike local storms, planet-encircling storms tend to develop from specific locations at specific seasons: the southern subtropics, during spring and summer (fall and winter in the north). The Hellas and Argyre basins, along with Noachis and Solis Planum, are the most active regions. At this season and latitude, the available solar radiation is near its maximum yearly value and the strength of the general circulation is high (Pollack et al. 1990). However, planet-encircling storms do not occur each Martian year (Martin and Zurek 1993; Zurek and Martin 1993). During the Viking mission, they were observed during the first and fourth year of observations, but not during the second and third years. Since then, no dust storms greater than regional in size have been observed. Prior to the era of spacecraft exploration, only one planet-encircling storm—the 1956 storm—was observed. From observations at all favorable oppositions since 1905, one conclues that planet-encircling storms may be episodic. However, the possibility that such storms went unobserved at other times cannot be ruled out. A summary of the occurrence of major dust storms since 1905 is given in Fig. 3.

Particle Properties. Because no *in-situ* measurements of dust particles in

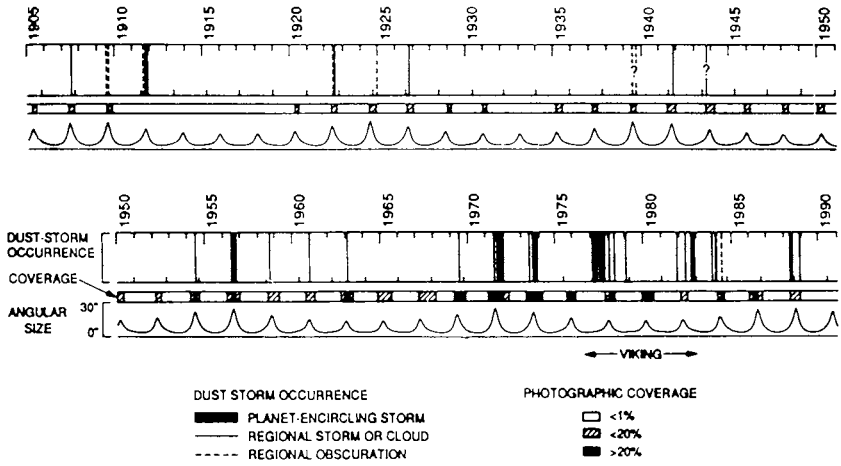


Figure 3. Time line of observed major dust storms on Mars (Zurek and Martin 1993).

the Martian atmosphere have been made, their composition, size distribution, and shape are not well constrained. Instead, inferences about these characteristics are based on remote sensing measurements which cannot uniquely separate the various contributions to the observed opacity. Synthesized thermal emission spectra of the 1971 global dust storm compare favorably with the observed spectra for particles dominated by igneous silicates, or weathering products like clays. Some basalt may also be present. Relatively high SiO_2 contents (>60%) have been inferred from these data (Hanel et al. 1972)—in contrast to the lower silicate abundance (~40%) inferred from soil measurements at the Viking Lander sites (Toulmin et al. 1977; Clark et al. 1982). The density of silicate materials on Earth is typically 2 to 3 g cm^{-3} and is probably a good estimate for Martian dust densities as well.

The synthesized spectra also imply a particle size distribution having a mode radius of 0.4 μm, and a cross-section weighted-mean radius of 2.5 μm (Toon et al. 1977). Using this size distribution, a density of 3 g cm^{-3}, and assuming the dust particles are uniformly mixed with a scale height $H = 10$ km, the concentration of dust particles as a function of height z can be approximately obtained from

$$N \sim N_0 \tau \exp(-z/H) \tag{1}$$

where $N_0 = 6\,cm^{-3}$ is the number density at the surface when the optical depth $\tau = 1$. Thus, during a major dust storm ($\tau = 5$) there are approximately 30 particles cm^{-3} at the surface. It should be noted, however, that this estimate is sensitive to the assumed particle size distribution. Other distributions with cross-sectional mean radii ranging form 0.4 μm (Pollack et al. 1977; Clancy and Lee 1991) to 10 μm (Leovy et al. 1972) have been estimated. These distributions imply that N_0 varies between 38 and 1.5 particles cm^{-3}.

Thus, there is significant uncertainty in actual concentrations. In addition, concentrations can be much higher in regions of active dust raising. In this sense, Eq. (1) can be regarded as a lower limit.

The dust loading of the atmosphere can also be estimated though it is subject to the same uncertainties as the concentration. For a 2.5 μm mean particle radius, Pollack et al. (1979) show that the mass loading (g cm^{-2}) is related to the optical depth by

$$m = 5.0 \times 10^{-4} \tau \tag{2}$$

Thus, a column with unit optical depth contains about 5×10^{-4} g cm^{-2} of dust—enough to form a layer $\sim$2 μm thick. During a global dust storm there is an equivalent of $\sim$10 μm of dust in the atmosphere. Thus, the amount of dust that settles out of the atmosphere following a single storm may be quite small. This is consistent with changes in the surface brightness seen by the Viking Landers (Arvidson et al. 1983). However, it may have a significant effect on the performance of solar cells (see Sec. V).

TABLE III
Dust Particle Single Scattering Properties

Wavelength, μm	ω_o	g	Q_{ext}
0.35	0.598	0.736	3.052
0.40	0.627	0.682	3.177
0.45	0.659	0.634	3.294
0.50	0.695	0.592	3.40
0.556	0.735	0.556	3.48
0.669	0.840	0.497	3.58
0.80	0.901	0.492	3.30
1.0	0.739	0.585	2.50
Solar Average[a]	0.751	0.576	3.06

[a] Weighted by the solar spectrum and the products integrated over wavelength.

The shape of Martian dust particles is also uncertain. Sky brightness measurements from the Viking Lander sites suggest the particles are nonspherical but equidimensional (Pollack et al. 1977). In addition, disk-shaped particles, because they settle out of the atmosphere more slowly are better able to reproduce several observed characteristics of dust particles in the Martian atmosphere (Murphy et al. 1993). These characteristics include the relatively small temporal variation in the particle size distribution (Toon et al. 1977), and the small temporal and spatial variation of the visible-to-infrared opacity (Martin 1986). However, the latter can also be reproduced with spherical particles, though not uniquely (Clancy and Lee 1991). In the absence of more definitive measurements, the particle shape remains uncertain.

Radiative Properties. The size and composition of Martian dust particles permits them to interact efficiently with solar radiation. That interaction can be quantified given knowledge of the particles' single scattering properties and optical depth. The latter has been directly measured by the Viking Landers (Colburn et al. 1989); the former have been estimated from the angular variation of sky brightness (Pollack et al. 1979). The single scattering properties of interest are the single-scattering albedo ω_o, the particle asymmetry factor g, and the extinction efficiency Q_{ext}. The values of these parameters for each of the six different filters on the Lander cameras, as well as their solar average value, are listed in Table III.

B. Water-Ice Clouds

Occurrence and Distribution. Water vapor in the Martian atmosphere is controlled by saturation so that clouds are fairly common. They have been observed as wave clouds, convective clouds, ground fogs, high altitude hazes, and the so called "polar hood" (Leovy et al. 1972,1973; Briggs and Leovy 1974; French et al. 1981; Kahn 1984). However, the water content of Martian clouds is much less than their terrestrial counterparts because of the colder thermal environment. Consequently, optically thick clouds on Mars are less widespread than they are on Earth.

In general, Martian clouds appear to form more frequently in the northern hemisphere than in the southern hemisphere, and during fall and winter rather than spring and summer. This is especially true of the polar hoods which appear as deep ($\sim$10 km) polar-encircling fogs. This asymmetry is related to the fact that the north residual cap is water ice (and therefore acts as a source for atmospheric water) while the south residual cap is CO_2 ice (and therefore acts as a sink for atmospheric water). In addition, clouds form more easily at high latitudes than at low latitudes because saturation is more readily achieved. The exception is the upland regions of Tharsis during northern summer when frequent convective clouds resembling fields of large diffuse cumulus have been observed. Also, ground fogs have been observed in tropical craters and canyons but because they sublimate well before noon, they are not an issue for solar power. Some high-altitude tropical hazes may also be diurnally varying (Colburn et al. 1989)

Physical and Radiative Properties. Very little is known about the physical properties of Martian clouds. Spectral measurements of a single cloud indicate particles in the μm-size range (Curran et al. 1973). On theoretical grounds, a variety of particle sizes are expected depending on altitude, dust loading, humidity, etc. (Kahn 1990). Higher clouds generally contain smaller particles and are less opaque than lower clouds. Estimates of the mean particle size (radius) range from sub-μm for the high clouds, to several μm for the low clouds. The corresponding visible opacities range from near 0 to about 0.5, though higher values are certainly possible.

Without knowledge of the particle shape and size distribution the single scattering properties of Martian clouds cannot be determined. However, it

is likely that they are fairly conservative ($\omega_o \sim 1$) and moderately forward scattering ($g \sim 0.6$). In estimating the transmission properties of Martian clouds, we have adopted these values as a first approximation.

III. ATMOSPHERIC EFFECTS ON SOLAR TRANSMISSION

A. Solar Radiation at Mars

The first step in the assessment of solar power on Mars is to determine the maximum amount of sunlight S available at a given location, season, and time of day. S is therefore the irradiance of solar radiation incident on the top of the atmosphere. It can be calculated from

$$S = \mu S_0 \left(\frac{\bar{r}}{r}\right)^2 \tag{3}$$

where $\mu = \cos z$ is the cosine of the solar zenith angle z,

$$\cos z = \sin\theta \sin\delta + \cos\theta \cos\delta \cos h. \tag{4}$$

$S_0 = 590\ \mathrm{W\ m^{-2}}$ is the solar irradiance at Mars' mean distance from the Sun $\bar{r}$ ($= 1.52$ AU), θ and δ are the latitude and solar declination, $h = 2\pi t/P$ is the hour angle, t is time measured from local noon, and $P = 88775$ s is the length of the Martian solar day. The solar declination angle depends on Mars' obliquity ϵ (25°.2) and orbital position L_s. The latter is an angular measure: $L_s = 0°$ corresponds to northern vernal equinox; $L_s = 90°$ corresponds to northern summer solstice, and so on. The solar declination angle is related to ϵ and L_s by

$$\sin\delta = \sin\epsilon \sin L_s. \tag{5}$$

Finally, the Sun-Mars distance r can be found from

$$\left(\frac{\bar{r}}{r}\right) = \frac{1 + e\cos(L_s - L_s^p)}{1 - e^2} \tag{6}$$

where $e = 0.0934$ is the orbit eccentricity, and $L_s^p = 250°$ is the areocentric longitude at perihelion. It is worth noting that the high eccentricity of Mars' orbit results in a significant seasonal variation in the available solar irradiance. At perihelion the maximum available irradiance is $S = 717\ \mathrm{W\ m^{-2}}$, while at aphelion the maximum is $S = 493\ \mathrm{W\ m^{-2}}$. This is a much greater variation than experienced on Earth.

A useful quantity is the daily averaged insolation $\bar{S}$. Integrating Eq. (3) from Sunrise to Sunset gives

$$\bar{S} = \frac{S_0}{\pi}\left(\frac{\bar{r}}{r}\right)^2 [\cos\delta \cos\theta \sin H + H \sin\delta \sin\theta] \tag{7}$$

where H is the half-day length (radians) which can be found from

$$\cos H = \begin{cases} -1, & \theta > \pi/2 - \delta \text{ or } \theta < -\pi/2 - \delta \\ -\tan\theta \tan\delta, & |\theta| < \pi/2 - |\delta| \\ +1, & \theta > \pi/2 + \delta \text{ or } \theta < -\pi/2 + \delta \end{cases} \quad (8)$$

Note that $\cos H = -1$ when the Sun never sets, and $\cos H = +1$ when it never rises.

The daily averaged insolation as a function of latitude and season is shown in Fig. 4. Maximum insolation occurs at the poles where the increased day length more than compensates for the low zenith angles. The south pole has a higher seasonal maximum than the north pole because southern summer occurs near perihelion. Of course solar-powered systems would not be able to provide continuous power all year long at these latitudes because of the long polar nights. At the equator, however, seasonal variations are small and continuous power is possible all year long.

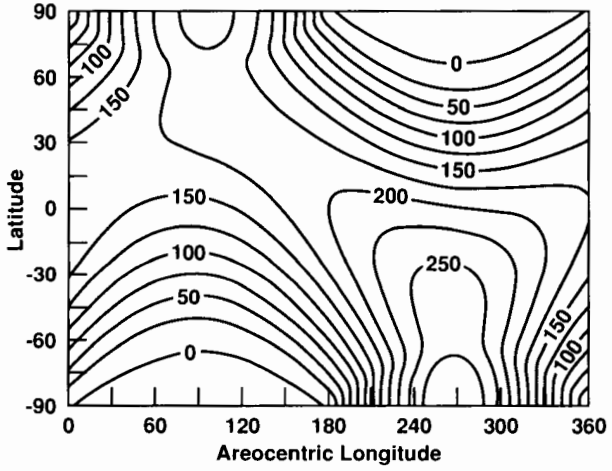


Figure 4. Daily averaged solar insolation at the top of the atmosphere. Contour intervals are 25 W m^{-2}.

The annually averaged insolation as a function of latitude is given by

$$\langle \overline{S} \rangle = \frac{S_0}{2\pi^2}(1 - e^2)^{\frac{1}{2}} \int_0^{2\pi} [1 - (\sin\theta \cos\epsilon - \cos\theta \sin\epsilon \sin\lambda)^2]^{\frac{1}{2}} d\lambda \quad (9)$$

where the integration is over all longitudes λ. $\langle \overline{S} \rangle$ is a maximum at the equator and is hemispherically symmetric (Fig. 5). However, due to the eccentricity

of Mars' orbit, the maximum and minimum of $\overline{S}$ are not symmetric; the minimum is sharply peaked in latitude, whereas the maximum is not. The highest minimum occurs at 15°N, while the lowest maximum occurs over a broad range of latitudes between 30 to 60°N. Thus, the low latitudes of the northern hemisphere experience the least seasonal variation, while the high latitudes of the southern hemisphere experience the greatest seasonal variation.

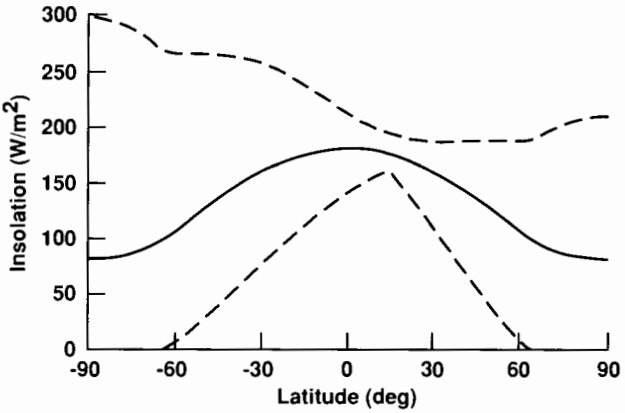


Figure 5. Mars annual insolation. At each latitude, the daily averaged insolation varies during the course of the year. Maximum (top dashed), minimum (bottom dashed), and mean (middle solid) daily averaged insolation are illustrated in this figure.

B. Atmospheric Extinction

Gases. Kuhn and Atreya (1979) have discussed the molecular extinction of solar radiation on Mars. Of the gases listed in Table I, CO_2, H_2O, O_2, and O_3 are of most interest. Except in the far ultraviolet, scattering by these gases contributes very little to the total extinction which is dominated instead by absorption. The effectiveness of these gases as absorbers is given by the product of their column density and absorption cross section. Their approximate time and globally averaged column densities (molecules cm^{-2}) are $[CO_2] = 2 \times 10^{23}$, $[H_2O] = 4 \times 10^{19}$, $[O_2] = 10^{20}$, and $[O_3] = 5 \times 10^{15}$. From 200 to 1500 nm the absorption cross section for CO_2 is negligible (Banks and Kockarts 1973) and radiation at these wavelengths undergoes very little absorption by this gas. Water vapor on the other hand, has many absorption features in this region, but is too scarce ($<10^{-3}$ of that found on Earth) to affect much absorption. This is also true for O_2 and O_3. Consequently, there is very little molecular absorption of radiation between 200 and 1500 nm in the Martian atmosphere. Shortward of 200 nm, however, the CO_2 cross sections become large enough ($>>5 \times 10^{-24}$ cm^2) to yield significant opacity. Water vapor and O_2 also begin to absorb shortward of 200 nm, but their contribution

is small compared to the more abundant CO_2. In any case, virtually no radiation at wavelengths <200 nm reaches the Martian surface.

That part of the ultraviolet irradiance that does reach the surface (200–350 nm) constitutes about 4.5% of the total solar energy. Therefore, the distribution of incident ultraviolet radiation at the surface can be found by multiplying 0.045 times the value shown in Fig. 5. Exceptions to this occur when the ozone column density departs significantly from the time and globally averaged value quoted above. Such departures occur mainly near the edge of the winter polar cap where column densities as high as 150×10^{15} molecules cm^{-2} have been observed (Barth 1985). As the absorption cross section for O_3 is 10^{-17} cm^2 at 260 nm (Daumont et al. 1983), such concentrations can produce a measurable opacity. However, the solar zenith angle is so large in this situation that very little solar energy is available anyway (at least for horizontal collectors). In the summer hemisphere and at midlatitudes, O_3 concentrations are typically $<5 \times 10^{15}$ molecules cm^{-2} and very little attenuation of ultraviolet radiation occurs.

In general, therefore, a clear Mars atmosphere is transparent to solar radiation at wavelengths >200 nm, and opaque to solar radiation at wavelengths <200 nm. If the atmosphere is dusty, however, additional near-ultraviolet opacity could be provided. Indeed, if the ultraviolet opacity is comparable to the visible opacity (as would be theoretically expected because the particles are much larger than ultraviolet wavelengths), then dust could be the greatest source of opacity in this region of the spectrum. Ultraviolet reflectance measurements are consistent with this interpretation (Hord et al. 1972). However, in the absence of more definitive measurements, damage due to radiation in the 200 to 350 nm region should be considered an issue for solar collectors and other equipment exposed on the Martian surface.

Dust. The main concern for solar-powered systems is the effect of suspended dust particles on the amount of sunlight reaching the surface. These effects can be calculated given the particles' single scattering properties discussed earlier. Pollack et al. (1990) have used these properties in a doubling/adding code to calculate the net irradiance as a function of altitude, total optical depth, surface albedo and zenith angle. The results were stored in a four-dimensional look-up table which was intended to speed up heating rate calculations in a general circulation model. For solar power studies, however, it is the total downward irradiance—not the net irradiance—that is of interest. The total downward irradiance, F_{tot}, can be obtained from the Pollack et al. data using the following formula

$$F_{tot} = \frac{Sf(\tau, \mu, A)}{(1 - A)} \tag{10}$$

where A is the surface albedo, and $f(\tau, \mu, A)$ is a normalized net irradiance function whose value as a function of optical depth and zenith angle is given in Tables IV and V. Previous studies of solar power on Mars have utilized

TABLE IV

Normalized Net Irradiance Function for a Surface Albedo = 0.1

Optical Depth, τ	Solar Zenith Angle, Degrees									
	0	10	20	30	40	50	60	70	80	85
0.100	0.883	0.883	0.881	0.879	0.875	0.868	0.855	0.830	0.757	0.640
0.200	0.866	0.865	0.862	0.857	0.850	0.836	0.813	0.768	0.651	0.508
0.300	0.848	0.847	0.842	0.835	0.826	0.806	0.773	0.712	0.571	0.433
0.400	0.830	0.829	0.823	0.814	0.801	0.776	0.736	0.663	0.510	0.385
0.500	0.813	0.811	0.804	0.793	0.778	0.748	0.701	0.619	0.462	0.351
0.600	0.796	0.793	0.785	0.772	0.755	0.721	0.668	0.579	0.422	0.325
0.700	0.778	0.775	0.766	0.752	0.732	0.695	0.638	0.546	0.393	0.306
0.800	0.761	0.758	0.748	0.732	0.711	0.670	0.609	0.514	0.366	0.288
0.900	0.744	0.741	0.730	0.713	0.690	0.646	0.583	0.486	0.345	0.274
1.000	0.728	0.724	0.712	0.694	0.669	0.623	0.557	0.459	0.324	0.261
1.100	0.712	0.708	0.696	0.676	0.649	0.602	0.535	0.438	0.309	0.250
1.200	0.695	0.691	0.679	0.658	0.630	0.581	0.513	0.416	0.295	0.240
1.300	0.679	0.675	0.662	0.640	0.611	0.560	0.491	0.395	0.280	0.230

1.400	0.664	0.660	0.646	0.624	0.594	0.542	0.473	0.379	0.270	0.222
1.500	0.649	0.644	0.630	0.607	0.576	0.524	0.455	0.364	0.260	0.215
1.600	0.634	0.629	0.615	0.591	0.559	0.506	0.437	0.348	0.249	0.207
1.700	0.619	0.614	0.599	0.575	0.542	0.489	0.420	0.333	0.240	0.200
1.800	0.605	0.600	0.585	0.561	0.527	0.475	0.460	0.321	0.232	0.194
1.900	0.591	0.586	0.571	0.546	0.513	0.460	0.392	0.310	0.225	0.188
2.000	0.578	0.572	0.557	0.532	0.498	0.445	0.377	0.298	0.217	0.182
2.250	0.544	0.539	0.523	0.498	0.463	0.411	0.346	0.272	0.200	0.168
2.500	0.514	0.509	0.493	0.467	0.432	0.381	0.319	0.252	0.187	0.157
2.750	0.483	0.478	0.462	0.436	0.401	0.352	0.293	0.231	0.173	0.146
3.000	0.457	0.452	0.436	0.411	0.376	0.329	0.273	0.216	0.162	0.137
3.250	0.431	0.426	0.411	0.386	0.352	0.307	0.254	0.202	0.152	0.129
3.500	0.405	0.400	0.385	0.361	0.328	0.285	0.236	0.187	0.142	0.120
4.000	0.363	0.358	0.344	0.321	0.290	0.251	0.208	0.166	0.127	0.107
5.000	0.289	0.285	0.273	0.253	0.227	0.196	0.162	0.131	0.101	0.086
6.000	0.229	0.226	0.216	0.200	0.178	0.154	0.128	0.104	0.081	0.069

this formula (Appelbaum and Flood 1989*a*, *b*; Geels et al. 1989; Landis and Appelbaum 1990; Appelbaum and Flood 1990). Table IV is for an albedo of 0.1 and Table V is for an albedo of 0.4. A linear interpolation can extend the results to surfaces of arbitrary albedo.

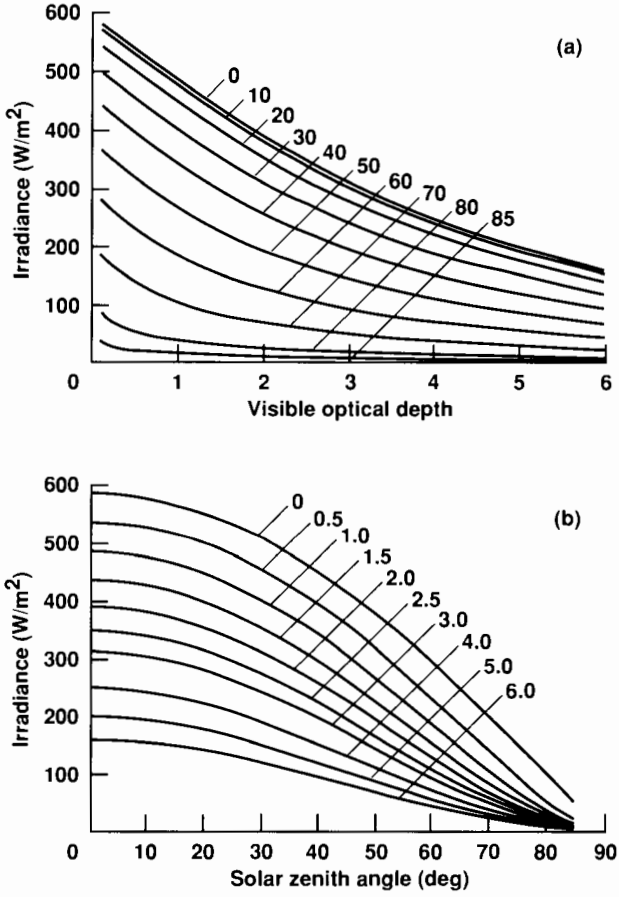


Figure 6. Top: Total irradiance (direct + diffuse) as a function of optical depth for 10 different zenith angles. The zenith angles correspond to those given in Tables IV and V. Bottom: Total irradiance as a function of zenith angle for 10 different optical depths. All irradiances are computed using the mean irradiance of 590 W m^{-2} at the top of the atmosphere.

Values of F_{tot} as a function of optical depth and zenith angle for $A = 0.25$—typical for Mars—are shown in Fig. 6. It can be seen from this figure that substantial energy reaches the surface even at high optical depths. This is contrary to suggestions that dust storms would render solar-powered systems ineffective. At normal incidence, for example, almost 200 W m^{-2} are still

available at $\tau = 5$; even with the Sun only 30° above the horizon, 50 W m^{-2} are available. The reason for this is that Beer's law, which would reduce the energy by $e^{-5} = 0.007$ in this case, is not appropriate for a scattering atmosphere. Consequently, there is significant diffuse component of the radiation field that partially compensates for the reduction of the direct beam.

To illustrate this more quantitatively, the total downward irradiance can be separated into its direct, F_{dir}, and diffuse, F_{diff}, components. To make this separation using Pollack et al.'s data, subtract the direct irradiance

$$F_{\text{dir}} = Se^{-\frac{\tau}{\mu}} \tag{11}$$

from the total irradiance to obtain the diffuse irradiance, i.e,

$$F_{\text{diff}} = S\left[\frac{f(\tau, \mu, A)}{(1-A)} - e^{-\frac{\tau}{\mu}}\right]. \tag{12}$$

Note that F_{dir} and F_{diff} apply to a horizontal surface. As illustrated in Figs. 7 and 8, the direct component declines sharply and monotonically with increasing optical depth, while the diffuse component increases sharply, then falls off slowly with increasing optical depth. For both components, the incidence angle dependence is modest at normal incidence, and strong at grazing incidence. Thus, at large optical depths much of the direct beam is removed and converted into scattered radiation. It is this property of Martian dust that greatly enhances the viability of solar power on Mars. An example of how the direct, diffuse, and total radiation fields vary diurnally at the Viking Lander 1 site using the data of Pollack et al. is shown in Fig. 9. Again, the diffuse component dominates at high opacity, and is nontrivial at low opacity.

The total daily insolation on a horizontal surface and its direct and diffuse components can be calculated by integrating Eqs. (10)–(12) over the period from Sunrise to Sunset. The results are

$$\overline{F}_{\text{tot}} = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^H \mu \frac{f(\tau, \mu, A)}{(1-A)} \mathrm{d}h \tag{13}$$

$$\overline{F}_{\text{dir}} = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^H \mu e \frac{-\tau}{\mu} \mathrm{d}h \tag{14}$$

$$\overline{F}_{\text{diff}} = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^H \mu \left[\frac{f(\tau, \mu, A)}{(1-A)} - e^{\frac{-\tau}{\mu}}\right] \mathrm{d}h. \tag{15}$$

Using the optical depth data from Fig. 1, Eqs. (13)–(15) can be used to determine the seasonal variation of the total daily irradiance at the Viking Lander sites. The results are shown in Fig. 10 along with the percentage of the total irradiance contributed by the diffuse component. During the spring and summer when the optical depth is relatively low, the diffuse component

TABLE V

Normalized Net Irradiance Function for a Surface Albedo = 0.4

Optical Depth, τ	Solar Zenith Angle, Degrees									
	0	10	20	30	40	50	60	70	80	85
0.100	0.594	0.593	0.592	0.591	0.588	0.583	0.575	0.558	0.509	0.430
0.200	0.586	0.585	0.583	0.580	0.576	0.566	0.551	0.520	0.441	0.344
0.300	0.578	0.577	0.574	0.569	0.562	0.549	0.527	0.485	0.389	0.295
0.400	0.569	0.567	0.564	0.557	0.549	0.532	0.504	0.454	0.350	0.264
0.500	0.560	0.558	0.553	0.546	0.535	0.515	0.483	0.426	0.318	0.242
0.600	0.550	0.548	0.543	0.534	0.522	0.498	0.462	0.401	0.292	0.225
0.700	0.540	0.538	0.532	0.522	0.509	0.483	0.443	0.379	0.273	0.213
0.800	0.531	0.528	0.522	0.511	0.496	0.467	0.425	0.358	0.256	0.202
0.900	0.521	0.518	0.511	0.499	0.483	0.452	0.408	0.341	0.242	0.193
1.000	0.511	0.509	0.501	0.488	0.470	0.438	0.392	0.323	0.228	0.184
1.100	0.501	0.499	0.490	0.476	0.458	0.424	0.377	0.309	0.219	0.177
1.200	0.492	0.489	0.480	0.465	0.446	0.411	0.363	0.295	0.209	0.171
1.300	0.482	0.479	0.469	0.454	0.433	0.398	0.349	0.281	0.200	0.164

1.400	0.472	0.469	0.459	0.444	0.422	0.386	0.337	0.271	0.193	0.159
1.500	0.463	0.460	0.450	0.433	0.411	0.374	0.325	0.260	0.186	0.154
1.600	0.453	0.450	0.440	0.423	0.400	0.363	0.313	0.250	0.180	0.149
1.700	0.444	0.440	0.430	0.413	0.389	0.352	0.302	0.240	0.173	0.144
1.800	0.435	0.431	0.421	0.403	0.379	0.342	0.292	0.232	0.168	0.140
1.900	0.426	0.422	0.411	0.394	0.370	0.332	0.283	0.224	0.163	0.136
2.000	0.417	0.413	0.402	0.384	0.360	0.322	0.273	0.217	0.158	0.133
2.250	0.395	0.391	0.380	0.362	0.336	0.299	0.252	0.199	0.147	0.124
2.500	0.375	0.371	0.360	0.341	0.316	0.279	0.234	0.185	0.138	0.116
2.750	0.354	0.350	0.339	0.320	0.295	0.259	0.216	0.171	0.128	0.108
3.000	0.336	0.332	0.321	0.303	0.277	0.243	0.202	0.161	0.121	0.102
3.250	0.319	0.315	0.304	0.285	0.261	0.288	0.189	0.151	0.114	0.096
3.500	0.301	0.297	0.286	0.268	0.244	0.212	0.176	0.141	0.107	0.091
4.000	0.271	0.267	0.257	0.240	0.217	0.189	0.157	0.125	0.096	0.082
5.000	0.219	0.216	0.207	0.192	0.173	0.149	0.124	0.101	0.078	0.066
6.000	0.176	0.173	0.166	0.154	0.137	0.119	0.099	0.081	0.063	0.054

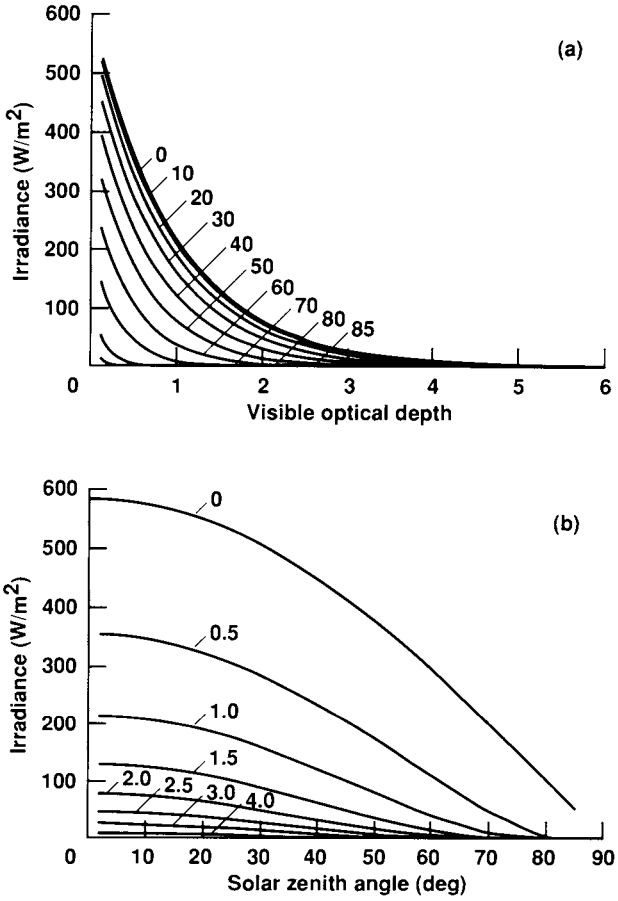


Figure 7. Same as in Fig. 6, but for the direct irradiance.

contributes almost half of the total irradiance at both sites; it rises to almost 90% during fall and winter when the optical depth is relatively high.

The advantage of using the data of Pollack et al. is that they are based on an accurate radiative transfer code. The disadvantage is that the dust optical properties are fixed. As discussed above, there are uncertainties in these properties and so it would be useful to know how well simpler transmission models compare with Pollack et al.'s in order to gain some flexibility in the calculations. For this purpose, two additional models have been chosen: a 2-stream model based on the Delta-Eddington approximation (Joseph et al. 1976), and a purely scattering model in which no absorption takes place at all ($\omega_o = 0$). As will be shown, the latter has a very simple analytical expression that is useful for quick estimates; it is also a much better representation of transmission than pure absorption (Beer's law).

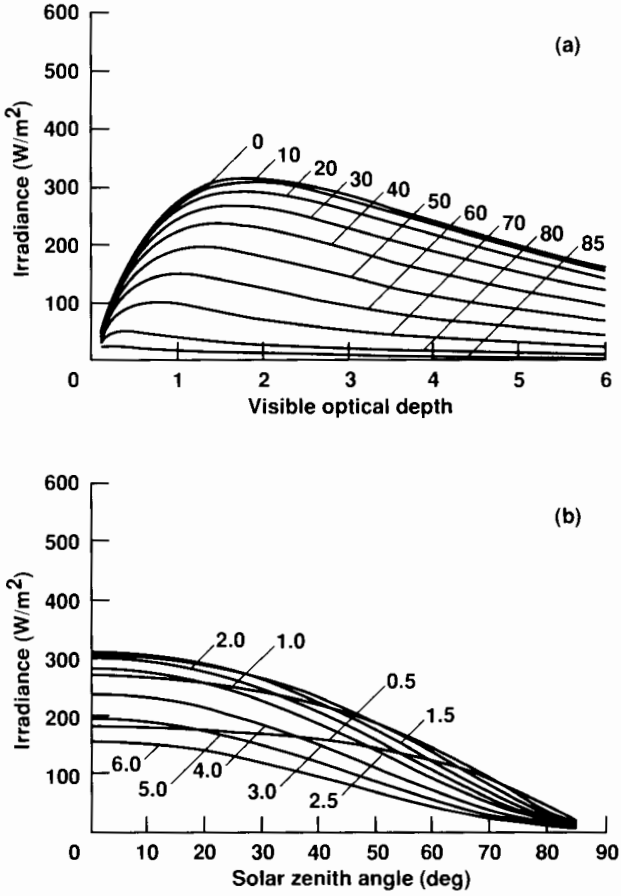


Figure 8. Same as in Fig. 6, but for the diffuse irradiance.

For a semi-infinite atmosphere overlying a reflecting (Lambertian) surface, the transmissivity (normalized by the μ-weighted incident irradiance) of solar radiation given by the Delta-Eddington approximation is

$$T(\mu, \tau, A) = c_1 e^{-k\tau}(1+P) + c_2 e^{+k\tau}(1-P) - (\alpha+\beta-1)e^{-\frac{\tau}{\mu}} \quad (16)$$

where the constants c_1 and c_2 are

$$c_1 = -\frac{(1-P)e^{-\frac{\tau}{\mu}}B - (\alpha+\beta)e^{+k\tau}C}{(1+P)e^{+k\tau}C - (1-P)e^{-k\tau}D} \quad (17)$$

and

$$c_2 = +\frac{(1+P)e^{-\frac{\tau}{\mu}}B - (\alpha+\beta)e^{-k\tau}D}{(1+P)e^{+k\tau}C - (1-P)e^{-k\tau}D} \quad (18)$$

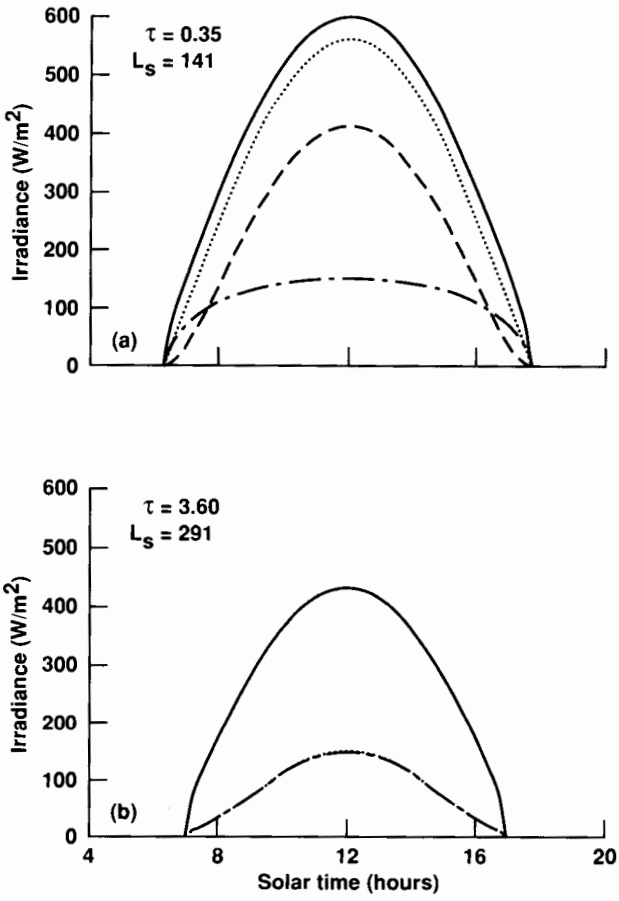


Figure 9. Diurnal variation of solar radiation at the Viking Lander 1 site for a midsummer season when the dust loading is minimal (top), and for an early winter season when it is large (bottom). The solid line represents the energy available to a horizontal collector at the top of the atmosphere; the dotted line represents the energy available at the surface and is therefore a measure of the extinction by dust. The dashed and solid-dotted curves show the direct and diffuse components, respectively. Virtually all of the radiation arriving at the surface in the winter case comes from the diffuse component.

where $B = A + (1 - A)\alpha - (1 + A)\beta$, $C = (1 - A) + P(1 + A)$ and $D = (1 - A) - P(1 + A)$.

$$k = [3(1 - \omega_0)(1 - g\omega_0)]^{\frac{1}{2}} \tag{19}$$

$$P = \frac{2}{3}\left[\frac{3(1 - \omega_0)}{(1 - g\omega_0)}\right]^{\frac{1}{2}} \tag{20}$$

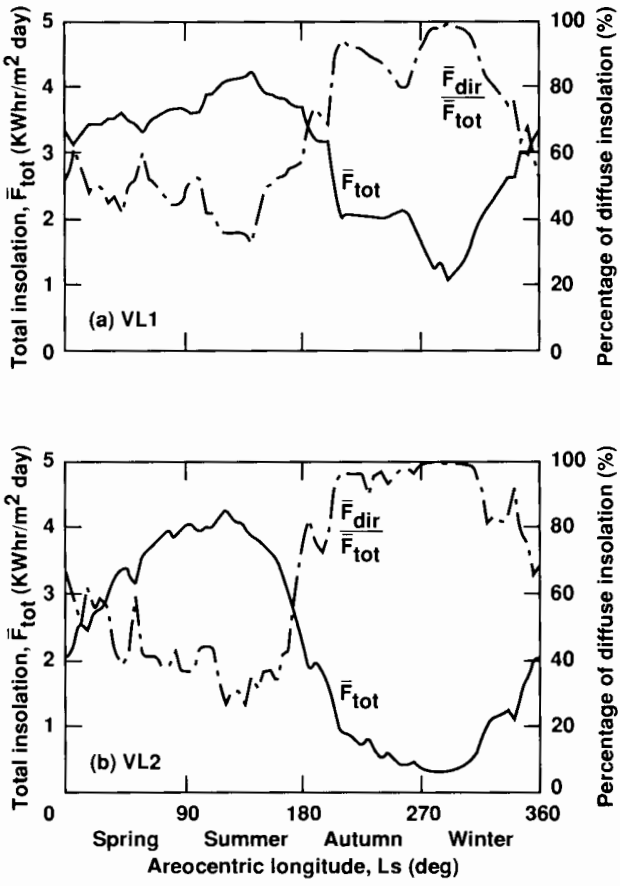


Figure 10. Seasonal variation of insolation at the two Viking Lander sites. Solid lines represent the daily-averaged total irradiance, dashed lines represent the percentage of that radiation due to the diffuse component.

$$\alpha = \frac{3}{4}\mu\omega_0\left[\frac{1+g(1-\omega_0)}{1-(\mu k)^2}\right] \tag{21}$$

$$\beta = \frac{1}{2}\mu\omega_0\left[\frac{\left(\frac{1}{\mu}\right)+3g\mu(1-\omega_0)}{1-(\mu k)^2}\right]. \tag{22}$$

A comparison of the Delta-Eddington and doubling code transmissivities is shown in Fig. 11. In general, the Delta-Eddington solution agrees well with the doubling code, but overestimates the transmissivity at low optical depths and underestimates it at high optical depths. These differences are mostly due to the multispectral nature of the doubling code calculations, as well as its accounting for the non-spherical shape of the dust particles. Also shown are

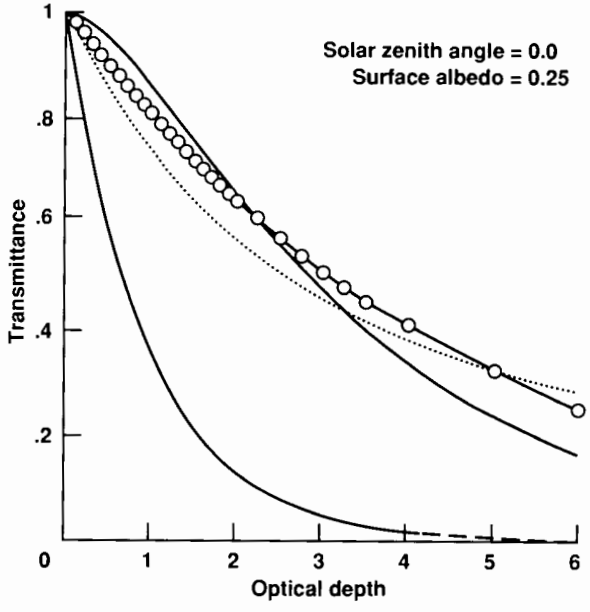


Figure 11. Dust transmissivities from several different radiation models. The solid curve with circles corresponds to the model of Pollack et al. (1990). The solid curve without circles is based on the Delta-Eddington 2-stream solution (Eq. 16). The dotted curve is based on a pure scattering model (Eq. 23), while the lower curve corresponds to Beer's law ($T = e^{-\tau/\mu}$).

the transmissivities for pure isotropic scatterers ($\omega_o = 1$) and pure absorbers ($\omega_o = 0$). The pure scattering transmissivity is

$$T(\mu, \tau) = \left(1 + \frac{\tau}{2\mu}\right)^{-1} \tag{23}$$

while that for pure absorption is simply $e^{-\tau/\mu}$ (Beer's law). In deriving Eq. (23) we have neglected reflection from the surface ($A = 0$). Clearly, the pure scattering approximation is much better than Beer's law.

The spectral dependence of the dust transmissivity is shown in Fig. 12. This dependence is calculated from Eq. (16) using the optical properties listed in Table III. In general, the spectral dependence is weak, but transmission does show a maximum near 800 nm. The sensitivity of existing photovoltaic cells has a similar frequency dependence, but with a response that is more sharply peaked and which maximizes at slightly longer wavelengths (900–1000 nm).

Clouds. A comparison of dust transmissivities with those for a postulated Martian water ice cloud ($\omega_o = 1$, $g = 0.6$) is shown in Fig. 13. Because of their higher single scattering albedos, clouds absorb less sunlight than dust particles and for any given optical depth, they transmit more sunlight to the surface. In designing solar-powered systems, therefore, a conservative

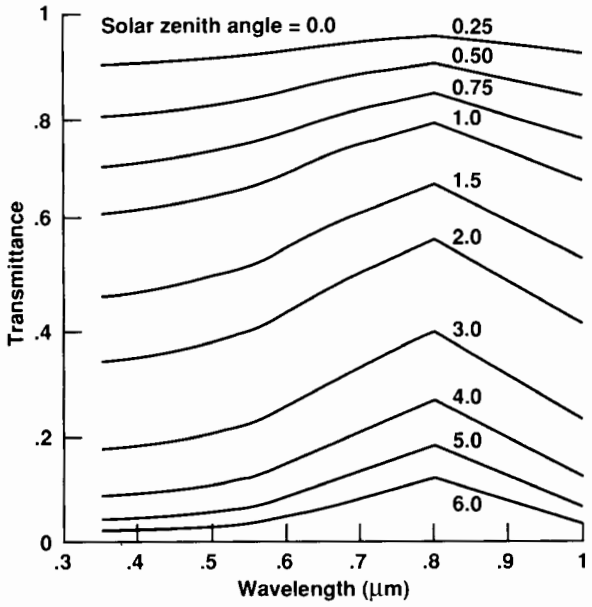


Figure 12. Spectral dependence of the dust transmissivity at visible wavelengths for several different optical depths. Transmissivities are calculated from Eq. (16).

approach would be to assume that dust rather than clouds will be the principal agent of extinction.

IV. POINTING TRADEOFFS

Because of atmospheric scattering by dust particles, tracking the direct beam may not always be the the optimum collection strategy. Even with modest amounts of dust in the atmosphere, the diffuse component can be substantial. For Martian dust, the daily integrated diffuse irradiance falling on a horizontal collector becomes comparable to the direct beam irradiance well before the optical depth reaches unity (Fig. 14). At the equator, the two components are comparable at optical depths of 0.6 to 0.8. However, at high latitudes the longer slant path and higher zenith angle lowers the value to between 0.2 and 0.3. As a solar-pointed collection system will lose some fraction of the diffuse component (it will not be able to see the full 2π steradians of the sky), there will be a "breakpoint" at which it is no longer advantageous to point the collectors.

In this section, the performance characteristics of four pointing systems of two general classes are discussed: stationary (horizontal and equator/meridian) and tracking (one-axis Sun-tracking and two-axis Sun-tracking). Stationary systems have the advantage of simplicity, but cannot realize the full potential of the direct beam. Tracking systems make up for this disadvan-

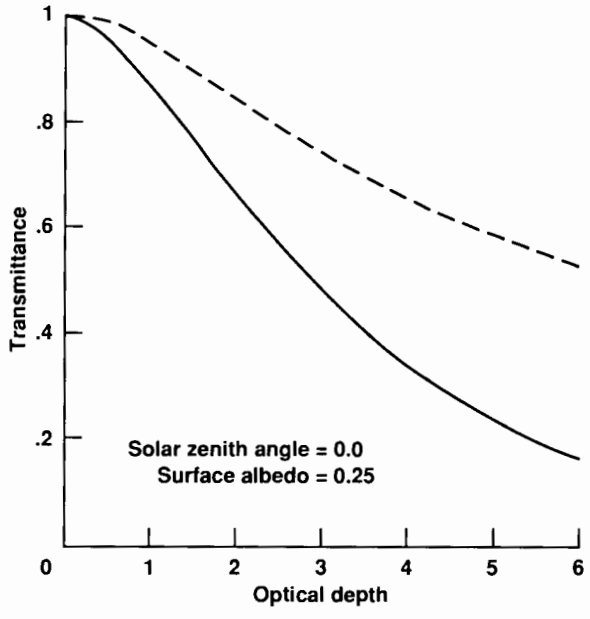


Figure 13. Dust (solid line) and water-ice cloud (dashed line) transmissivities for a single value of the zenith angle (0°). Transmissivities are calculated from Eq. (16).

tage, but not without complexity (power supplies, control systems, tracking electronics, etc.). The engineering tradeoffs between mass, complexity, reliability, etc., will depend on the particular application and are beyond the scope of this chapter.

A. Generalized Design Equations for Tilted Collectors

The equations for the total daily irradiance falling on horizontal collectors have already been presented (Eqs. 13–15). In this subsection a generalized set of design equations for tilted collectors is presented. The equations are expressed in terms of α, the angle the collector is tilted with respect to the horizontal, and α_s the angle between the collector normal and Sun direction.

In clear sky conditions (no dust in the atmosphere), the daily averaged surface insolation is simply

$$\overline{F}_{\text{tot}} = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^H \cos(\alpha_s)\mathrm{d}h \tag{24}$$

In dusty conditions, however, tilted collectors will receive an additional contribution due to reflection from the ground such that the instantaneous total irradiance is

$$F_{\text{tot}} = F_{\text{dir}} + F_{\text{diff}} + F_{\text{gnd}} \tag{25}$$

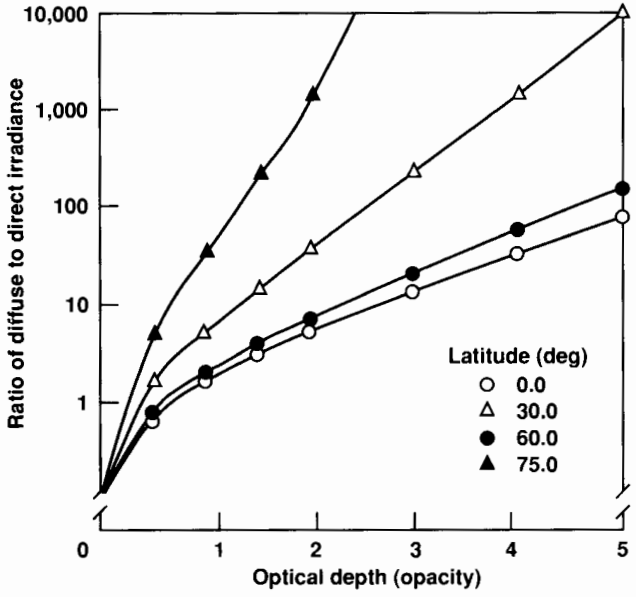


Figure 14. Ratio of the daily integrated diffuse/direct irradiance as a function of optical depth for several different latitudes at northern spring equinox.

where the direct irradiance is

$$F_{\text{dir}} = S_0 \left(\frac{\bar{r}}{r}\right)^2 \cos(\alpha_s) e^{-\frac{\tau}{\mu}} \tag{26}$$

and the diffuse irradiance is

$$F_{\text{diff}} = \mu S_0 \left(\frac{\bar{r}}{r}\right)^2 \left[\frac{f(\tau, \mu, A)}{(1-A)} - e^{-\frac{\tau}{\mu}}\right] \cos^2\left(\frac{\alpha}{2}\right). \tag{27}$$

The reduced sky coverage of a tilted collector is represented by the $\cos^2(\alpha/2)$ term in Eq. (27). This term arises by weighting the contribution from each strip of sky at angle β with respect to the collector axis by $\cos(\beta)$, and integrating over the $2(\pi - \alpha)$ steradians viewable by the collector. It is also worth noting that Eq. (26) contains an implicit dependence of the solar zenith angle through Beer's law.

The irradiance due to ground reflection depends on the surface albedo A and the total (direct+diffuse) irradiance incident on the ground. The ground reflection contribution to collector irradiance is

$$F_{\text{gnd}} = A\mu S_0 \left(\frac{\bar{r}}{r}\right)^2 \frac{f(\tau, \mu, A)}{1-A)} \sin^2\left(\frac{\alpha}{2}\right) \tag{28}$$

where $\sin^2(\alpha/2)$ represents the effect of projecting the ground irradiance onto the collector.

The daily-averaged surface insolation for dusty conditions is therefore

$$\overline{F}_{\text{tot}} = \frac{1}{\pi}\int_0^H (F_{\text{dir}} + F_{\text{diff}} + F_{\text{gnd}})\mathrm{d}h. \tag{29}$$

B. Equator/Meridian System

The Equator/Meridian (E/M) system is pointed toward the intersection of the local meridian and the equator and is inclined to the horizon at an angle equal to its latitude. At the equinoxes, therefore, the E/M system receives the full direct beam at local noon. At other seasons, the angle between the collector normal and Sun direction at noon is equal to the solar declination, i.e., $\alpha_s = \delta$. E/M systems have an advantage over horizontal systems in that the daily averaged collector/Sun angle is lower at latitudes away from the equator. However, this advantage is offset by the loss in diffuse irradiance due the limited view of the sky afforded E/M systems, particularly at high latitudes. In addition, the E/M system does not perform well during spring and summer when the Sun is higher in the sky. For the E/M system

$$\alpha = \theta \tag{30}$$

$$\cos(\alpha_s) = \cos(\delta)\cos(h). \tag{31}$$

For clear sky conditions, the daily averaged irradiance is

$$\overline{F}_{\text{tot}} = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^H \cos(\delta)\cos(h)\mathrm{d}h = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \cos(\delta)\sin(H), H < \frac{\pi}{2} \tag{32}$$

or

$$= \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \int_0^{\frac{\pi}{2}} \cos(\delta)\cos(h)\mathrm{d}h = \frac{S_0}{\pi}\left(\frac{\overline{r}}{r}\right)^2 \cos(\delta), H \geq \frac{\pi}{2}. \tag{33}$$

C. One-Axis Sun Tracking System

The one-axis system consists of a single drive on a polar mount rotating at an angular velocity equal to the rotation of the planet, but in the opposite direction. This enables the collector to track the diurnal motion of the Sun's projection on the equator. However, seasonal changes in the solar declination cannot be followed with one-axis systems. Thus, the maximum collector/Sun angle will be equal to the planet's obliquity. For the one-axis system,

$$\cos(\alpha) = \cos(\theta)\cos(h) \tag{34}$$

$$\alpha_s = \delta. \tag{35}$$

For clear sky conditions, the daily averaged surface insolation is

$$\overline{F}_{\text{tot}} = \frac{S_0}{\pi}\left(\frac{\bar{r}}{r}\right)^2 \int_0^H \cos(\delta)\mathrm{d}h = \frac{S_0}{\pi}\left(\frac{\bar{r}}{r}\right)^2 \cos(\delta)H. \tag{36}$$

D. Two-Axis Sun Tracking Systems

Two-axis systems simultaneously follow the diurnal and seasonal motion of the Sun. In this case, $\alpha_s = 0$ and the tilt angle is the same as the solar zenith angle,

$$\cos(\alpha) = \sin(\theta)\sin(\delta) + \cos(\theta)\cos(\delta)\cos(h). \tag{37}$$

For clear sky conditions, the daily averaged surface insolation is

$$\overline{F}_{\text{tot}} = \frac{S_0}{\pi}\left(\frac{\bar{r}}{r}\right)^2 \int_0^H \mathrm{d}h = \frac{S_0}{\pi}\left(\frac{\bar{r}}{r}\right)^2 H. \tag{38}$$

E. Results

The daily averaged surface insolation (including the diffuse and ground reflected components) as a function of season, dust opacity, and latitude for each of the four different pointing systems is shown in Figs. 15–19. In general, tracking systems outperform stationary systems when the opacity is low ($\tau < 0.2$). While this is true for all latitudes and seasons, the break-even point does have a latitudinal dependence. At high latitudes, for example, the Sun is never very high on the horizon and a tracking system (one- or two-axis) will annually gather about five times the energy of a horizontal collector. At the equator, on the other hand, Sun-trackers boost the annual gain by only 50% over a horizontal system. Thus, high latitude operations clearly favor a tracking system—if atmospheric scattering is negligible.

However as the optical depth increases, tracking systems loose their advantage and horizontal systems yield better performance. Importantly, this changeover occurs at optical depths that frequently occur on Mars ($\tau \leq 1$). At the poles, the longer slant path (increased scattering) and higher zenith angles (reduced direct irradiance) result in a lower break-even optical depth ($\tau \sim 0.7$) than at lower latitudes ($\tau \sim 1.0$). For all latitudes and seasons, however, horizontal systems are superior to tracking systems when the optical depth exceeds unity.

V. ENVIRONMENTAL CONCERNS

Aside from the availability of sunlight, photovoltaic performance on Mars will also depend on temperature, wind, dust accumulation and abrasion. The annual mean surface temperature of the planet is ~215 K. However, surface temperatures can vary from as low as 130 K in the winter polar regions, to as high as 300 K in the southern hemisphere subtropics (Kieffer et al. 1977). The

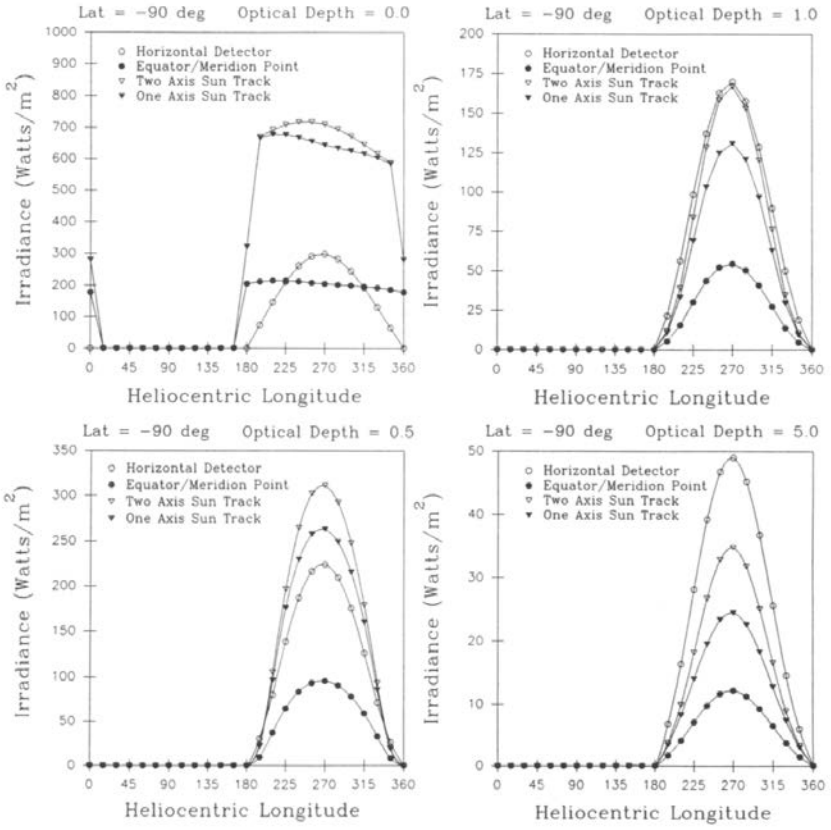


Figure 15. The daily averaged insolation collected by the different pointing systems at the south pole as a function of season and optical depth. The different systems are: horizontal detector (open circles), Equator/Meridion Pointing (filled circles), two-axis Sun tracker (open triangles), one-axis Sun tracker (filled triangles).

latter occurs at noon when the planet is close to perihelion ($L_s = 250°$); at night in these regions, the surface cools to about 180 K. Thus, the diurnal range of Martian surface temperatures can also be quite large. In general, solar cells perform better in low temperatures with peak performance occurring between 150 and 200 K; at lower temperatures their efficiency declines.

Solar arrays on Mars will have to withstand wind loads. Winds have been measured at the Viking Lander sites at a height of 1.6 m above the surface (Hess et al. 1977). Like temperature, they exhibit significant diurnal and seasonal variability. During summer, the strongest winds occurred at midday: about 6 to 7 m s^{-1} at Viking Lander 1 site, and about 3 to 4 m s^{-1} at Viking Lander 2 site. During winter, stronger mean winds developed with occasional gusts in excess of 25 m s^{-1} at Viking Lander 1 and 15 m s^{-1} at Viking Lander 2. Winds of this magnitude were relatively infrequent, and

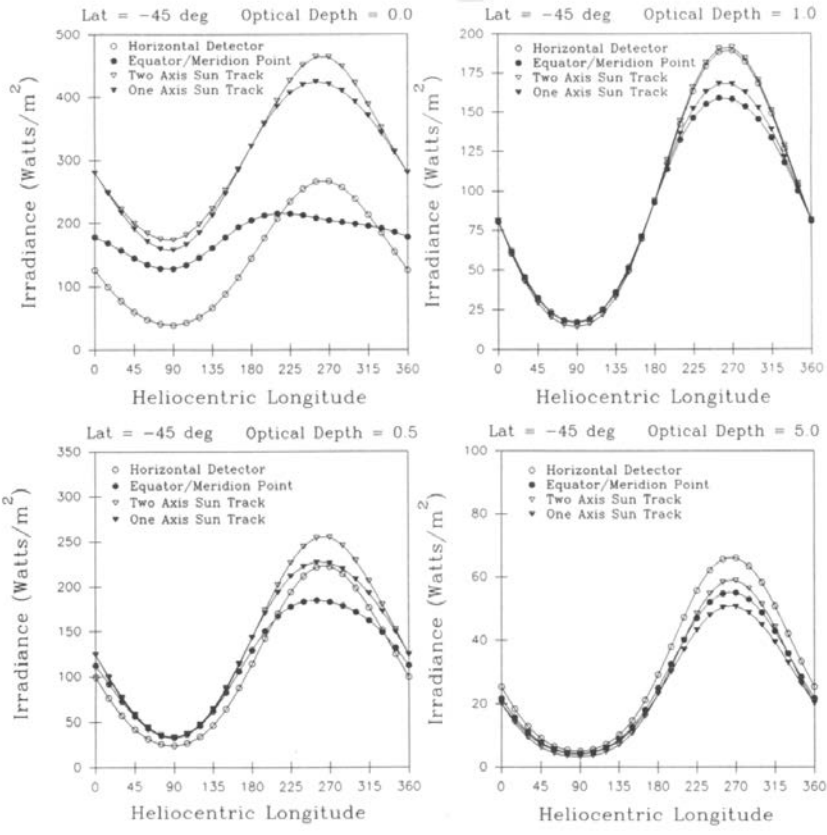


Figure 16. Same as Fig. 15, but for 45°S latitude.

occurred less than 0.01% of the time. The dynamic pressures associated with these winds are 5 to 10 Nt m^{-2}.

Dust accumulation is a potentially significant problem for solar power. During a great dust storm the atmosphere suspends as much as 10 μm of dust (see Sec. II). If the mean particle size is several μm, then 10 μm of dust represents several monolayers of potential accumulation. The performance of solar cells would be seriously degraded if covered by several monolayers of dust. Thus, occasional cleaning is required. Recent wind tunnel tests show that this might be accomplished by tilting the arrays into the wind (Gaier et al. 1990). The most efficient cleansing occurred for tilt angles of 45°, though the angular dependence was not particularly sharp. However, cleansing was minimal for all angles at wind speeds <30 m s^{-1}. As surface winds of this magnitude are relatively infrequent, a more direct method for cleansing solar arrays is required.

Abrasion is also a potentially significant problem. Abrasion by wind-blown dust of the approximate elemental composition and size as found on

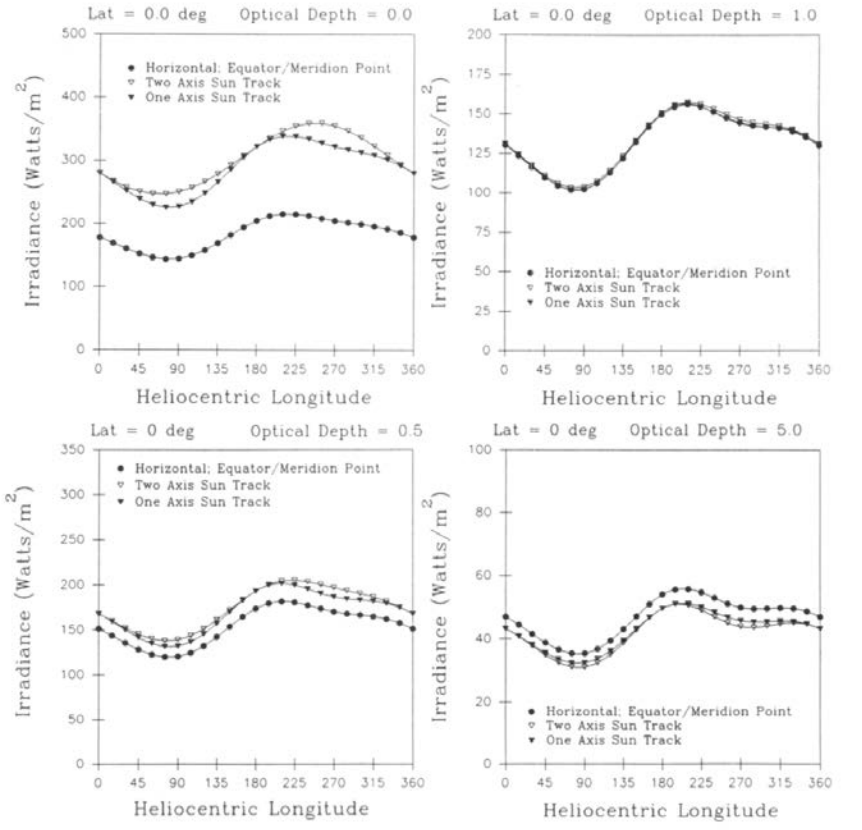


Figure 17. Same as Fig. 15, but for the equator.

Mars has been measured for typical solar cell coverglass (Gaier and Perez-Davis 1992). The results suggest that performance degradation is a concern only for very large wind speeds (85 m s^{-1}), and only for vertically oriented arrays. Worst-case abrasive loses were at the several percent level and did not seriously degrade power production. To some extent, this finding is corroborated by the relative lack of abrasion damage on the Viking Lander optical elements. Thus, as long as the arrays are glass-coated, abrasion may not be a significant problem.

However, as an additional precaution against abrasion it would be desirable to position the arrays above the saltation height. Saltation is the process believed to be responsible for lifting dust particles into the Martian atmosphere (Greeley and Iverson 1985). Winds first set into motion sand-sized particles ($\sim$100 μm in diameter) which follow modified ballastic trajectories and fall back to the surface. Upon impacting the surface, they suspend the smaller dust-sized particles which are carried to great heights. For Mars, sand-sized particles will "saltate" to about 20 cm above the surface. Positioning solar

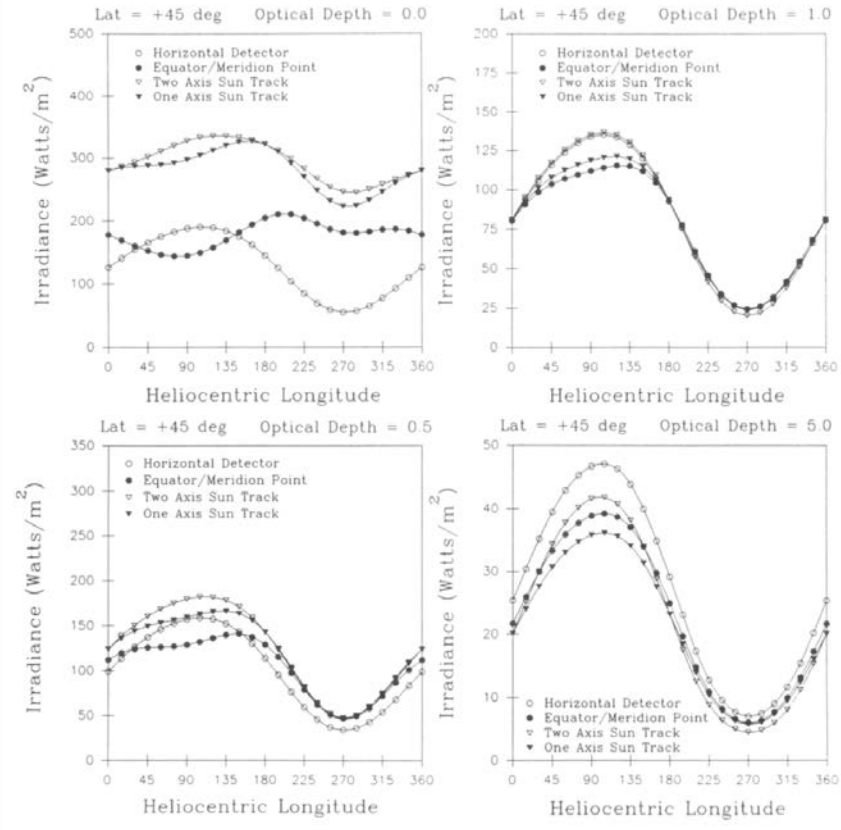


Figure 18. Same as Fig. 15, but for 45°N latitude.

arrays above this height will further reduce the risk of abrasion by wind-blown sand.

VI. APPLICATION TO SURFACE SYSTEMS

The previous sections have laid out a procedure for calculating the amount of solar radiation reaching the surface of Mars at any given latitude, season, dust loading, and pointing configuration. In this section, the state-of-the-art in photovoltaic power system technology is briefly reviewed, and the power requirements of a hypothetical mission and a resource application are analyzed.

A. Solar Cell Technology

There are three approaches to photovoltaic power: conventional flat-plate arrays, thin-film solar cells, and concentrator systems. A discussion of these

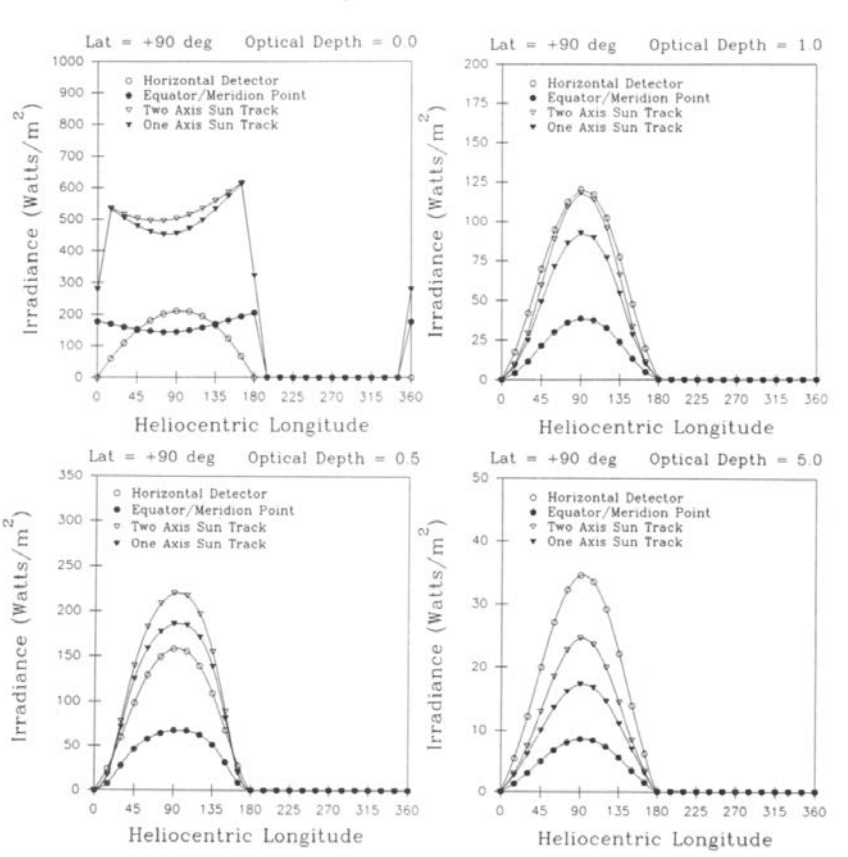


Figure 19. Same as Fig. 15, but for the north pole.

approaches can be found in Green (1982), Buresch (1983), Van Overstraaten and Mertes (1986), and Landis and Appelbaum (1991).

Conventional arrays for space flight use either crystalline silicon (Si) or gallium arsenide (GaAs) solar cells. The Si arrays have been used the most and are well developed. Their characteristics are well known and they have demonstrated conversion efficiencies of ~14% in space. Some Si cells have up to 20% conversion efficiencies, but these have not yet been flight qualified. GaAs cells are smaller and more efficient than Si cells, but are heavier and more brittle. The GaAs cells currently on the market have an average conversion efficiency of 18.5%, although higher efficiencies (21.5%) have been achieved in the laboratory. Several technologies are under development to make GaAs cells lighter. The most advanced of these is CLEFT, where a thin (5 μm) large-area cell has been separated from a single-crystal substrate (Fan et al. 1981).

Of importance to power system analysis is the specific power (power output per unit mass). At present, the specific power of Si arrays has been

demonstrated on the Solar Array Flight Experiment (SAFE) aboard the Space Shuttle at 66 W kg^{-1}. A recent design, but not yet flown, achieved 130 W kg^{-1}. This design, from the Advanced Photovoltaic Solar Array (APSA) program, is based on 2 mil thick Si cells. The long-term goals are to combine the APSA and CLEFT technologies to develop specific powers of 300 W kg^{-1}. All of these designs are intended for the zero gravity conditions of low Earth orbit and must be scaled for application to Mars. Furthermore, these figures measure the specific power at the array level which includes the cell, blanket, and array structure masses. Currently, the array structure and its power management and distribution system account for almost three-quarters of the total power system mass. Clearly, there is powerful incentive to reduce the masses of these components.

Thin-film solar cell technology is relatively recent and is not yet flight qualified. In this approach, the arrays are constructed from thin layers (1–5 μm) of highly absorbing photovoltaic materials that are deposited onto a flexible substrate. Three materials have been used: amorphous silicon (α-Si), copper indium diselenide ($CuInSe_2$), and cadmium telluride (CdTe). The advantage of thin-film cells is their potential for high specific power (1–15 kW kg^{-1}) in spite of their low inherent conversion efficiencies (5–10%). Unfortunately, the development of this technology has been driven by surface-based applications on Earth. Consequently, very little research into developing lightweight systems for space applications has been carried out. Improvements in conversion efficiency and cell performance degradation will greatly benefit this approach.

Concentrator systems focus light onto small, extremely high-efficiency solar cells. This approach has been tested in space only on small-scale experiments. Although conversion efficiencies over 30% have been demonstrated, concentrator systems have inherent limitations for surface power on Mars because they focus the direct beam only. If a power system must be designed to operate under worst-case conditions, then concentrator systems can be ruled out.

Many resource processing technologies require thermal processing. One method of achieving high temperatures is the "solar furnace" using concentrating mirrors or lenses. Because a solar furnace, like a concentrator PV system, only utilizes the direct beam component of solar radiation, the low direct beam insolation on Mars makes the utility of this approach much lower than in other envirnoments.

B. Energy Storage

For human outposts, continuous power will be required. Because Mars is subject to a 12 hr night and prolonged dust storms, some of the energy collected during the day must be stored. The amount of storage will depend on location and season. As the base location is moved away from the equator, the night length decreases during summer and increases during winter. The

night length in Earth hours is

$$T_n = 24 - T_d \tag{39}$$

where

$$T_d = 0.1369 \cos^{-1}(-\tan\theta \tan\delta) \tag{40}$$

is the number of daylight hours.

Conventional storage systems are likely to be inadequate for human outposts (McKissock et al. 1990). Their capacities range from 10 W-hr kg^{-1} for nickel-cadmium batteries operating at 25% depth of discharge (Halpert 1985), to about 80 W-hr kg^{-1} for lithium/SO_2 batteries operating at 40% depth of discharge (Linden 1984). To deliver a 25 kW hr^{-1} load would require battery masses of the order of 300 to 2500 kg. A better candidate is the hydrogen-oxygen cell (RFC) with either pressurized gas or cryogenic liquid storage. RFC energy storage densities have been shown to approach 1000 W-hr kg^{-1} (Martin et al. 1985).

C. Sample Case: PV Power System for a Hypothetical Human Outpost

The primary use of the solar radiation data presented here is to permit a more accurate assessment of photovoltaic power-system sizes and masses for system analysis and tradeoff studies. An example of such an analysis is given in Table VI for a human outpost located at the Viking Lander 1 site with a crew size of 15, and a power demand of 25 kWe during the day and 12.5 kWe at night. The analysis is carried out for two different solar cell options (α-Si and GaAs), and two different energy storage options (Li-SO_2 batteries and hydrogen-oxygen RFCs). The cell efficiencies of both types are near their optimum values. Both cell types are assumed to be integrated into a suitable lightweight flexible blanket assembly with a specific mass of 0.5 kg m^{-2} for the GaAs cell, and 0.04 kg m^{-2} for the α-Si cell. Added to these is the array structure specific mass which is 0.56 kg m^{-2} for each type. The storage system efficiency takes into account the storage "round-trip" efficiency as well as the power management and distribution efficiencies. For Li/SO_2 batteries, the round trip efficiency is 0.9 and 40% depth of discharge. For each system, the array is assumed to be fixed in the horizontal plane with no pointing capability as might be the case if operational simplicity is desired for deployment. For a relatively short mission—one that arrives and departs during midsummer ($L_s = 141°$)—Table VI indicates that at least 660 m^2 of solar panels and 845 kg of total mass (panels + energy storage) are required.

D. Sample Case: PV Power System for the "Mars Direct" Scenario

In-situ propellant generation on Mars is an option for drastically reducing the cost of Mars expeditions. Baker and Zubrin (1990) and Zubrin et al. (1991) have proposed that 107 tons of methane/oxygen propellant can be produced on Mars from 5.7 tons of hydrogen brought from Earth plus CO_2 from the

TABLE VI

Mars Mission Power System Comparison

Performance at $L_s = 141$, 22°3 N, $\tau = 0.35$

	Day length:				13.18 hr		
	Night length:				11.47 hr		
	Average Daytime Insolation:				305 W m^{-2}		
	Energy Storage Requirement:				143 kW-hr		
System	Cell Efficiency	Array Size (m^2)	Array sp. Mass (kg m^{-2})	Array Mass (kg)	Storage Capacity (W-hr/kg)	Storage Mass (kg)	System Mass (kg)
α-Si/Li-SO_2	0.12	1210	0.6	726	200	1790	2516
GaAs/Li-SO_2	0.22	660	1.06	700	200	1790	2490
α-Si/RFC	0.12	1210	0.6	726	1000	145	871
GaAs/RFC	0.22	660	1.06	700	1000	145	845

Martian atmosphere. They estimate that the energy requirements for this scenario could be met with a 100 kW reactor running continuously (day and night) for 155 days. This is equivalent to 370 MW-hr of electrical energy.

To determine if these requirements can be met with a surface photovoltaic array, the following assumptions are made:

1. The production plant is located at the Viking Lander 1 site.
2. Production is completed during spring and summer.
3. The production process stops at night.
4. The arrays are fixed and horizontal (no tracking).
5. Solar array specific mass = 0.9 kg m^{-2} (APSA technology).
6. Solar array efficiency = 20% at Mars (GaAs technology).
7. No mass allocation for energy storage.

The global insolation on a horizontal surface at Viking Lander 1 averages slightly more than 3 kW-hr m^{-2} per day during spring and summer. This period is chosen because of its relatively low atmospheric dust content. Running the system for this time (half a Mars year) gives a total insolation of about 1000 kW-hr m^{-2}. Thus, the arrays will produce about 200 kW-hr m^{-2} of electrical energy, and the required array area and mass is 1850 m^2 and 1.67 metric ton. This is comparable to the baseline power system of the 100 kW nuclear reactor at 3.96 metric tons considered in the Zubrin/Baker architecture.

VII. CONCLUSION

The utilization of solar energy as a power source for surface-based operations on Mars depends on the attenuating effects of its atmosphere as well as the location and season of interest. In this chapter, we have summarized what is known about the Martian atmosphere as it relates to solar power, and have presented a procedure for calculating the amount of solar radiation available at the surface as a function of latitude, season, time of day, and atmospheric dust load (optical depth).

A major concern for solar power on Mars has been the dust storms which can occur on local to global scales, and which can last for several months. The concern is not only for their effect on array output, but also for their potential to cover and/or abrade the arrays themselves. The former appears to be more of an issue and may require some kind of cleansing device. The latter is less of an issue as long as the arrays have a glass overcoat. However, as far as the attenuating effect of suspended dust particles is concerned, solar power remains a viable energy source on Mars. As has been pointed out by others, dust particles suspended in the Martian atmosphere scatter as well as absorb, and it is this property that allows enough sunlight to reach the surface to justify photovoltaic systems.

Acknowledgments. This work was support by the Planetary Atmospheres Program of the National Aeronautics and Space Administration and was

performed in part at NASA's Ames and Lewis Reserach Centers.

REFERENCES

Appelbaum, J., and Flood, D. 1989*a*. *Photovoltaic Power System Operation in the Mars Environment*, NASA TM-102075.

Appelbaum, J., and Flood, D. 1989*b*. Solar radiation on Mars. *Solar Energy* 45:353–363, and NASA TM-102299.

Appelbaum, J., and Flood, D. 1990. *Solar Radiation on Mars—Update 1990*, NASA TM-103623.

Arvidson, R. E., Guiness, E. A., Moore, H. J., Tillman, J. E., and Wall, S. D. 1983. Three Mars years: Viking Lander 1 imaging observations. *Science* 222:463–468.

Baker, D., and Zubrin, R. M. 1990. Mars direct: Combining near-term technologies to achieve a two-launch manned Mars mission. *J. British Interplanet. Soc.* 43:519–523.

Banks, P. M., and Kockarts, G. T. 1973. *Aeronomy Part A* (New York: Academic Press).

Barth, C. A. 1985. Photochemistry of the atmosphere of Mars. In *The Photochemistry of Atmospheres, Earth, the Other Planets, and Comets*, ed. J. Levine (Orlando, Fl.: Academic Press).

Briggs, G. A., and Leovy, C. B. 1974. Mariner 9 observations of the Mars north polar hood. *Bull. Amer. Meteorol. Soc.* 55:278–296.

Buresch, M. 1983. *Photovoltaic Energy Systems: Design and Installation* (New York: McGraw-Hill).

Clancy, R. T., and Lee, S. W. 1991. A new look at dust clouds in the Mars atmosphere: Analysis of emission-phase-function sequences from global Viking IRTM observations. *Icarus* 93:135–158.

Clark, B. C., Baird, A. K., Weldon, R. J., Tsusaki, D. M., Schnabel, L., and Candelaria, M. P. 1982. Chemical composition of Martian fines. *J. Geophys. Res.* 87:10059–10067.

Colburn, D., Pollack, J. B., and Haberle. R. M. 1989. Diurnal variations in optical depth at Mars. *Icarus* 79:159–189.

Curran, R. J., Conrath, B. J., Hanel, R. A., Kunde, V. G., and Pearl, J. C. 1973. Mars: Mariner 9 spectroscopic evidence for H_2O clouds. *Science* 182:381–383.

Daumont, D., Brion, J., and Malicet, J. 1983. Measurement of total atmospheric ozone: Consequences entailed by new values of ozone absorption cross-sections at 223K in the 310–350 nm spectral range. *Planet. Space Sci.* 31:1229–1234.

Fan, J. C. C., Bozler, C. O., and McClelland, R. W. 1981. Thin film Ga-As solar cells. In *Proc. of 15th IEEE Photovoltaic Specialists Conf.* (New York: Inst. of Electrical and Electronic Engineers), pp. 666–672.

French, R. G., Gierasch, P. J., Popp, B. D., and Yerdon, R. J. 1981. Global patterns in cloud forms on Mars. *Icarus* 45:468–493.

Gaier, J. R., and Perez-Davis, M. E. 1992. *Effect of Particle Size of Martian Dust On the Degradation of Photovoltaic Cell Performace*, NASA TM-105232.

Gaier, J. R., Perez-Davis, M. E., and Marabito, M. 1990. *Aeolian Removal of Dust from Photovoltaic Surfaces on Mars*, NASA TM-102507.

Geels, S., Miller, J. B., and Clark, B. C. 1989. Feasibility of using solar power on Mars: Effects of dust storms on incident solar radiation. In *Case for Mars III*,

ed. C. R. Stoker (San Diego: Univelt), pp. 505–516.
Greeley, R., and Iversen, J. D. 1985. *Wind as a Geological Process* (Cambridge: Cambridge Univ. Press).
Green, M. A. 1982. *In Solar Cells: Operating Principles, Technology and System Application* (Englewood Cliffs, N. J.: Prentice-Hall).
Haberle, R. M., Leovy, C. B., and Pollack, J. B. 1982. Some effects of global dust storms on the atmospheric circulation of Mars. *Icarus* 50:322–367.
Halpert, G. 1985. The design and application of nickel-cadmium batteries in space. *J. Power Sources* 15:119–140.
Hanel, R., Conrath, B., Hovis, W., Kunde, V., Lowman, P., Maguire, W., Pearl, J., Pirraglia, H., Prabhakara, C., Schlachman, B., Levin, G., Straat, P., and Burke, T. 1972. Investigation of the Martian environment by infrared spectroscopy on Mariner 9. *Icarus* 17:423–442.
Hess, S. L., Henry, R. M., Leovy, C. B., Ryan, J. A., and Tillman, J. E. 1977. Meteorological results from the surface of Mars: Viking 1 and 2. *J. Geophys. Res.* 82:4559–4574.
Hord, S. W., Barth, C. A., and Stewart, A. I. 1972. Mariner 9 ultraviolet spectrometer experiment: Photometry and topography of Mars. *Icarus* 17:443–456.
Hunt, G. E. 1979. Thermal infrared properties of the martian atmosphere. 4. Predictions of the presence of dust and ice clouds from Viking IRTM spectral measurements. *J. Geophys. Res.* 84:2865–2874.
Jaquin, F., Gierasch, P., and Kahn, R. 1986. The vertical structure of limb hazes in the Martian atmosphere. *Icarus* 68:442–461.
Joseph, J. H., Wiscombe, W. J., and Weinman, J. A. 1976. The delta-Eddington approximation for radiative flux transfer. *J. Atmos. Sci.* 28:833–837.
Kahn, R. 1984. The spatial and seasonal distribution of Martian clouds, and some meteorological implications. *J. Geophys. Res.* 89:6671–6688.
Kahn, R. 1990. Ice haze, snow, and the Mars water cycle. *J. Geophys. Res.* 95:14677–14694.
Kahn, R., Goody, R., and Pollack, J. 1981. The martian twilight. *J. Geophys. Res.* 86:5833–5838.
Kahn, R., Martin, T. Z., Zurek, R. W., and Lee, S. W. 1992. The Martian dust cycle. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 1017–1053.
Kieffer, H. H., Martin, T. Z., Peterfreund, A. R., Jakosky, B. M., Miner, E. D., and Palluconi, F. D. 1977. Thermal and albedo mapping of Mars during the Viking primary mission. *J. Geophys. Res.* 82:4249–4291.
Kuhn, W. R., and Atreya, S. K. 1979. Solar radiation incident on the Martian surface. *J. Molec. Evol.* 14:57–64.
Landis, G., and Appelbaum, J. 1990. Design considerations for Mars photovoltaic power systems. In *Proc. of 21st IEEE Photovoltaic Specialists Conf.* (New York: Inst. of Electrical and Electronic Engineers), pp. 1263–1270.
Landis, G., and Appelbaum, J. 1991. Photovoltaic power options for Mars. *Space Power* 10:225–237.
Leovy, C. B., Briggs, G., Young, A., Smith, B., Pollack, J., Shipley, E., and Wildey, R. 1972. The Martian Atmosphere: Mariner 9 television experiment progress report. *Icarus* 17:373–393.
Leovy, C. B., Briggs, G. A., and Smith, B. A. 1973. Mars atmosphere during the Mariner 9 extended mission: Television results. *J. Geophys. Res.* 78:4252–4266.
Linden, D. 1984. *Handbook of Batteries and Fuel Cells* (New York: McGraw-Hill).
Martin, L. J., and Zurek, R. W. 1993. An analysis of dust activity on Mars. *J. Geophys. Res.*, in press.

Martin, R. E., Garow, J., and Michaels, K. B. 1984. *Regenerative Fuel Cell Energy Storage System for a Low Earth Orbit Space Station*, NASA CP-174802.

Martin, T. Z. 1986. Thermal infrared opacity of the Martian atmosphere. *Icarus* 66:2–21.

McElroy, M. B., and Donahue, T. M. 1972. Stability of the Martian atmosphere. *Science* 177:986–988.

McKissock, B. I., Kohout, L. L., and Schmitz, P. C. 1990. *A Solar Power System for an Early Mars Expedition*, NASA TM-103219.

Murphy, J. R., Haberle, R. M., Toon, O. B., and Pollack, J. B. 1993. Martian global dust storms: Zonally symmetric numerical simulations including size dependent particle transport. *J. Geopys. Res.*, in press.

Owen, T., Biemann, K., Rushneck, D. R., Biller, J. E., Howarth, D. W., and Lafleur, A. L. 1977. The composition of the atmosphere at the surface of Mars. *J. Geophys. Res.* 82:4635–4639.

Parkinson, T. D., and Hunten, D. M. 1972. Spectroscopy and Aeronomy of O_2 on Mars. *J. Atmos. Sci.* 29:1380–1390.

Peterfreund, A. P. 1985. Contemporary Aeolian Processes on Mars: Local Dust Storms. Ph. D. Thesis, Arizona State Univ.

Pollack, J. B., Colburn, D., Kahn, R., Hunter, J., Van Camp, W., Carlston, C., and Wolfe, M. 1977. Properties of aerosols in the martian atmosphere, as inferred from Viking Lander imaging data. *J. Geophys. Res.* 82:4479–4496.

Pollack, J. B., Colburn, D., Flasar, F. M., Kahn, R., Carlston, C., and Pidek, D. 1979. Properties and effects of dust particles suspended in the martian atmosphere. *J. Geophys. Res.* 84:2929–2945.

Pollack, J. B., Haberle, R. M., Schaeffer, J., and Lee, H. 1990. Simulations of the general circulation of the martian atmosphere: I. Polar processes. *J. Geophys. Res.* 95:1447–1473.

Toon, O. B., Pollack, J. B., and Sagan, C. 1977. Physical properties of the particles composing the martian dust storm of 1971–1972. *Icarus* 14:235–244.

Toulmin, P., III, Baird, A. K., Clark, B. C., Keil, K., Rose, H. J., Jr., Christian, B. P., Evans, P. H., and Kelliher, W. C. 1977. Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* 82:4625–4634.

Van Overstraaten, R. J., and Mertes, R. P. 1986. *Physics Technology and Use of Photovoltaics* (London: Adam Hilger).

Wells, R. A. 1979. In *Geophysics of Mars* (Amsterdam: Elsevier).

Zubrin, R. M., Baker, D. A., and Gwynne, O. 1991. Mars Direct: A Simple, Robust, and Cost Effective Architecture for the Space Exploration Initiative. AIAA Paper 91-0326.

Zurek, R. W. 1981. Inference of dust opacities for the 1977 martian great dust storms from Viking Lander 1 pressure data. *Icarus* 45:202–215.

Zurek, R. W. 1982. Martian great dust storms: an update. *Icarus* 50:288–310.

Zurek, R. W., and Martin, L. J. 1993. Interannual variability of planet encircling dust storms on Mars. *J. Geophys. Res.*, in press.

CHEMICAL AND PHYSICAL PROPERTIES OF THE MARTIAN SATELLITES

JEFFREY F. BELL and FRASER FANALE
University of Hawaii

and

DALE P. CRUIKSHANK
NASA Ames Research Center

Data bearing on the composition of Phobos and Deimos are reviewed. Infrared spectroscopy indicates that neither satellite has significant amounts of bound water in its surface minerals. However, magnetospheric data from the Phobos 2 spacecraft can be interpreted as suggesting that the satellites are outgassing water vapor. Theoretical models indicate that if Phobos originally formed with an ice component, it could still retain an icy "permafrost" core at depths less than 100 m near the poles. A significant ice component is consistent with the surprisingly low density (1.9 g cc^{-1}) measured by Phobos 2; but a very high porosity (~50%) could also explain this result. The most likely composition for Phobos and Deimos is a "CM3" carbonaceous-chondrite-like assemblage of anhydrous silicates, carbon, organic compounds and ice. This composition suggests that the Martian satellites were originally formed in the outer asteroid belt and later captured by Mars. Our knowledge of Phobos and Deimos is still inadequate to evaluate intelligently their usefulness as resource bases, and another unmanned mission to Phobos is essential before any serious planning for its use is carried out.

I. INTRODUCTION

The Martian satellites Phobos and Deimos have long attracted space mission planners as possible stepping stones to Mars. Their convenient orbital location, together with their suspected water-rich carbonaceous chondrite composition, has led to many proposals that rocket propellants extracted from the surface minerals of these bodies could be used to refuel Mars-bound spacecraft (see, e.g., Cole and Cox 1964). The most useful compound which could be found is water, which not only is essential to life support in space, but is the raw material for fueling hydrogen-oxygen rocket engines. In 1988, the Soviet Union recognized the importance of Phobos and Deimos by launching the first two examples of a new generation of highly sophisticated unmanned space probes to Phobos. These spacecraft had as their primary mission determining the water content of Phobos. Despite the failure of both probes before the

planned close encounter and landing, Phobos 2 did return valuable data.

By coincidence, in the fall of 1988 Mars passed through the most favorable opposition for Earth-based observations in the remainder of this century, encouraging renewed telescopic observations. This chapter reviews our current knowledge of the chemical and physical state of Phobos and Deimos, emphasizing the new results published since 1988, and their implications for resource utilization.

A. Relevant Data

Spectroscopy of sunlight reflected from asteroid surfaces has proven to be one of the most valuable methods of determining the composition of planetary bodies, and this technique has been of particular use in determining the mineralogy of asteroids (see, e.g. Gaffey et al. 1989). Despite the proven value of such techniques when applied to asteroids they have not significantly improved our knowledge of the superficially similar Martian satellites for two reasons: (1) the spectral observations of Phobos and Deimos made by spacecraft before 1988 usually employed instruments which were not optimized for the satellites, but rather for the surface and atmosphere of Mars; (2) the proximity of these objects to Mars as seen from Earth introduces severe observational problems into any groundbased observing programs due to the scattered light halo from the planet. Instruments and telescopes which are designed to observe isolated point sources like asteroids cannot usually produce accurate photometry near a bright extended source like Mars.

Despite these problems, a sizeable body of observational data has accumulated. The orbital motions of Earth and Mars interact so that favorable oppositions recur at ~16-year intervals, and attempts to observe Phobos and Deimos follow this cycle. The first broadband photometric data for the Martian satellites was obtained by Kuiper at the 1956 perihelic opposition (Harris 1961). Similar data in the UBV system was obtained by Zellner (1972) and Zellner and Capen (1974). Barth et al. (1972) obtained a spectrum of Phobos from 0.26 to 0.35 μm with the Mariner 9 UV spectrometer. Finally, Pollack et al. (1978) measured Phobos from the surface of Mars in three broad filters with the Viking Lander cameras. Together, these data sets cover the wavelength range from about 0.2 μm to 0.9 μm. The composite spectrum spectrum is very red shortward of about 0.4 μm and is flat at longer wavelengths (Fig. 1); it is very similar to the spectrum of the asteroid Ceres (Pang et al. 1978). Spectra from the KRFM instrument on Phobos 2 are generally consistent with this earlier data, despite some uncertainty in the calibration (Ksanfomality et al. 1991).

The favorable opposition in the fall of 1988 allowed a new generation of telescopic instruments to be employed in attempts to observe the satellites. Bell et al. (1989*a*) obtained the first near-infrared reflection spectra of Deimos, allowing a direct measurement of the hydration state of the surface minerals. Lucey et al. (1989) attempted to obtain high-resolution CCD spectra of both Phobos and Deimos. While the shape of these latter spectra are generally

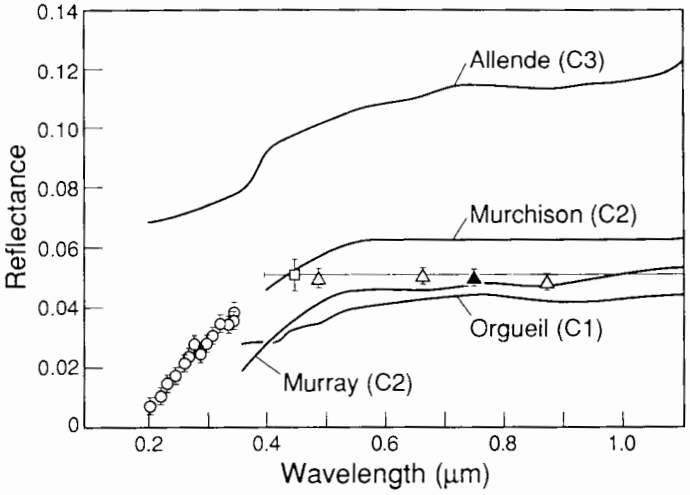


Figure 1. Mariner 9, Viking Lander, and pre-1988 groundbased ultraviolet-visible photometry of Phobos compared with lab spectra of various carbonaceous chondrites (figure from Pollack et al. 1978).

similar to the data obtained in the 1970s, the ultraviolet absorptions appear impossibly deep for any real geologic material, and much deeper than that implied by the Mariner 9 data. For these reasons we will not consider the Lucey et al. data further.

The Soviet spacecraft Phobos 1 and Phobos 2 were the first space vehicles intended primarily to study an asteroidal body; they carried a wide variety of instruments specifically designed to determine the chemical composition and subsurface geology of Phobos. The detailed study of Phobos would have comprised 6 main elements: (1) measurements of magnetospheric effects induced by Phobos; (2) visible, near-infrared and thermal infrared spectral mapping of the surface; (3) long- and short-wavelength radar sounding of the regolith; (4) a low "terrain-following" pass by the orbiter during which regions of the upper regolith would have been vaporized by means of laser and particle beams, and the resulting vapor sampled by instruments on the orbiter; (5) a long-lived lander which would have attached itself to the surface with a harpoon-like device; and (6) a mobile lander or "hopper" which would have moved from place to place on the surface to search for compositional variations. Unfortunatly, most of this ambitious program was not carried out. Phobos 1 was lost during interplanetary cruise due to an incorrect command from Earth, and Phobos 2 lost contact with Earth before it had approached closer than 200 km to Phobos. However, it had already carried out several sessions of distant observations with three optical remote-sensing instruments. Also, tracking of the spacecraft provided a much improved density for Phobos. We will discuss these results in detail below.

B. Do Phobos and Deimos Contain Hydrated Silicates?

Ceres and many similar asteroids possess a broad absorption band at 2.7 to 3.2 μm due to bound water in the mineral structure of clays (Lebofsky 1980; Lebofsky et al. 1981; Jones et al. 1990); the depth of this band appears to be correlated with the depth of the absorption below 0.4 μm (Feierberg et al. 1985). Thus the similarity of Phobos and Deimos to Ceres in the 0.2 to 0.9 μm range pointed out by Pang et al. (1978,1980) has led to the general belief that they have a composition dominated by hydrated clay minerals, probably analogous to the CI and CM meteorites. This interpretation suggested that the Martian satellites could be mined for water to support future space activities near Mars. This possibility has excited space travel planners for over 20 years (see, e.g., Cole and Cox 1964).

The clay-rich CI-like interpretation of Phobos reigned unchallenged in the 1978–88 period; indeed many discussions of space resources during this time took the presence of water-rich minerals on Phobos as virtually certain. During the 1988 opposition, both Earth-based telescopes and the Phobos 2 spacecraft were used in attempts to unambiguously detect hydrated silicates on the Mars satellites using the spectroscopic techniques which have proved so successful with low-albedo asteroids.

Bell et al. (1989*a*) reported 1.2 to 3.2 μm spectral data of Deimos obtained on 7 October 1988 with the 3-m NASA Infrared Telescope Facility at Mauna Kea, Hawaii. Figure 2 shows the IRTF data plotted along with models of the total flux assuming the "standard thermal model" long used for asteroids, and the spectral reflectance of various meteorites. The solid curve assumes that Deimos has the spectrum of Orgueil, one of the most hydrated carbonaceous chondrites. It is clear that Deimos is much drier than Orgueil. The dashed model curve was generated with the reflectance of Karoonda, a totally anhydrous chondrite. This comparison demonstrates that Deimos probably has no water-bearing clay minerals in its optical surface at all.

Another interesting aspect of the 1988 IRTF observations is that the short-wavelength data points fall well below the model curves, implying a very red reflectance curve in the 1.2 to 2.2 μm region. Direct comparison of these colors with those of asteroids reveals that Deimos resembles the D-class asteroids found mostly in the Trojan asteroid clouds, rather than the C-class asteroids in the main belt which are a close spectral match to Phobos and Deimos in the visible wavelength region.

Phobos 2 carried a joint French-Soviet instrument, the Mapping Infrared Spectrometer (ISM), which obtained spectra of spatial pixels in the range 0.8 to 3.1 μm covered by arrays having a total of 128 PbS detectors (Bibring et al. 1989). The ISM was the first near-infrared imaging spectrometer to be flown on a planetary mission. The overall slope of the spectrum of Phobos from 0.8 to 3.1 μm is quite red (Bibring et al. 1991; Cruikshank et al. 1991), bearing a close similarity to the color of the average of several D-type asteroids and appearing much redder than the rather neutral color of C-type asteroids in

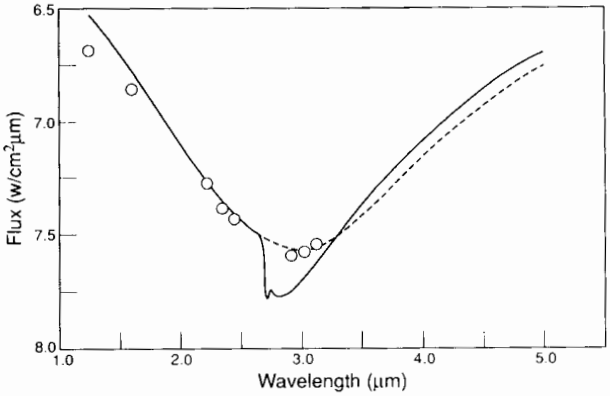


Figure 2. Telescopic observations of Deimos by Bell et al. (1989*a*) (circles) compared with flux models assuming standard asteroid thermal model and spectral reflectance of hydrous (solid line) and anhydrous (dashed line) carbonaceous chondrites. Telescope data are scaled to models at 2.2 μm.

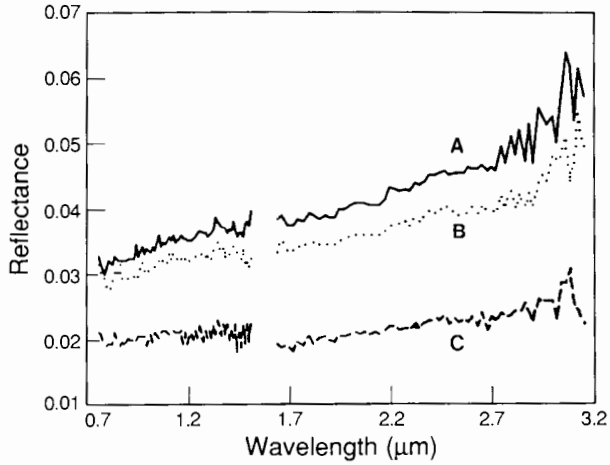


Figure 3. Spectral reflectances of three regions on Phobos measured by ISM instrument on Phobos 2 (figure from Langevin et al. 1991). Note red continuum slope resembling D-class asteroids and lack of any hydrated silicate absorption near 3 μm.

this spectral region (see Fig. 3). This spectrum is also a close match for the spectrum of Deimos measured by Bell et al. (1989*a*). The three spectra in Fig. 3 illustrate the albedo variations on Phobos known earlier from Mariner 9 and Viking Orbiter images, which are mostly attributible to particle-size variations (French et al. 1988).

The ISM instrument on the Phobos 2 spacecraft offered the first oppor-

tunity to search for hydrated minerals on Phobos. Langevin et al. (1991) summarize the negative search for the 2.75 μm spectral signature of the hydroxyl ion or bound water molecules. They note that the major variations in the spectra around 2.7 μm of spatially resolved regions on Phobos are caused by variations in the thermal radiation from the surface, which contributes up to 25% of the total flux near the center of the image and is diminished toward the terminator and limb. Such changes in the continuum slope as are observed appear to be related to changes in grain size or to the effect of the small solar incidence angle. These results rule out a significant component of hydrated silicates on the surface of Phobos.

C. Do Phobos and Deimos Contain Ice?

The apparent lack of hydrated minerals on the optical surface of both Phobos and Deimos is part of a larger question: why do the surfaces of the dark, "primitive" asteroids in the main belt apparently become drier with increasing distance from the Sun (Jones et al. 1990)? This is inconsistent with the traditional picture in which hydrated silicates form in the solar nebula by the reaction of water in the gas phase on pre-existing anhydrous silicate grains. In this concept, the degree of hydration in primitive chondritic material should increase with distance from the Sun, due to lower nebular temperatures which favor this reaction. In addition, ice should be an important constituent of outer belt and Trojan asteroids, if condensation continued there down to temperatures of ~200 degrees K. There are three proposed explanations for why the asteroid observations to date do not agree with the classical picture:

1. The clay absorption bands are suppressed by the increasing abundance of dark red organic polymers. Laboratory studies investigating this possibility suggest that the fundamental OH stretch band at 2.9 μm is too strong to be completely suppressed in this way (Jones et al. 1989).
2. The abundance of water in the original solar nebula declined with distance from the Sun, at least in the region of the outer asteroid belt. It is difficult to reconcile this with the Jovian satellites Ganymede and Callisto, which are at least one-half water ice.
3. The water in all asteroids was originally ice, but only in the middle asteroid belt were bodies heated enough after accretion to melt the ice and create hydrated silicates through the action of "groundwater" protected from the vacuum of space by a permafrost layer. In the P-class and D-class asteroids, ice is still present and was never mobilized.

Most recent studies of this issue have accepted (3) as the correct explanation (Jones et al. 1990). This model is supported by the distribution of asteroid types in the inner asteroid belt, which indicates a drastic fall-off in heating with solar distance. The zone of hydration in the middle belt is thus a zone of mild metamorphism between the completely melted asteroids of the inner belt and the primitive material of the outer belt (Bell et al. 1989*b*). Furthermore, recent studies of the complex clay minerals in CI and CM me-

teorites demonstrate that they formed in a ground-water environment on their parent asteroids, not from gas-solid reactions in the nebula (Zolensky and McSween 1988). Besides this observational evidence, theoretical studies of such nebular reactions suggest that they are too slow under likely nebular conditions to produce significant amounts of hydrated silicates (Prinn and Fegley 1987,1988). Thus there is currently wide agreement that clay minerals known to exist in asteroids possessing the 2.9 μm absorption band are the result of post-accretion heating, and that truly primitive asteroids (like comets?) would *not* contain hydrated silicates, but rather would contain any water component in the form of ice. In fact, it is possible that the depth of the 2.9 μm band is *anticorrelated* with the *total* abundance of water in the low-albedo asteroids. In this conceptual structure, the observed lack of any 2.9 μm clay absorption band on Phobos and Deimos does not in any way prohibit the presence of a sizeable fraction of ice in the deep interiors. The absence of detectable ice absorption bands in the spectrum is not a problem either, because any ice near the surface would quickly evaporate at the temperatures of Phobos and Deimos (which are much warmer than main-belt asteroids).

Considerations such as these motivated Fanale and Salvail (1989*b*) to construct a quantitative model of ice loss/retention for Phobos. The model was intended to answer the following questions: if as suggested by the above arguments, Phobos originally contained randomly distributed ice in its interior, then (a) what would be the depth as a function of latitude to which the ice interface would have receded after 4.5 Gyr, and (b) what would be the current water flux? This model assumed a steady-state heat flow condition within Phobos, including internal heat flow from equator to pole. The model was run assuming porosities of 0.1 and 0.5, and pore sizes of 1 to 10 μm in the ice-free region overlying the receding ice interface. It was found that the depth to the ice interface could currently be as low as 10 to 100 m at 80 deg latitude and ~200 m at the equator, and that the current water flux from Phobos would be 1×10^{-13} to 1×10^{-14} g cm^{-2} s^{-1}. In a later study (Fanale and Salvail 1990), a more sophisticated model was used in which the previous assumption of zero obliquity for Phobos was replaced with the actual obliquity history of Mars, which Phobos is currently constrained to follow. In addition, the earlier study had incorrectly assumed a tortuosity value of 2.0 (typical of nearly straight pores) whereas the new model used the canonical value for real geologic materials of 5.0. The improved model suggested ice depths of 20 to 60 m at 80 deg latitude and 200 to 700 m at the equator (see Fig. 4). It also predicted a global water flux as high as 3 g s^{-1} from Phobos.

The Phobos 2 spacecraft carried a radar sounding device similar to those which have been used to measure permafrost depths in arctic regions of the Earth, and it was hoped that this instrument would provide a definitive test of these model predictions. Unfortunatly the spacecraft failed before reaching useful radar range. However, two other discoveries by Phobos 2 are consistent with the existence of buried ice.

First, the density of Phobos was measured at only 1.90±0.10 (Avanesov

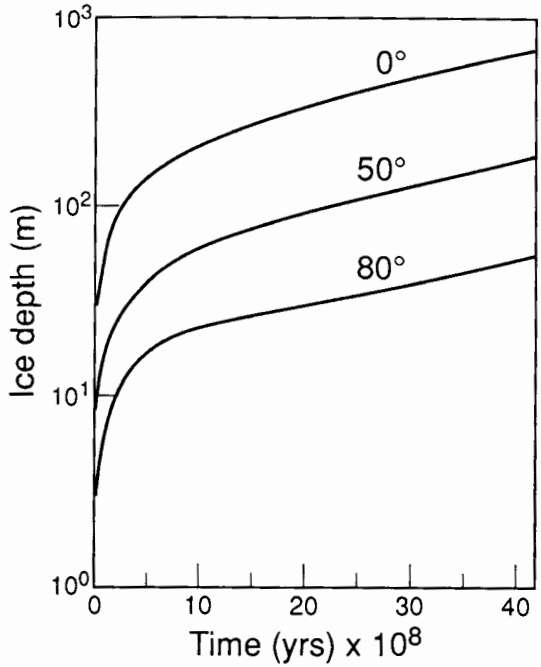


Figure 4. Depths to top of permafrost layer in Phobos over geologic time predicted by model of Fanale and Salvail (1990). Note predicted shallower depths at higher latitudes.

et al. 1991), which is significantly lower than that of any known meteorite and requires a sizeable proportion of ice and/or empty pore space to exist in the interior of Phobos. Well-compacted anhydrous chondritic meteorites have a density of about 3.5 g cm^{-3}. Lower densities are found in other chondrites due to abundant pore space and/or water of hydration. The IRTF and ISM results cited above indicate that water of hydration does not contribute to the low density of Phobos. To explain the observed density with pore space alone would require Phobos to consist of at least 50% empty space. The only rocks with this much porosity are some lunar breccias which are soil rewelded by impact. Thus an ice-free Phobos would have to be a rubble-pile consisting entirely of regolith. This appears unlikely in view of the existence of large topographic features such as the walls of Stickney crater, and several distinct ridges. At the other extreme, a Phobos composed of chondritic silicates plus ice in the "cosmic" or cometary ratio (2/1 ice/rock by mass) would also explain the density without any pore space. (One might think a Phobos of cometary composition should resemble an active comet, but in fact the Fanale and Salvail (1990) model predicts that the current water flux out of Phobos should not increase with ice contents above ~0.1%.) The true situation on Phobos is probably somewhere between these two extreme models.

Furthermore, studies of the particles and fields data returned by Phobos 2 have offered evidence that Phobos is actually emitting water molecules into space, at the theoretically expected rate. Dubinin et al. (1990) state that "The interaction of Phobos with the solar wind appears quite similar to that of a comet. The outgassing of matter from Phobos and Deimos is also suggested by plasma observations in the wake/tail of the Martian satellites." Ip and Banaszkewicz (1990) state that "The presence of a gas ring with a neutral number density substantially higher than the hot atomic oxygen background is possible only if the Phobos moon maintains a certain level of outgassing with $Q > 10^{23}$ molecules/sec." Further evidence for gas and/or dust emission from Phobos is given in Dubinin et al. (1991). Although consistent with a water flux, the measurements could also be explained by emission of CO_2, MgO, O_2, or Si ions. Of these only CO_2 would seem a reasonable alternative to H_2O.

The flux of 3 g s^{-1} predicted in Fanale and Salvail (1990) corresponds to 1.0×10^{23} mol s^{-1}, exactly equal to the Ip and Banaszkewicz lower limit. [Dubinin et al. interpret their data as suggesting 30 to 100 times the flux predicted by Fanale and Salvail. This seems impossibly large, because it would completly deplete Phobos of any volatile in a time shorter that the age of the solar system.] While this agreement offers some feeble indirect evidence that Phobos does in fact contain ice, there is no way to establish for certain what caused the magnetospheric anomalies noted by Phobos 2. The search for deep ice and escaping water vapor on Phobos must remain the top priority for any future Mars/Phobos mission.

II. MINERALOGY AND METEORITICAL ANALOGS OF PHOBOS AND DEIMOS

From the data outlined above, the reflectance spectrum of both Phobos and Deimos appears to combine a steep increase in the ultraviolet, a flat section in the visible, and a steep red slope in the near-infrared. The details of the transitions between these three wavelength regions are obscure, and producing a continuous spectrum from the various nonoverlapping data sets requires a certain amount of judgment as to the relative scaling of data taken under different photometric conditions. For example, in Pang et al. (1978) and Pang et al. (1980), slightly different offsets between the Mariner 9 ultraviolet spectrum and the visible photometry were used. This shifts the wavelength of the beginning of the ultraviolet absorption feature, which is one of the most important criteria in comparing spectra of dark meteorites and asteroids. However, it seems clear that both Phobos and Deimos may be described in terms of asteroids as having C-class spectra out to about 1 μm, and D-class spectra from there to 2.5 μm. This is a very rare (possibly nonexistent) type of spectrum among observed asteroids. However, a look at the meteorite spectroscopy literature provides a straightforward explanation. Johnson and Fanale (1973) obtained laboratory spectra of several carbonaceous chondrite

meteorites. As the nature of asteroidal regoliths was poorly known at the time, these authors measured several particle size fractions of each meteorite. In the case of CM2 chondrites (called C2 chondrites in 1973), changing particle size has a strong effect on the reflection spectrum. As shown in Fig. 5, the smallest size fractions show a steep red slope in the infrared combined with the "normal" C-class spectrum in the ultraviolet and visible wavelengths. This is the type of spectrum observed in both Phobos and Deimos. Most C-class asteroids have flat infrared spectra corresponding to the larger particle size fractions measured by Johnson and Fanale (1973). Indeed, more recent studies of carbonaceous chondrite spectra have usually measured only large size fractions for this reason (see, e.g., Gaffey 1976). But in the case of Phobos and Deimos, ejecta recycling due to their unique location inside the gravity well of Mars is likely to produce a more finely pulverized regolith than on asteroids.

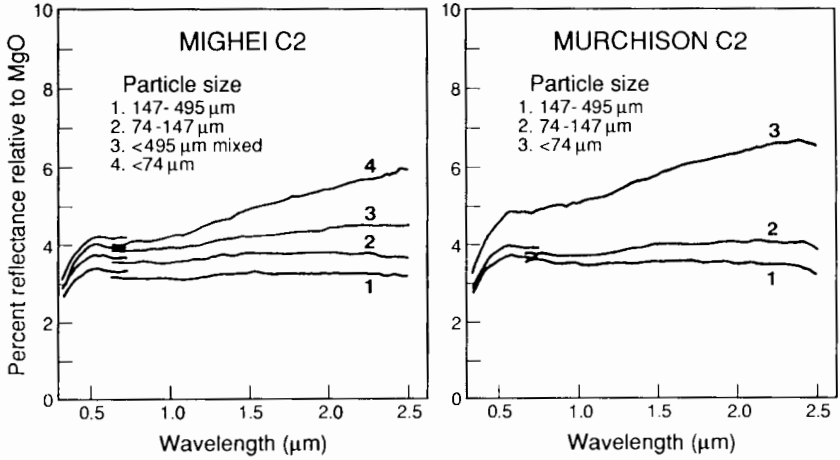


Figure 5. Laboratory reflection spectra of various size fractions of powdered CM2 carbonaceous chondrites Mighei and Murchison (figure from Johnson and Fanale 1973). The most finely pulverized samples (all particles <74 μm in diameter) are good spectral matches for the unique combination of C-like visible spectra and D-like infrared spectra found in the Mars satellites.

Therefore, it appears that the most likely composition of Phobos and Deimos is something that might be described by the imaginary meteorite type "CM3," because it would closely resemble the unmetamorphosed parent material of the familiar CM2 meteorites. This material would combine anhydrous silicates like olivine and pyroxene with some unknown amount of water ice. Because Phobos and Deimos did not undergo strong post-accretional heating, the water ice did not melt and form the liquid water which was responsible for hydration of the silicates in CM2 and CI1 chondrites (and also many asteroids). Instead this ice component has gradually sublimed away from

the upper layers of the satellites, leaving a large component of pore space. The outgassing continues at a low level today as indicated by the Phobos 2 magnetospheric instruments. In this model, the deep interior of Phobos still contains water in the form of "chondritic permafrost" even though the surface is completely anhydrous.

III. ORIGIN OF PHOBOS AND DEIMOS

In summary, the available spectral data on Phobos and Deimos, plus the extraordinarly low density of Phobos, suggests that these objects are very similar to the dark carbonaceaous asteroids. This is consistent with the common suggestion that Phobos and Deimos are asteroids (or fragments of a single asteroidal parent body) captured into orbit around Mars by gas drag in the very early days of the solar system, after which strong tidal evolution has brought them to their present orbital positions (Pollack et al. 1979; Hunten 1979; Zharkov et al. 1984; Sasaki 1990*a*, *b*). This same mechanism has also been often proposed to account for the two families of small satellites orbiting Jupiter, and the retrograde Saturn satellite Phoebe.

There are some problems with this neatly unified scenario, however. Phobos and Deimos are in circular, low-inclination orbits tightly bound to Mars, quite unlike the irregular, loosely bound orbits of the other supposedly captured satellites. A large amount of post-capture orbital evolution due to tidal forces is necessary to produce the current orbits of Phobos and Deimos, and modeling of this process suggests that this process may be very difficult (Burns 1992).

Furthermore, the interpretation of the satellites' composition as essentially CM-like limits their original formation location to the region beyond 2.5 AU from the Sun. In addition, all of the captured satellites of Jupiter and Saturn seem to be C-class objects originally derived from this same general region. This poses some difficult problems. The distinct stratigraphy preserved in the asteroid belt seemingly indicates that there has been little diffusion of orbits since the mineralogical differences between asteroid classes were established. So why did objects now orbiting Mars originate in the outer belt? How did they cross the intervening zones dominated by S, M and E asteroids? Similarly, how did the former C-class asteroids now orbiting Jupiter and Saturn cross the zones dominated by P and D objects? Why is there not a single captured satellite from another asteroid class? Some thought has been given to these questions (see Hartmann 1987), but they are still unanswered in detail.

Also, if the asteroidal heat source is in fact electromagnetic heating as suggested by the asteroid data (Bell et al. 1989*b*), capture must have occurred after the T Tauri phase of the Sun. At this time the residual nebular gas required in most capture scenarios were presumably dispersed by the same intense solar wind that melted the inner asteroid belt. Thus the capture origin

of Phobos and Deimos (or their common parent body) does not fit simply into the currently popular picture of asteroid formation.

IV. IMPLICATIONS FOR RESOURCE UTILIZATION

The evidence reviewed above suggests that while the common conception of the Martian satellites as carbonaceous-chondrite-like bodies is generally correct, the vision of easily extracting large amounts of water from the surface regolith of these bodies is exaggerated. The 1988 IRTF observations of Deimos and the similar work done by Phobos 2 seem to rule out any significant amount of water in the surface of either satellite. However, there are indirect and theoretical indications suggesting that they may contain large amounts of water ice in their deep interiors. The most important lines of future research are:

1. Send another spacecraft to Phobos equipped to directly detect the water vapor suspected to be leaking out of Phobos and map the depth to the top of the permafrost layer at many points over the surface.
2. Make improved models of the behavior of ice inside Phobos and Deimos over time.
3. Using the data obtained from (1) and (2), work out realistic methods for extracting water from the interior of Phobos at the regions where the permafrost layer is shallowest (probably at the poles).

Of these tasks, (1) is by far the most important and difficult. It is impossible to make realistic plans for exploiting the resource potential of the Martian satellites without another unmanned mission, properly equipped to answer the key questions remaining. Any serious program for human exploration and exploitation of space should include a search for water on both Mars and Phobos as its earliest objective.

Note Added in Proof

The theoretical model for ice retention and water vaporization on Phobos (Fanale and Salvail 1989*b*,1990) had earlier been applied to the main-belt asteroid Ceres (Fanale and Salvail 1989*a*). Recently A'Hearn and Feldman (1992) have reported observations with the International Ultraviolet Explorer spacecraft which they interpret as indicating a water flux from Ceres virtually identical to that predicted by the Fanale and Salvail (1989*a*) model. This observations supports the accuracy of the model in predicting the history of ice in Phobos if it in fact included an ice component at the time it entered its current orbit.

REFERENCES

A'Hearn, M. F., and Feldman, P. 1992. Water vaporization on Ceres. *Icarus* 98:54–60.

Avanesov, G., Zhukov, B., Ziman, Ya., Kostenko, V., Kuzmin, A., Murav'ev, V., Fedotov, V., Bonev, B., Mishev, D., Petkov, D., Krumov, A., Simeonov, S., Boycheva, V., Uzunnov, Yu., Weide, G.-G., Halmann, D., Possel, W., Head, J., Murchie, S., Schkuratov, Yu. G., Berghangel, R., Danz, M., Mangoldt, T., Pihan, U., Weidlich, U., Lumme, K., Muinonen, K., Peltoneimi, J., Duxbury, T., Murray, B., Herkenhoff, K., Fanale, F., Irvine, W., and Smith, B. 1991. Results of TV imaging of Phobos (Experiment VSK-FREGAT). *Planet. Space Sci.* 39:281–295.

Barth, C. A., Hord, C. W., Stewart, A. I., and Lane, A. L. 1972. Mariner 9 ultraviolet spectrometer experiment: Initial results. *Science* 175:309–312.

Bell, J. F., Piscitelli, J. R., and Lebofsky, L. A. 1989*a*. Deimos: Hydration state from infrared spectroscopy. *Lunar Planet. Sci.* XX:58–59 (abstract).

Bell, J. F., Davis, D. R., Hartmann, W. K., and Gaffey, M. J. 1989*b*. Asteroids: The big picture. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 921–947.

Bibring, J.-P., Combes, M., Langevin, Y., Cara, C., Drossart, P., Encrenaz, Th., Erard, E., Forni, O., Gondet, B., Ksanfomality, L., Lellouch, E., Masson, Ph., Moroz, V., Rocard, F., Rosenqvist, J., Sotin, C., and Soufflot, A. 1989. Results from the ISM experiment. *Nature* 341:591–593.

Bibring, J.-P., Langevin, Y., Moroz, V. I., Ksanfomality, L. V., Grigoryev, A. V., Khatuntsev, I. V., Nikolsky, Yu. V., Zharkov, A. V., and Combes, M. 1991. Composite KRFM-ISM spectrum of Phobos (0.315–3.1 μm). *Lunar Planet. Sci.* XXII:99–100 (abstract).

Burns, J. A. 1992. Contradictory clues as to the origin of the Martian moons. In *Mars*, eds. H. H. Kieffer, B. M. Jakosky, C. W. Snyder and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 1283–1301.

Cole, D. M., and Cox, D. W. 1964. *Islands in Space: The Challenge of the Planetoids* (Philadelphia: Chilton Books).

Cruikshank, D. P., Bartholemew, M. J., and Roush, T. L. 1991. Composition of the surface of Phobos: Results from the PHOBOS mission. *Lunar Planet. Sci.* XXII:263–264 (abstract).

Dubinin, E. M., Lundin, R., Pissarenko, N. F., Barabash, S. V., Zakaharov, A. V., Koskinen, H., Schwingenshuh, K., and Yeroshenko, Ye. G. 1990. Indirect evidence for a gas/dust torus along the Phobos orbit. *Geophys. Res. Lett.* 17:861–864.

Dubinin, E. M., Pissarenko, N. F., Bararbash, S. V., Zakaharov, A. V., Lundin, R., Pellinen, R., Schwengenshuh, K., and Yeroshenko, Ye. G. 1991. Plasma and magnetic fields associated with Phobos and Deimos tori. *Planet. Space Sci.* 39:113–121.

Fanale, F. P., and Salvail, J. R. 1989*a*. The water regime of asteroid (1) Ceres. *Icarus* 82:97–110.

Fanale, F. P., and Salvail, J. R. 1989*b*. Loss of water from Phobos. *Geophys. Res. Lett.* 16:287–290.

Fanale, F. P., and Salvail, J. R. 1990. Evolution of the water regime of Phobos. *Icarus* 88:380–395.

Feierberg, M. A., Lebofsky, L. A., and Tholen, D. J. 1985. The nature of C-class asteroids from 3-micron spectrophotometry. *Icarus* 63:183–191.

French, L. M., Veverka, J., and Thomas, P. 1988. Brighter material on Deimos: a particle-size effect in a carbonaceous material? *Icarus* 75:127–132.

Gaffey, M. J. 1976. Spectral reflectance characteristics of the meteorite classes. *J. Geophys. Res.* 81:905–920.

Gaffey, M. J., Bell, J. F., and Cruikshank, D. P. 1989. Reflectance spectroscopy and asteroid surface mineralogy. In *Asteroids II*, eds. R. P. Binzel, T. Gehrels and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 98–127.

Harris, D. L. 1961. Photometry and colorimetry of planets and satellites. In *Planets and Satellites*, eds. G. P. Kuiper and B. Middlehurst (Chicago: Univ. of Chicago Press), pp. 272–342.

Hartmann, W. K. 1987. A satellite-asteroid mystery and a possible early flux of scattered C-class asteroids. *Icarus* 71:57–68.

Hunten, D. M. 1979. Capture of Phobos and Deimos by protoatmospheric drag. *Icarus* 37:113–123.

Ip., W.-H., and Banaszkiewicz, M. 1990. On the gas/dust tori of Phobos and Deimos. *Geophys. Res. Lett.* 17:857–860.

Johnson, T., and Fanale, F. 1973. Optical properties of carbonaceous chondrites and their relationship to asteroids. *J. Geophys. Res.* 78:8507–8518.

Jones, T. D., Lebofsky, L. A., and Lewis, J. S. 1989. Mid-IR reflectance spectra of carbonaceous chondrites: Applications to low-albedo asteroids. *Meteoritics* 24:282.

Jones, T. D., Lebofsky, L. A., Lewis, J. S., and Marley, M. S. 1990. The composition and origin of the C, P, and D asteroids: Water as a tracer of thermal evolution in the outer belt. *Icarus* 88:172–192.

Ksanfomality, L., Murchie, S., Britt, D., Duxbury, T., Fisher, P., Goroshkova, N., Head, J., Kuhrt, E., Moroz, V., Murray, B., Nikitin, G., Petrova, E., Pieters, C., Soufflot, A., Zharkov, A., and Zhukov, B. 1991. Phobos: Spectroradiometry between 0.3 and 0.6 μm and IR-radiometry. *Planet. Space Sci.* 39:311–326.

Langevin, Y., Bibring, J.-P., Gondet, B., and Cruikshank, D. P. 1991. ISM observations of the spectral characteristics of Phobos in the near infrared. *Lunar Planet. Sci.* XXII:781–782 (abstract).

Lebofsky, L. A. 1980. Infrared reflection spectra of asteroids: A search for water of hydration. *Astron. J.* 85:573–585.

Lebofsky, L. A., Feierberg, M. A., Tokunaga, A. T., Larson, H. P., and Johnson, J. R. 1981. The 1.7- to 4.2-micron spectrum of asteroid 1 Ceres: Evidence for structural water in clay minerals. *Icarus* 48:453–459.

Lucey, P. G., Bell, J. F., and Piscitelli, J. R. 1989. High spectral resolution spectroscopy of the Martian moons. *Lunar Planet. Sci.* XX:598–599 (abstract).

Pang, K. D., Pollack, J. B., Veverka, J., Lane, A. L., and Ajello, J. M. 1978. The composition of Phobos: Evidence for carbonaceous chondrite surface from spectral analysis. *Science* 199:64–66.

Pang, K. D., Rhoads, J. W., Lane, A. L., and Ajello, J. M. 1980. Spectral evidence for a carbonaceous chondrite surface composition on Deimos. *Nature* 283:277–278.

Pollack, J. B., Veverka, J., Pang, K., Colburn, D., Lane, A. L., and Ajello, J. M. 1978. Multicolor observations of Phobos with the Viking lander cameras: Evidence for a carbonaceous chondrite composition. *Science* 199:66–69.

Pollack, J. B., Burns, J. A., and Tauber, M. E. 1979. Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* 37:587–611.

Prinn, R. G., and Fegley, B., Jr. 1987. The atmospheres of Venus, Earth, and Mars: A critical comparison. *Ann. Rev. Earth Planet. Sci.* 15:171–212.

Prinn, R. G., and Fegley, B., Jr. 1988. Solar nebular chemistry: Origin of planetary, satellite, and cometary volatiles. In *Origin and Evolution of Planetary and Satellite Atmospheres*, eds. S. K. Atreya, J. B. Pollack and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 78–136.

Sasaki, S. 1990*a*. Origin of Phobos—Aerodynamic drag capture by the primary atmosphere of Mars. *Lunar Planet. Sci.* XXI:1069–1070 (abstract).

Sasaki, S. 1990*b*. Orbital evolution of a captured satellite within an extended pri-

mary atmosphere—a case for Phobos. *Bull. Amer. Astron. Soc.* 22:1117–1118 (abstract).

Zellner, B. H. 1972. Minor planets and related objects. VIII. Deimos. *Astron. J.* 77:183–185.

Zellner, B. H., and Capen, R. C. 1974. Photometric properties of the martian satellites. *Icarus* 23:437–444.

Zharkov, V. N., Kosenko, A. V., and Maeva, S. V. 1984. Structure and origin of the satellites of Mars. *Astron. Vestnik* 18(2):88–89.

Zolensky, M. E., and McSween, H. Y. 1988. Aqueous alteration in meteorites. In *Meteorites and the Early Solar System*, eds. J. F. Kerridge and M. S. Matthews (Tucson: Univ. of Arizona Press), pp. 114–143.

MISSION AND TRANSPORTATION APPLICATIONS OF IN-SITU PROPELLANT PRODUCTION IN THE MARS SYSTEM

BENTON C. CLARK
Martin Marietta Civil Space & Communications

Mars is endowed with an abundance of compounds containing most of the elements needed to manufacture propellants for transportation, whether by rocket, rover, or other means. Selection of the most suitable propellants involves a careful engineering evaluation, based upon the application (especially the energy requirements thereof) and the complexities that are involved in extracting the indigenous compounds and converting them to more useful ones.

The earliest flyby spacecraft of Mars returned images that were strikingly like the Moon, leaving the impression of a cold, dry, and barren planet in spite of the fact that theoretical arguments had been made (Lewis 1972) that Mars should be volatile rich. Following the breakthroughs of the Mariner 9 and Viking missions, it was realized that the red planet indeed possesses light element resources far beyond that of the Moon or Mercury. One of the first proposed uses of such resources was the manufacture of rocket propellants for the return flight to Earth (Ash et al. 1978; Clark 1979). Resource utilization for propellant production is obviously of great importance because it is by far the largest single commodity which must be delivered to Mars on any round-trip mission. Its elimination could either reduce the requirements on mass launched from Earth, or free up cargo mass for delivery of high-technology equipment and instrumentation rather than transportation of bulk chemicals (propellants) of very low technological content. The amount of "leverage" on the mission mass achieved by using planetary propellants is strongly dependent upon the specific mission profile. Mars ascent vehicles can have propellant-to-dry mass ratios of greater than 5:1. The total leveraging by using *in-situ* propellants is typically much less, however, because of the masses of hardware that must be transported one-way to Mars (e.g., lander habitats, entry systems, communications satellites, and other mass that remains in Mars orbit). Because any amount of leveraging has the potential for significant cost payback, especially when the investment is amortized over the longer term and all life-cycle costs are taken into account, the utilization of Mars propellants has received considerable attention and support (see, e.g., Stancati et al. 1979; Meyer 1981; Ash et al. 1982; Clark 1985; French 1985,1989; O'Leary 1985; Clapp and Scardera 1987; Frisbee 1987; Lewis and Lewis

1987; Cordell 1989,1990; Ramohalli et al. 1989; Zubrin 1989; Linne 1991; Zuppero, in preparation). Relative value, cost, and risk of utilizing propellant production at the planet must be assessed so that a decision can be made at the highest level, i.e., in the context of a total systems analysis. Nonetheless, it is clear that resource production has a promising long-term potential. Besides a more thorough understanding of Martian resources, the work that remains undone is to obtain clearer definition of the engineering systems and investments needed to produce such propellants and selection of optimum approaches for various mission scenarios.

I. PROPELLANT AVAILABILITY

A survey of liquid chemical rocket propellants with good performance, good handling properties, or both, leads to the conclusion that nearly all candidate chemicals are composed of one or more of the following six elements: H, C, N, O, F, or Cl. As a virtual storehouse for many "volatiles," Mars is already known to contain all of these elements at its surface, with the exception of fluorine.

On Earth, fluorine is mined from deposits containing fluorite mineral (and less often as a byproduct of phosphate rock processing). Finding such ore bodies on Mars is not likely during the early exploration phases. The F content of SNC meteorites is only about 40 ppm. The halogens Cl and Br were found in unexpectedly high abundance in the martian soil, however (Clark et al. 1982), and fluorine could not have been measured by the instrument flown. On Earth, the chief repository of halogen elements is the ocean, where the F/Cl ratio is 1 in 20,000. If this ratio holds for Mars, the fluorine content of the soil would be <1 ppm, but the low sea-water concentrations are probably due to the poor solubility of the fluorides compared to the equivalent chlorides. Furthermore, the cosmic abundance of F is about 1/10 that of Cl (wt/wt) and in SNC meteorites, the F concentration can be equal to or higher than Cl. Thus, although probably unlikely, there is some reason to presume that F could be as high as 0.5% in the Martian soil. Even so, F-containing oxidizers have been found difficult to apply beneficially to large propulsion systems on Earth and it therefore remains difficult to forecast their use on Mars.

All of the other six key propellant elements are abundant in the soil or atmosphere, or both (see Chapter by Stoker et al.). For example, H atoms and O atoms can be obtained from H_2O, which occurs as vapor in the atmosphere, as condensate at the polar caps, and in various forms as an adsorbate or chemically incorporated into soil minerals. The CO_2, CO and N_2 in the atmosphere provide an enormous mass of dependable resource; carbonates and nitrates may also be present in the soil. One or more chlorine compounds are present in all Martian soils sampled to date. Thus, from the standpoint of plentiful starting resources, propellant production would seem technically quite promising for Mars.

II. PRINCIPLES OF APPLICATION

Suitability and utility of any particular rocket propellant must be evaluated simultaneously with respect to its performance, its producibility, and the logistics factors. The performance requirement is intimately tied to the application for which the propellant is intended.

Performance for rocket flight is measured in terms of specific impulse, I_{sp}, the amount of propulsive impulse (force integrated over a period of time) which can be produced by a unit mass of the selected propellants (Table I). This parameter is useful because a lesser quantity of propellant is required to achieve the same velocity increment, ΔV, during the rocket burn (see, e.g., Hill and Peterson 1970). When compared to the total ΔV required, this parameter also indicates whether a single stage or multiply-staged rocket will be needed.

TABLE I

Chemical Rocket Propellant Candidates—Performance as Measured by Specific Impulse (I_{sp})

Propellant Components		Typical
Fuel	Oxidizer	I_{sp} (m s^{-1})[a]
Hydrogen	Oxygen	4500
Methane	Oxygen	3700
Monomethyl Hydrazine (MMH)	Oxygen	3700
Methanol	Oxygen	3050
Carbon monoxide	Oxygen	2540
RP-1	Hydrogen peroxide	2700
Monomethyl Hydrazine (MMH)	Nitrogen Tetroxide (NTO)	3330
	same, pressure fed	3150
Hydrazine	Hydrogen Peroxide	3000
UDMH hydrazine	NTO	3050
Hydrazine monopropellant		2160
Hydrogen peroxide monopropellant		1600
Cold gas (nitrogen)		780
Hydrogen monopropellant (solid core nuclear rocket)		8300

[a] Same as N-s kg^{-1}. To convert to the more conventional units of "seconds" (actually, lbf-s/lbm), multiply by 0.102.

To assess these requirements, we must examine the ΔVs in the Mars system. Going to Mars can be accomplished by either the minimum energy transfer strategy, called the conjunction-class trajectory, or by the strategy of

minimization of total mission time, resulting in the opposition-class trajectory. The latter must accomplish at least one leg of flight by passing inside the orbit of Earth, typically approaching the orbit of Venus and sometimes using this to advantage to obtain a Venus gravity assist to reduce propulsion requirements. For conjunction class, all transfers occur outside Earth's heliocentric orbit and ΔVs are significantly smaller, but the round-trip times are 2.5 to 3 yr, compared to the opposition cases of 1.2 to 2 yr.

Mars propellant would most efficiently be used to ascend back into orbit, or possibly even directly back to Earth. The ΔVs for ascending to Mars orbit from the Martian surface and the various return-to-Earth options span the range ~1000 to 8000 m s^{-1}, as shown in Table II. Note that the choice of Mars orbit is important, i.e., whether the orbit is a circular low Mars orbit (LMO), such as used by the Mariner 9 and Mars Observer spacecraft; a high circular, equatorial orbit (such as for Phobos, Deimos, or 24.6 hr areostationary orbits); or a highly eccentric Mars orbit (HEMO) such as the 24.6 hr synchronous elliptical orbits of the Viking missions.

TABLE II
Major Propulsion Requirements (ΔV in m s^{-1}) for Five Different Mars Orbits[a]

	LMO	HEMO	Phobos	Deimos	Areostationary Orbit[b]
Descent to landing[c]	464	364	929	1022	1025
Ascent to Mars orbit[d]	4200	5400	5300	5850	5800
Trans-Earth inject (TEI)					
Cn (optimum)	2162	984	—	—	—
Cn, 180 d transfer	3164	1966	—	—	—
Opposition	5167	3942	—	—	—
Sprint	7726	6481	—	—	—

[a] No inclination changes included (see Table I). LMO = 500 km circular; HEMO = 250 km × 1 sol (33,838 km). Return to Earth from circular, equatorial orbits not generally an option because of need to adjust inclination to match outbound declination requirements leads to transfer to a highly eccentric orbit, such as the HEMO adopted here.
[b] Circular equatorial orbit with 24.623 hr period (17,045 km altitude)
[c] ΔV for periapsis lowering to 10 km altitude, plus 350 m s^{-1} for hovering, cross-range, and terminal descent.
[d] Ascent from Mars surface to Mars orbit.
[e] Example C3's of 8, 20, 50, and 100 km^2 s^{-2} for the four cases, respectively. Cn = conjunction class trajectory. At C3 = 20, the majority of Cn interplanetary transfers occur in <180 d, although some take up to 15% longer.

A direct return to Earth from the Martian surface can be taken as a summation of surface → elliptic orbit and a trans-Earth injection (TEI) from this orbit, except that launching into an orbital inclination that is co-planar with a specific departure asymptote can entail a small additional ΔV. Likewise,

the LMO orbit must be co-planar with the asymptote. In addition for the elliptical orbit, the argument of periapse must be uniquely positioned.

From LMO, the opportunities (launch windows) for the TEI burn to return to Earth are at most three orbits per day, and the window occurs only a few minutes over a specific part of each orbit. The number of successive orbits that can accommodate the launch will vary with the specifics of the LMO. Regression of the orbital plane will allow the ascent only during a 2 to 3 day period. For opposition-class trajectories, with short stay times at Mars, these orbits are highly constrained. Generally, there may be only a few such orbits before the amount of regression has become unacceptable. Launch windows may therefore be extremely limited by usual standards unless multiple burns and ΔV penalties for changes in inclination and argument of periapsis are made prior to departure.

The HEMO elliptical orbit is less constraining, with a launch window of a few minutes once each sol but lasting for a period of several weeks. A TEI from the Martian surface may have a window of as much as an hour or more each day, over a period of a few weeks. On the other hand, TEI from HEMO requires the least mass initially transported to Mars (mass at Mars arrival, MAMA) whereas LMO requires considerably more and a direct return from the Martian surface requires the largest MAMA. The general reason for this is that each successive scenario transports the heavy return vehicle deeper into the Mars gravity potential well. For the same reason, however, the use of Martian propellant will have even a greater relative benefit for these scenarios.

The strategy of whether to use a single rocket or a multi-staged rocket depends upon the I_{sp} of the propellant relative to the ΔV to be accomplished. Generally speaking, once the ΔV exceeds the I_{sp}, it will be advantageous to split the rocket into two or more stages. This rule is affected, of course, by payload mass relative to the dry mass of the propulsion system, as well as by considerations such as higher expense and lower reliability (and hence decreased safety) generally resulting from multiple staging. For example, using a pump-fed UDMH/NTO propellant pair for ascent to HEMO, a single-stage rocket was preferred by Geels (1990) for reasons of simplicity and safety even though the $\Delta V/I_{sp}$ ratio to reach HEMO was 1.6.

Orbital transfer ΔVs at Mars are tabulated in Table III. It is particularly noteworthy that transfers to and from Phobos and Deimos require quite significant ΔVs, especially when major inclination (i) changes must be made. For example, a roundtrip from HEMO at = 45° to either martian satellite is a ΔV penalty on the order of 2 km s^{-1}, i.e., equal to or greater than the ΔV to return to Earth in some cases. Large inclination changes will be generally required to visit Phobos or Deimos for several reasons. First, when a vehicle arrives at Mars from Earth, it will do so at some declination, which can be as high as 48°. Because it is not possible to avoid a ΔV penalty unless the inclination of the orbital plane is at least as large as the incoming declination, this factor will be important for all but the very rare transfer opportunities when arrival declinations are fortuitously low. Second, the vehicle may for

TABLE III
Interorbit Transfers within the Mars System, ΔV (m s^{-1})

			HEMO		LMO	
	Areostat	Deimos	Phobos	i=45°	i=0°	i=45°
LMO (i=0°)	1615	1661	1130	1509	1163	2543
LMO (i=45°)	2156	2145	2302	1163	1509	
HEMO (i=0°)	648	636	793	346		
HEMO (i=45°)	994	982	1139			
Phobos	666	749				
Deimos	96					

[a] Assumes optimum interorbit transfers, using multiburn via HEMO when appropriate (e.g., from the natural satellites to a LMO at 45° inclination). Taking advantage of an aerobrake instead of propulsion for lowering apoapsis allows reductions in certain propulsive ΔVs: HEMO⟶Phobos becomes ΔV = 560 m s^{-1} total; Deimos⟶LMO (0°) becomes ΔV = 740 m/s^{-1}; and Phobos⟶LMO (0°) is ΔV = 640 m s^{-1}.

many other reasons require injection into a relatively highly inclined orbit, even polar. These other reasons can include planetary latitude viewing requirements, landing site accessibility, communications networking, lighting and thermal constraints, etc. Third, when returning to Earth, the vehicle must return from its near-equatorial orbit to an inclined orbit in order to match the declination required for the outgoing launch asymptote. Incoming and outgoing declinations are generally quite different. For comparison, inclinations of both 0° and 45° are included in Table III.

III. PRODUCTION PATHWAYS AND MISSION UTILIZATION

The producibility of Martian propellants is dependent, first of all, upon the availability of the necessary ingredients. Although trace quantities of a few natural propellant chemicals are known or predicted to exist at the Martian surface (e.g., CO, O_2, H_2, H_2O_2), direct extraction to accumulate large quantities may be more difficult than chemical conversion from other, more-abundant molecules.

The only chemical form in which H atoms are reasonably abundant on Mars is as H_2O, as organic compounds are apparently absent. Water may be "mined" from either the atmosphere (Meyer and McKay 1984; Clapp 1985) or from the soil (Clark 1984) at energy costs ranging from <1 to >100 kW-hr kg^{-1}, where the energy expended is mainly electrical/electromechanical, for atmospheric extraction, or thermal energy for release from the regolith fine-grained soils.

Direct electrolysis of H_2O produces the "H/O" propellant pair of H_2 and O_2, as shown in reaction (1).

$$2H_2O \longrightarrow 2H_2 + O_2 \qquad (1)$$

Electrolysis is quite energy-expensive (about 5.3 kWe-hr kg^{-1} of H_2O); additionally, liquefaction of the gases will require considerable additional energy—at least 1.6 kW-hr kg^{-1} for H_2 (Ash et al. 1978). These processes of manufacture of H/O propellant are among the most straightforward that can be envisioned, especially as H_2O electrolysis has already received much development effort for space applications because of its utility in generating breathing oxygen in life support systems. Liquefaction techniques could borrow from commercial practice or incorporate advanced techniques, perhaps capitalizing on the lower ambient atmospheric temperatures on Mars, especially during nighttime. As advantageous as H/O might be for Mars, the difficult logistics of LH2 (storing and handling of a hard cryogen) leads to interest in other propellant combinations, as pointed out in the landmark paper on Mars *in-situ* propellant production by Ash et al. (1978).

The next best-performing propellant combination is CH_4/O_2 (methane/-LOX, or "Me/O"). In the original Ash et al. concept, H_2O is extracted from the soil (by heating), electrolzed, and the hydrogen reacted with atmospheric CO_2 to produce methane and more water:

$$CO_2 + 4H_2 \longrightarrow CH_4 + 2H_2O. \tag{2}$$

The advantages of using methane when compared to hydrogen are its greater density and storability at higher temperatures. The leverage on reduction in initial mass in low Earth orbit (IMLEO) by using locally produced propellant was estimated by Ash and coworkers as a factor of 2 for the conjunction class trajectory they assumed.

A major concern with this Me/O scenario has been the presumed variations in the availability of H_2O from site to site on Mars, with the success of such an approach being dependent upon the peculiarities of each particular landing location. However, it should be pointed out that these authors assumed only 2% of H_2O released per unit mass of heated soil, and showed that if waste thermal energy from power supply inefficiencies and recovered heat from spent soil can be exploited, then no additional power drain for soil processing would be required other than acquisition.

In the approach by French (1985,1989), it was proposed that the propellant combination carbon monoxide and oxygen, CO/O_2, be utilized in order to avoid the problem of locating water on Mars. Both fuel and oxidizer can be produced from a single reaction:

$$2CO_2 \longrightarrow 2CO + O_2. \tag{3}$$

Thermal decomposition, followed by isolation of the O_2 using a zirconia electrolytic membrane at high temperatures, has been suggested by Frisbee (1987) and pursued more recently in the laboratory by Ramohalli et al. (1991). The performance of CO/O_2 as a rocket propellant pair has recently been verified by Linne (1991). The chief drawback of this combination is the quite

low performance (Table I), with a $\Delta V / I_{sp}$ ratio of more than 1.6 required just for achieving LMO. Stancati et al. (1979), Ash et al. (1982) and Frisbee et al. (1987) proposed transportation of CH_4 from Earth, with *in-situ* production of O_2 from CO_2. Ash et al. (1982) also considered transporting H_2 from Earth, but stated that the trade-off of improved performance vs the weight to store liquid hydrogen using active refrigeration led them not to consider this approach viable. However, Zubrin et al. (1991) proposed transporting LH_2 to Mars but reacting it immediately upon landing with CO_2 to produce the more easily stored CH_4. The amount of LH_2 required is only 5.5% of the total final mass of Me/O propellants produced. The sequence is to react H_2 with Martian atmospheric CO_2 to produce methane and H_2O [reaction (2)] and then electrolyze the H_2O according to reaction (1) to store the O_2 and recycle the hydrogen to produce more methane. Because the amount of O_2 generated is less than that optimum for Me/O propellant (taken as a 1:3.5 mixing ratio, wt/wt), the proposal is to produce additional O_2 via raction (3). An alternate approach would be to bring more LH_2 and produce sufficient H_2O for all required O_2 synthesis and an excess amount of methane (stored for use later as a propellant or as a feedstock for any number of organic synthesis or biochemical processes). This would require 9.7% of the Me/O final propellant mass be imported as LH_2, but 1.7 times the mass of the imported hydrogen would be available as surplus methane. If methane were imported and all oxygen generated locally by reaction (3), as previous authors have suggested, the mass of imported fuel would have to be 22.2% of the final propellant mass needed.

Zubrin et al. (1991) provide an entire manned mission scenario utilizing Me/O via transport of LH_2 to Mars, followed by reactions (1)–(3). They and Meyer (1981) also suggest an alternative to reaction (3) to extract O_2 from CO_2, via the reaction of H_2 with CO_2 under conditions suitable to produce CO and H_2O—followed by reaction (1) to produce O_2. Also considered as a potential method is recovery of some H_2 by pyrolysis of CH_4, to again produce more H_2O, followed by (1) to produce more O_2.

The widely used storable propellants based upon hydrazines and oxidizers such as nitrogen tetroxide have been suggested (Clark 1978; Meyer 1981; LeCompte and Stets 1991). Although all the required ingredients (N, H, O, C) are present on Mars, the manufacture of these compounds can be relatively complex and their specific impulses are lower than those obtainable with the cryogenic propellants. However, these materials should not be necessarily ruled out at this time just because of the lack of a serious search for space-applicable manufacturing processes. Their great advantage, of course, is that they have no need for active cooling in order to store them. Indeed, to use them, it would be necessary in most cases to unfreeze them.

A number of propellant fuels are being considered for lunar *in-situ* propellant production, including metallics, silane, and other compounds that are attractive because of the greater availability of their starting materials relative to the low quantities of H and C in lunar materials. Any or all could also be

manufactured on Mars. See the Chapter by Repic et al. for a discussion of these materials and their processes, including the extraction of oxygen from surface minerals.

IV. ROVER FUELS

Unmanned rovers on Mars can be powered by nuclear energy sources, such as radioisotope thermoelectric generators (RTGs) or dynamic isotope power systems (DIPS). However, both RTG and DIPS power sources are unlikely to receive serious consideration for manned sorties because of the extremely high penalties for radiation shield mass that would have to be incorporated to reduce the radiation hazard to acceptable levels. Electric vehicles for Mars also would be intrinsically heavy because of the very large battery complement needed to achieve a reasonable range. Solar cell power arrays would have to be many times larger in area than the footprint of the rover, would be susceptible to contamination by dust kicked up by the rover wheels, could be susceptible to damage from wind gusts, and would entail complicated Sun-tracking hardware because of the constantly changing attitude and orientation of the rover itself. Chemical fuels, on the other hand, being energy-dense and independent of the Sun, are a most likely choice for rover exploration by humans on Mars, especially for the pressurized, long-range rovers that will be needed to accomplish exploration on the regional scale (hundreds of km).

A number of rover fuels could be manufactured on Mars. For example, hydrogen peroxide (Clark 1978), methane (Meyer 1981; Zubrin et al. 1991), methanol (Meyer 1981; Clark et al. 1991), carbon monoxide (Clapp and Scardera 1987), and carbon disulfide (Clark 1984) have been proposed. Each has its own advantages and disadvantages. Although the mass leveraging to the mission is less, it is clear that for a vigorous exploration or a permanent base, the ability to manufacture transportation fuels on-site will be of great advantage. It is often suggested that a first application of *in-situ* propellant production on Mars might be for ground transportation rather than for rocket flight. It would be convenient if the rover propellants were the same as planned for rocket use, in order to demonstrate the production on a smaller scale and also because the rovers could then exploit the availability of large quantities of needed consumables. However, engineering differences between rockets and rover engines might dictate a better solution to be to utilize different components. This is clearly an area for future study and strategic planning.

V. PRODUCTION OPERATIONS: TECHNOLOGICAL CHALLENGES AND LOGISTICS

To produce propellants on Mars, a number of technical components will be required. A full systems analysis would define the mass, volume, power, reliability and cost requirements for each candidate approach. Based upon

available data a survey of the required power levels, one of the main design drivers, has already been undertaken (Friedlander and Cole 1989). The derivation of additional requirements and the detailed technical implementation requires individual design analysis of the various steps that must be undertaken to accomplish each candidate production process.

By the term "production" is meant all phases that must be accomplished in going from a starting resource to a final, ready-to-use propellant product. From a generic viewpoint, these phases include: *mining* the raw material (from the surface or atmosphere); *beneficiation* via physical concentration and separation into desired feedstocks; *extraction* to isolate intermediate products from less valuable components; *chemical conversions* in one or more steps until the end-product is formed; *purification* to remove deleterious minor or trace components; and *conditioning* into a form most appropriate for final storage. Each of these steps will require special equipment, as well as controls of the physical and chemical state of the various intermediates.

Mining the regolith may target the soil fines, the duricrust, the abundant rocks, or local outcrops of bedrock. For propellant production from regolith resources, the most likely target is H_2O and the most likely mode of extraction will be via thermal release. In terrestrial and lunar soils, water can occur in many forms: as interstitial ice ("permafrost ice"); chemically, in hydroxide minerals; adsorbed on grain surfaces; trapped in fluid inclusions and pores; as tightly held interlayers in smectite clays; as water of crystallization (especially in salts); absorbed into hygroscopic compounds; and as implanted H and O ions. The last form is important for the Moon, courtesy of the solar wind, but not for Mars. In contrast, all of the other above forms are possible for Mars, but unknown or rare for lunar surface material.

Even though performed under highly nonoptimum conditions, the Viking organics analyzer did experimentally demonstrate the thermal release of H_2O vapor by simply heating Martian fines. Martian soils may contain several percent releasable H_2O, or even $>10\%$ if clays or salt hydrates are present. If extremely dry SNC meteorites are an appropriate indicator of the water content of Martian rocks, then the soil fines (and duricrust) will be the highly preferred source of H_2O, unless ice-laden permafrost or appropriate sedimentary rocks are also found. Soil fines not only are expected to contain more water, but also have the advantage of not requiring the high energy crushing and grinding operations that rocks would require preparatory to extraction. Beneficiation of duricrust or fines is a simple matter of sieving processes, although further knowledge of the occurrences of hydrated minerals could lead to more sophisticated approaches to select out water-rich material.

The soil and duricrust could also be mined for chloride salts if a Cl-based oxidizer were desired. Beneficiation would be achieved by simple aqueous extraction and fractional crystallization to isolate the chlorides from sulfates on other soluble components.

The atmosphere can also be "mined" for H_2O, although the content is so low that extraction would be slow and considerable energy expenditure would

be required for recovery. More accessible is CO_2, for conversion into O_2, CO, CH_4, CH_3OH, etc. The carbon dioxide can be extracted in many ways: by condensing as frost on a cold plate at 150 K; by adsorption/desorption on any of a number of suitable materiels (clays, activated carbon, molecular sieve [artificial zeolites], etc); by compression and liquefaction; by reversible chemical uptake; by selective membrane diffusion; or other physical techniques. Several of these processes also provide purification of the CO_2 and further enrichment can be accomplished by fractional phase change methods.

The Martian atmosphere also contains minor or trace quantities of noble gases, including Ar, Kr, and Xe for potential use in electric propulsion (EP) systems. The preferred EP gas, Xe, is present at only 80 ppb in the Martian atmosphere. This is still some 10 times the abundance of helium-3 isotope in the lunar regolith, and much easier to extract and beneficiate than the latter. However, transporting EP fuel into Mars orbit is a significant cost and generally may not be sufficiently leveraging of MAMA or IMLEO to justify its production.

Nitrogen gas or noble gases can serve as inert gas pressurants for propellant feed, although the mass of pressurization gas needed is typically quite small compared to the propellants themselves, and of relatively minor significance even compared to the mass of associated tanking and plumbing of the propulsion system. Thus, in many cases it may be more practical to simply transport the required quantity of gas in highly pressurized, lightweight tanks from Earth. All of these gases could be obtained by the physical separation processes listed above, but the procedures would differ considerably for each gas.

Once the desired chemical is obtained from the "ore," chemical conversion to the desired end-product will be necessary in many cases. Such conversion may be relatively straightforward, such as the electrolysis of H_2O to H_2 and O_2, or the conversion of CO_2 to CH_4 using a source of H_2; moderately complicated, such as the conversion of CO_2 into C and O_2; or challenging, such as the isolation of Mg metal from $MgSO_4$, the generation of H_2O_2, or manufacture of complex hydrazines. It is not that any of these conversions require exotic chemistry, but rather that they must be accomplished as multi-step processes, at high or low temperatures, high pressures, and/or at low yields. The need for highly automated equipment with many autonomous control and response functions in a very different environment than on Earth will result in one set of design problems. Constraints on safety for humans and robots performing process and maintenance functions will invoke others. Added to this are requirements for reliability, robustness, and recoverability, as well as low mass of the equipment and highest practicable power efficiency.

Chemical conversions typically employ catalysts, which also typically require replacement due to progressive poisoning by extraneous constituents. Reactants must often be recycled through the process stream to increase effluent concentrations, and working fluids (e.g., solvents) need to be recovered and reconditioned for further use.

Heat energy must often be recovered and used at a lower temperature, e.g., in pre-heating, to maximize process efficiency. On the other hand, excess heat may require special efforts for dissipation to keep reaction temperatures within range. For example, reaction of H_2 with CO_2 liberates on the order of 4000 kW-hr of thermal energy for each metric tonne of methane produced. On Mars, evaporation cooling towers are out of the question, but the sub-zero ambient temperatures facilitate radiators (Ash et al. 1988) and the atmosphere, although tenuous, will allow some convective thermal transfer.

Purification of the product may proceed via many alternative pathways. However, chemical extractions would involve one or more additional reagents, which would then require additional reclaiming and purification steps. Physical processes involving phase transformations may be more appropriate to the space setting. For example, fractional distillation or crystallization could result in a purer product with minimal complexity, although residual chemical liquor would need further processing, or simply disposal with loss of contents.

Even after obtaining the pure product, it may be necessary to condition it for containerization. At the very least, its temperature and pressure will have to be adjusted to that planned for storage. Some items (e.g., hydrazine, hydrogen peroxide, etc.) can decompose and must be protected from antagonistic catalysts. Stabilizers may be added to compensate for impurities or wall effects, or to terminate one or more intermediates in a decomposition chain reaction.

Logistics includes the storing, transporting, gauging and dispensing of propellant, as well as any necessary disposal of waste products. Liquid/gas containment, conditioning, and transfer in isolation from the low-pressure and low-temperature ambient environment of Mars will be challenging in terms of mass transfer efficiency, power consumption, speed, reliability, safety and cost-effectivity. The typical average temperature for the majority of the Martian surface (mid-latitudes and lower) ranges between $-35°$ and $-80°$ C, with a typical value of $-55°$ C in equatorial regions. Unfortunately, this is still more than 100° C too warm to store many cryogens, such as LOX, LN_2, CH_4, CO, etc., and much too warm for storing hydrogen. As indicated in Fig. 1, many other propellants would freeze at such temperatures, including hydrogen peroxide and most of the hydrazines. Active cooling would be needed to preserve the cryogens, including vacuum-jacketed tanks with multi-layer radiation shield insulation. Such refrigeration would be critically important to successful storage because of venting in order to counteract the input of heat to the tanks. Indeed, such systems would be inherently hazardous in the face of loss of refrigeration or loss of vacuum integrity. Freezing of propellants could be prevented, or reversed, by using foam-insulated tanks and modest electrical heating elements. Passive radioisotope heater units (RHU) also could be employed to bias the temperature of the storage tank into the liquid region of the phase diagram. By placing RHUs in the center of such tanks (e.g., Geels 1990), their radiation hazard could be reduced, especially during periods when the tank is full.

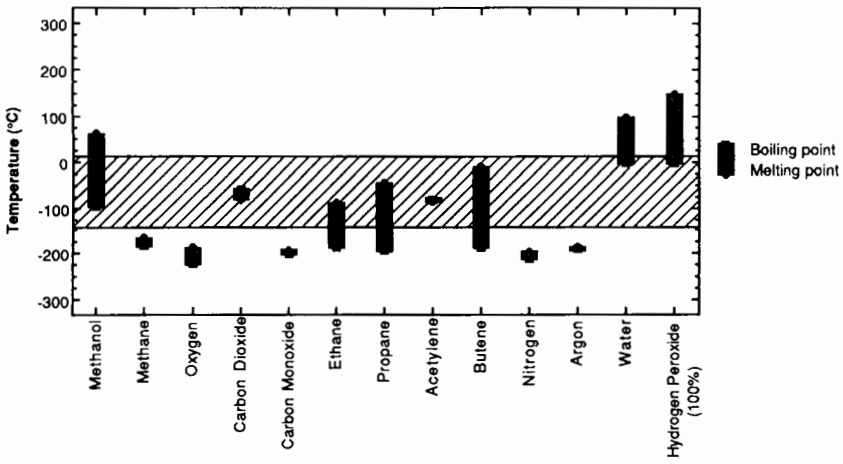


Figure 1. Storage temperatures and other data for Mars propellants.

As is shown in Fig. 1, certain propellants can be maintained in the liquid state with little or no thermal conditioning. The best of these are some of the alcohols (including methanol), CO_2 (under modest pressure), UDMH, and eutectic mixtures of H_2O_2/H_2O. For very long-term storage, and especially for transfers to other tanks over long runs (pipelines), these materials would be very attractive.

VI. PROPELLANT PRODUCTION AT PHOBOS OR DEIMOS

Ever since the Viking mission, which established the bulk densities of Phobos and Deimos as very low, these objects have been thought to possibly be representatives of the volatile-rich type of meteorites known as carbonaceous chondrites. If so, they could serve as a valuable source of organic compounds and water for production of propellants. For this reason, Phobos and Deimos have often been proposed as favorable locations for propellant production in the Mars system (O'Leary 1985; Lewis and Lewis 1987; Cordell 1989) and have even been considered for transportation of propellant from Mars to the Earth-Moon system (Cordell 1990; Zuppero, in preparation). Other than the very low albedo of these objects, which is also consistent with carbonaceous material, there is no evidence to confirm the actual presence of these compounds. More recently, the infrared spectrometer on the Soviet Phobos I satellite mission to Mars indicated higher water of hydration in Martian soils than in Phobos itself (Bibring et al. 1989). Exploration of one or both of these objects should obviously be high priority for one or more future missions to Mars, although for a variety of reasons, they are not appropriate for locating a human base or for serving as a mandatory waypoint.

Making propellant on either satellite would not, however, improve the accessibility to the Martian surface by a significant amount because aerobraking during entry obviates the need for propulsion except for terminal descent. Indeed, the ΔV to land on the Martian surface from either Phobos or Deimos is approximately 3 times as large as for landing from an elliptical orbit, Table III. Of course, the ascent leg could be supplied by propellants produced locally on the satellites, but a more mass-efficient way is to produce the propellants on the Martian surface, if all other factors are equal. However, as pointed out by Lewis and Lewis (1987), Phobos may be ideal for supplying the lower latitudes of Mars with water or propellant if water is indeed scarce at these locations compared to the more poleward regions where surface temperatures of the planet are lower and hence more effective in preserving near-surface permafrost ice.

Operating a base on Phobos or Deimos has advantages such as direct sunlight and low gravity. The latter may be an impediment, however, making necessary the deployment of deep anchors to fix the equipment to the surface. Some of the dust produced during mining operations could reach escape velocities and then through orbital evolution later impact the surface, base, or nearby vehicles at relatively high speed. In addition, observation of the Martian surface from Phobos is limited mainly to mid and low latitudes. Transportation access (to and from the surface) is limited to a very narrow band around the equator, unless the very major penalties for plane changes are taken.

As discussed above, ΔV accessibility to Phobos and Deimos is complicated by the orbital mechanics of arrival and departure declinations, as well as the energetics of reaching high circular orbits. A number of strategies for using Phobos as a propellant supply for return to Earth of an arriving spaceship have been examined (Clark and Miller 1988). The payback for transporting the entire vehicle to Phobos is generally unfavorable. Thus, even if the infrastructure were already in place, chemical propellant production at Phobos would not be beneficial by this scenario. However, other strategies are possible. The payback is closer to being favorable if the spaceship goes into a HEMO at Mars arrival, then deploys a lightweight tanker vehicle loaded only with sufficient propellants to transfer to Phobos. The tanker is filled and transports itself back to the mother spaceship. With this scenario, the payback is still unfavorable for typical manned mission examples if the inclination of the HEMO must be high. More favorable yet is when the tanker vehicle is stationed at Phobos itself. Net mass savings can be realized, as seen in Fig. 2. However, mass is only an indicator metric. A cost- and infrastructure-benefits analysis would also have to be performed to actually justify a decision to employ this or any other Phobos propellant base scenario.

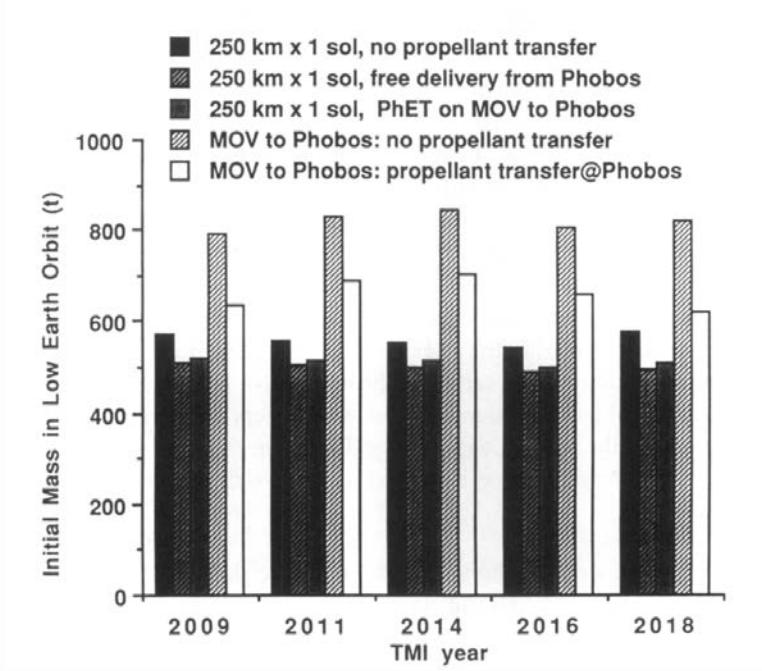


Figure 2. Mass in low Earth orbit (IMLEO) to accomplish the same manned Mars mission, with and without Phobos propellant.

VII. PROPELLANT PRODUCTION AT STRATEGICALLY LOCATED ASTEROIDS

A surprisingly large number of asteroids have reflectance spectra indicating the possible presence of water of hydration and/or carbonaceous materials on their surfaces. For transportation to Mars, it is theoretically possible, although by no means yet demonstrated, that one or more suitable asteroids may be in a celestial position to be of great advantage. One example would be an orbit with relatively low-energy access from Earth and also a low-energy access to Mars, such as a heliocentric orbit whose perihelion is at Earth's orbit and aphelion at Mars' aphelion. Such a target could be initially supplied from Earth with equipment for mining and propellant production. At appropriate launch opportunities, which may be unfortunately rare, the propellants could be transferred to Mars HEMO and cached for future arriving vehicles.

Yet another strategy to assist in Mars exploration would be to locate a chemically suitable asteroid in a most favorable location to Mars and then use propulsion to cause it, or a fragment of it, to aerocapture into HEMO. Once in the Mars system, mining and processing systems could produce material for transport to the surface or other Mars orbital location. As is pointed out in the Chapter by Lewis and Hutson, a similar scenario to emplace asteroidal material into HEEO or LEO at Earth also holds considerable promise for

future resource utilization, staged at areas of maximum use.

VIII. SUMMARY

As is undoubtedly all too obvious to the reader, this review provides only the barest outline of the technical promise and ramifications of propellant production in the Mars system. Further exploration of Mars can only enlarge the possibilities. But from what is currently known, it already would be possible to perform detailed assessments of what will be required to implement many candidate approaches. What is now needed are a series of serious design efforts, backed up by carefully selected laboratory projects and engineering demonstrations to evaluate critically the options for propellant production on Mars. A fruitful collaboration between government, academia, and industry to bring to bear the mining, chemical, and aerospace technologies, along with operational flight experience is needed to achieve the potential of planetary propellants.

Acknowledgments. The storage temperature data and figure were prepared by C. MacLeod. Useful inputs and comments were provided by L. Mason. The author is also indebted to R. Ash and M. Stancati for careful reviews of the submitted manuscript.

REFERENCES

Ash, R. L., Dowler, W. L., and Varsi, G. 1978. Feasibility of rocket propellant production on Mars. *Acta Astronautica* 5:705–724.

Ash, R. L., Richter, R., Dowler, W. L., Hanson, J. A., and Uphoff, C. W. 1982. Autonomous oxygen production for a Mars return vehicle. Paper presented at the 33rd Intl. Astronautical Congress, Sept. 27–Oct. 2, Paris. IAF-82-210 (Intl. Astronautical Federation).

Ash, R. L, Wiseman, M. A., and Jackson, C. C. 1988. Radiator design considerations for martian surface applications. Paper presented at the 33rd Intl. Astronautical Congress, Sept. 27–Oct. 2, Paris. ASME Special Publ. HTD, Vol. 101, pp. 147–152.

Bibring, J.-P., Combes, M., Langevin, Y., Soufflot, A., Cara, C., Drossart, P., Encrenaz, Th., Erard, S., Forni, O., Gondet, B., Ksanfomality, L., Lellouch, E., Masson, Ph., Moroz, V., Rocard, F., Rosenqvist, J., and Sotin, C. 1989. Results from the ISM experiment. *Nature* 341:591–593.

Clapp, W. M. 1985. Water supply for a manned base. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 557–566.

Clapp, W. M., and Scardera, M. P. 1987. Applications of in-situ CO-O propellent production at Mars. In *Case for Mars III*, ed. C. R. Stoker (San Diego: Univelt), pp. 513–534.

Clark, B. C., 1979. The Viking results—the case for man on Mars. In *The Future United States Space Program*, AAS 78-156, pp. 263–277.

Clark, B. C. 1984. Chemistry of the Martian surface—Resources for the manned exploration of Mars. In *The Case for Mars*, ed. P. J. Boston (San Diego: Univelt), pp. 197–208.

Clark, B. C. 1985. The H-atom resource on Mars. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), pp. 519–526.

Clark, B. C., and Miller, J. B. 1988. Phobos propellant utilization study. Paper presented at the 2nd Intl. Symp. on Solar System Exploration, Pasadena.

Clark, B. C., Baird, A. K., Weldon, R. J., Tsusaki, D. M., Schnabel, L, and Candelaria, M. P. 1982. Chemical composition of Martian fines. *J. Geophys. Res.* 87:10059–10067.

Clark, B. C., Kalkstein, J. L., and Meyer, S. 1991. The case for the methanol-powered planetary rover. Paper presented at the 42nd Intl. Astronautical Congress, Oct. 5–11, Montreal. IAF-91-447 (Intl. Astronautical Federation).

Cordell, B. M. 1989. Manned Mar mission overview. Paper presented at the 25th AIAA,ASME,SAE, and ASEE Joint Propulsion Conf., July 10–13, Monterey, Ca. AIAA-89-2766

Cordell, B. M. 1990. Transportation approaches for manned Mars missions. Paper presented at the AIAA Space Programs and Technologies Conf., Sept. 25–27, Huntsville, Al. AIAA-90-3892.

French, J. R. 1985. The impact of martian propellant manufacturing on early manned exploration. In *Case for Mars II*, ed. C. P. McKay (San Diego: Univelt), 527–535.

French, J. R. 1989. Rocket propellants from martian resources. *J. Brit. Interplanet. Soc.* 42:167–170.

Friedlander, A., and Cole, K. 1989. Power Requirements for Lunar and Mars Exploration Scenarios. SAIC Rept. 89/1000. Prepared under contract NASW-4214 to NASA Headquarters.

Frisbee, R. H. 1987. Mass and power estimates for Mars in-situ propellant production systems. Paper presented at the 23rd AIAA,ASME,SAE, and ASEE Joint Propulsion Conf., June 29–July 2, San Diego, Ca. AIAA-87-1900.

Frisbee, R. H., French, J. R., and Lawton, E. A. 1987. A new look at oxygen production on Mars—in-situ propellant production (ISPP). Paper presented at the 25th Aerospace Sciences Meeting, June 12–15, Reno, Nev. AIAA-87-0236.

Geels, S. A. 1990. System Design of a Mars Ascent Vehicle. S. M. Dissertation, Massachusetts Inst. of Technology.

Hill, P. G., and Peterson, C. R. 1970. *Mechanics and Thermodynamics of Propulsion* (Reading, Mass.: Addison-Wesley).

LeCompte, M. A., and Stets, J. P. 1991. Propellant manufacturing on Mars: Issues and implication. In *Space Manufacturing 8, Proc. of the 10th Princeton/AIAA/SSI Conf.: Energy and Materials from Space*, eds. B. Faughnan and G. Maryniak (New York: AIAA).

Lewis, J. S. 1972. Metal/silicate fractionation in the solar system. *Earth Planet. Sci. Lett.* 15:286–290.

Lewis, J. S., and Lewis, R. A. 1987. *Space Resources: Breaking the Bonds of Earth* (New York: Columbia Univ. Press).

Linne, D. L. 1991. *Carbon Monoxide and Oxygen Combustion Experiments: A Demonstration of Mars In Situ Propellants*, NASA TM-104473.

Meyer, T. R. 1981. Extraction of Martian resources for a manned research station. *J. Brit. Interplanet. Soc.* 34:285–288.

Meyer, T. R., and McKay, C. P. 1984. The atmosphere of Mars—Resources for the exploration and settlement of Mars. In *The Case for Mars*, ed. P. J. Boston (San Diego: Univelt), pp. 209–230.

O'Leary, B. 1985. Phobos and Deimos as resource and exploration centers. In *Case for Mars II*, ed. C. P. McKay, (San Diego: Univelt), pp. 62, 225–244.

Ramohalli, K., Lawton, E., and Ash, R. 1989. Recent concepts in missions to Mars: Extraterrestrial processes. *J. Propulsion* 5:181–187.

Sridhar, K. R., and Iyer, V. A. 1991. Thermal design of a Mars oxygen production plant. Paper presented at the 42nd Intl. Astronautical Congress, Oct. 5–11, Montreal. IAF-91-447 (Intl. Astronautical Federation).

Stancati, M. L., Niehoff, J. C., Wells, W. C., and Ash, R. L. 1979. Presented at the AIAA Conf. on Advanced Technology for Future Space Systems, Hampton, Va.

Zubrin, R. M. 1989. Indigenous martian propellant. *Aerospace America* 27:48–51.

Zubrin, R. M., Baker, D. A., and Gwynne, O. 1991. Mars direct: A simple, robust, and cost effective architecture for the Space Exploration Initiative. Paper presented at the 29th Aerospace Sciences Meeting, Jan. 7–10, Reno, Nev. AIAA-91-0329.

PLANETARY ENGINEERING

JAMES B. POLLACK
NASA Ames Research Center

and

CARL SAGAN
Cornell University

Assuming commercial fusion power, heavy lift vehicles and major advances in genetic engineering, we survey possible late-21st century methods of working major transformations in planetary environments. Much more Earth-like temperatures may be produced on Mars by generating low freezing-point greenhouse gases (e.g., CO_2, NH_3, CFCs) from indigenous materials or by transporting them from elsewhere; on Venus by cancelling the greenhouse effect with high-altitude absorbing fine particles, or by a sunshield at the first Lagrangian point, and/or by sequestering or transforming CO_2 at the surface; and on Titan by greenhouse and/or fusion warming. To produce global environments suitable for plants and animals, including humans, requires modifying the atmospheric composition and mass and altering the surface temperatures on these bodies. In general, engineering congenial worlds for plants is much easier than for humans, and is also a useful means of working further modification of the atmospheric composition, especially the establishment of several hundred mbar of O_2 from H_2O. Establishing global habitats suitable for humans will require the addition of at least several hundred mbar of N_2 and O_2 into the Martian atmosphere; the removal of most of the CO_2 in Venus' atmosphere (most plausibly by forming carbonate minerals) plus the addition of large amounts of water; and the addition of several hundred mbar of O_2 to Titan's atmosphere. Climatologically active abundances of some gases may be toxic to humans. It is not clear that any of these schemes are technically feasible (much less cost effective) with technologies projected for the end of the next century. They also raise disturbing questions about environmental ethics. Global warming on Earth has already led to calls for mitigation by planetary engineering—e.g., emplacement and replenishment of reflective or anti-greenhouse layers at high altitudes, or sunshields in space. But here especially we must be concerned about precision, stability, and inadvertent side-effects. The safest and most cost-effective means of countering global warming of the Earth—beyond, e.g., improved energy efficiency, CFC bans and alternative energy sources—is the continuing reforestation of $\sim 2.5 \times 10^7$ km^2 of the Earth's surface. This can be accomplished with present technology.

I. GENERAL PRINCIPLES

Human technology is now able to affect the terrestrial atmosphere and climate

on a global scale. Three mechanisms have drawn particular attention in recent years:

1. Global warming through a side-effect of modern technology—the increasing greenhouse effect. It arises chiefly from the burning of fossil fuels (releasing CO_2), from the industrial production of chlorofluorocarbons (releasing CFC s), and from cattle and rice paddies (which release CH_4). Global warming of several K over the next century might, it is feared, lead to drought, desertification in mid-latitude continental interiors, massive agricultural failure and relocation, a global rise in sea level and the flooding of coastal cities and island nations (see, e.g., Hansen et al. 1988; Strong 1989; MacCracken et al. 1990).
2. Ozonosphere depletion. It is thought to arise chiefly from the industrial production of chlorofluorocarbons (CFC s) and threatens—through a significant increase in the surface flux of solar near-ultraviolet radiation (WMO 1985; NASA 1988)—not just increases in skin cancer and a weakening of the human immune system, but an assault on the primary photosynthetic producers at the base of the food chain.
3. Nuclear winter. This is a long-overlooked climatic consequence of nuclear war. Even 1% of the 1992 global strategic arsenals—if targeted on city centers and, especially, petroleum refineries and storage depots—seems capable of reducing Northern Hemisphere temperatures by as much as 10 to 20 K in continental interiors in a time scale of the order of 10 days (and then relax back to unperturbed conditions on a time scale of months). The principal mechanism by which nuclear winter works is by undoing the greenhouse effect ("anti-greenhouse effect"), which happens when appreciable sunlight is absorbed above the greenhouse gases (Turco et al. 1983,1990; Pittock et al. 1986; Sagan and Turco 1990). Among the predicted consequences of nuclear winter following a full nuclear exchange in Northern spring or summer is the devastation of Northern Hemisphere agriculture and the subsequent deaths by starvation of vast numbers of people (Harwell and Hutchinson 1985).

These three processes demonstrate the general proposition that humans can now alter environments on a planetary scale. This conclusion follows even if we were to entertain serious reservations about the canonical values of the speed or severity of these effects within the range of uncertainty in current atmospheric modeling. There may be mitigating effects that have not yet been identified and that just counterbalance each of these catastrophes, but such effects have certainly not been demonstrated, and hoping for such a *deus ex machina* may be tantamount to exercising what psychiatrists call denial.

In none of these three examples is a carefully designed stable change brought about; instead, the changes are haphazard and inadvertent, and almost always constitute an unpleasant surprise to those responsible for the relevant technology. It seems possible, therefore, if present trends continue, that

within the not-too-distant future human technology should be capable of even more major alterations, both intentional and inadvertent. An important issue is whether any can cause improvements rather than deterioration in the planetary environment—perhaps with high-precision negative feedbacks.

After the discovery of the inclement high temperatures at the surface of Venus, it was natural to try to imagine a possible technological fix to make that planet more Earth-like (Sagan 1961). A wide variety of schemes have been suggested in the subsequent scientific literature, usually to transform Venus or Mars into a more Earth-like environment (see, e.g., Averner and MacElroy 1976; Burns and Harwit 1973; Fogg 1987,1989; Oberg 1981). This subject was earlier described as "terraforming," a term coined by the science fiction writer Jack Williamson under the pseudonym Will Stewart in a series of stories published in the 1940's. However, it is possible to imagine a range of environmental alterations other than nudging a hostile world into more Earth-like conditions, and we here use the more general phrase "planetary engineering" (Sagan 1961,1973).

The traditional objective of planetary engineering is to generate an Earth-like environment on other worlds of the solar system, so that humans and other inhabitants of the Earth can live there without special protective gear (spacesuits, large enclosed habitats, etc.). A continuum of intermediate cases exists, usually more readily achievable, between the current environment of an extraterrestrial body and present Earth ambient conditions. An important special case for planetary engineering is to reverse significant perturbations (e.g., global warming) in the environment of our own planet. This is surely easier than planetary engineering on other worlds: the circumstances under which humans can survive unaided represent at most very small variations from present Earth ambient conditions; but many microbes and plants grow under less severe constraints (cf. Table I).

In this chapter we examine the likelihood of accomplishing significant planetary engineering, particularly towards more Earth-like conditions, on Mars, Venus and Titan, although opportunities exist for a number of other worlds in the solar system as well. We focus chiefly on proposals by which clement temperatures (0–30°C) can be realized on a global scale. Such a temperature range meets many essential requirements for habitability, including the potential availability of surface liquid water. As stressed by Sagan (1961) and by McKay et al. (1991), introducing plants first may lead with comparatively little additional effort to other desired changes (e.g., production of significant quantities of photosynthetic O_2). At the end of the chapter we address schemes for countering or mitigating global warming on Earth. We adopt the spirit of Henry David Thoreau's remark, "What is the use of a house if you haven't got a tolerable planet to put it on?"

We postulate a future with abundant heavy-lift vehicles, commercially available and safe nuclear fusion, major practical advances in genetic engineering and a dedication to the exploration of the solar system. Over the next century all of these developments seem plausible, although they are by no

TABLE I
Limits of Habitability[a]

Parameter	Limits	Note
Global temperature	0–30°C	Earth mean temperature 15°C
Plants only:		
Total pressure	>10 mbar	Water vapor pressure plus O_2, N_2 and CO_2
CO_2	>0.15 mbar	Lower limit set by photosynthesis No sharp upper limit
N_2	>1–10 mbar	Nitrogen fixation
O_2	>1 mbar	Plant respiration
Human breathable total pressure,		
Pure O_2	>250 mbar	Lung water vapor, CO_2, and O_2
Air mixture	>500 mbar	Based on high elevation
	<5000 mbar	Buffer gas narcosis
CO_2	<10 mbar	Set by toxicity
N_2	>300 mbar	Buffer gas
O_2	>130 mbar	Lower limit set by hypoxia
	<300 mbar	Upper limit set by flammability

[a] Table from McKay et al. 1991.

means guaranteed. We also assume that fusion power will provide us with the technology to move at will asteroids ten kilometers in diameter and smaller around the inner solar system. The energy requirements, as we discuss below, may by then be wholly within reach. Transport of small worlds within the inner solar system has long been proposed for other reasons—e.g., supplying large quantities of precious metals—but much smaller worlds would suffice for that purpose. This technology may be developed for other reasons as well: to deflect errant near-Earth objects from impact trajectory with our planet (cf. Morrison et al. 1992). However, such technology would also give its possessors an unprecedented ability to destroy life on Earth, and it is not clear that a cost/benefit analysis favors its development (Sagan 1992), at least in the near future. But in keeping with the optimistic assumptions of this chapter, we postulate a future in which adequate safeguards against the misuse of

this technology exist.

We assume that fusion power is based on either reactions between deuterium and tritium (D-T) or between deuterium and ^{3}He (D-He). This choice is based on the extensive work done over the last several decades in developing practical schemes for fusion power (D-T) and plausible projections (D-He) (Cordey et al. 1992; Wittenberg et al. 1986). Because of the very low reaction cross sections of the *p-p* cycle that helps power the Sun, we ignore the much more efficient fusion power based on this cycle. In the case of D-T fusion, ^{6}Li may provide the source of the tritium through reactions occurring in the blanket of the fusion chamber (Cordey et al. 1992). For D-He fusion, ^{3}He may be derived from solar wind implanted in the regolith of the Moon (Wittenberg et al. 1986), or asteroids. We assume an adequate amount of ^{6}Li and ^{3}He for the scenarios considered in this chapter. We do, however, provide estimates of the amount of water required, which serves as the source of D. For D-T fusion, about 20 times the mass of rock is required to provide the needed ^{6}Li as the mass of water required to obtain a given amount of D. In either case, the amount of energy released per mole of reactant is about the same. Below, we assume that 50% of the mass deficit between reactants and products is available as useful energy. Thus, we use a fusion power of 4.2×10^{18} erg g^{-1} of D or 7.5×10^{13} erg g^{-1} of H_2O, assuming a terrestrial D/H ratio of 1.6×10^{-4}. Utilizing Martian water may increase the yield per gram of H_2O by as much as a factor of 5 because of the higher D/H ratio of water vapor in the Martian atmosphere.

We do not consider the possible use of thermonuclear weapons for planetary engineering, because the Outer Space Treaty of 1967—which the United States, the Soviet Union (and presumably its successor states), the nations of Western Europe and Japan, among many other nations, have ratified—specifically forbids installation of such weapons on other celestial bodies, and in "outer space" in general. The 1963 Limited Test Ban Treaty also prohibits the explosion of nuclear weapons anywhere except beneath the surface of the Earth (A. C. D. A.1982).

It seems possible (Sagan 1986) that between increasing the abundance of greenhouse gases and placing fine particles at high altitudes to provide an anti-greenhouse effect, we will have the means to heat or cool a planetary surface at will; and that, for the latter, it might suffice to pulverize a small asteroid or cometary nucleus. In addition, large partially opaque sunscreens emplaced between the Sun and a given world could be used to cool that world, and fusion technology could be used to warm it. In principle, anticipated advances in human technology might, within the next century, provide us with practical planetary thermostats.

We take particular note of the time scale for planetary engineering implicit in various methods. It would seem foolish to implement today a scheme that would occupy 10^4 or 10^6 yr, say, to achieve its effects, when in, say, 100 yr we will likely have much more powerful technologies for the purpose. Current political realities suggest that an expensive project without a significant payoff

in 25 yr or less is infeasible. As a rough figure of merit, any planetary engineering scheme that takes more than several decades to achieve significant results (if not full implementation) should be rejected on such grounds alone.

Projecting technological trends more than a few decades into the future is in any case a forlorn exercise. But advances in genetic engineering far beyond what we assume in the remainder of this chapter seem possible. We therefore note that, at some time in the future, a much more elegant way to overcome our parochial habitat restrictions may be to genetically engineer humans for other worlds than to physically engineer other worlds for humans (cf. Stapledon 1948).

We find below that planetary engineering generally entails severe ancillary environmental costs. This may be because we have not yet conceived of the really elegant methods. But these costs, especially, raise a number of issues: given that any planetary engineering scheme entails a balance of benefits against costs, how certain must we be that key scientific information will not thereby be destroyed? How much understanding of the world in question do we need to have before planetary engineering can be relied upon to produce the desired end state? Can we guarantee a long-term human commitment to maintain and replenish an engineered world when human political institutions are so short-lived? If a world is even conceivably inhabited—perhaps only by microorganisms—do humans have a right to alter it? What is our responsibility to preserve the worlds of the solar system in their present wilderness states for future generations—who may contemplate uses that we are today too ignorant to foresee? These questions may perhaps be encapsulated into a final question: can we who have made such a mess of this world be trusted with others? At the end of this chapter, we briefly re-address some of these issues.

II. MARS

Today Mars is a frigid, desert planet. The fact that conditions seem to have been warmer and wetter 4 Gyr ago when the solar luminosity was $\simeq 0.75$ its current value (Sagan and Mullen 1972; Pollack et al. 1987) suggests that it may be possible to move the Martian climate into a more temperate range. Although today temperatures near the equator can rise above the freezing point of water during the warmest times of the day, the diurnally averaged temperatures are everywhere well below freezing. Here, we consider ways of increasing the globally, diurnally, and seasonally averaged temperatures to values comparable to the annually averaged temperatures at mid-latitudes of Earth. We note from the start that even if this were done gently, it would disturb or destroy some major Martian landforms—e.g., the polar laminae.

The low mean temperatures on Mars are due both to it being further from the Sun than the Earth (a unit area on Mars absorbs roughly half the solar flux as an identical area on Earth) and to Mars having a very thin atmosphere—so that the greenhouse warming is only about 7 K, in contrast to about 33 K for the

Earth. On present-day Mars, CO_2, the major constituent of the atmosphere, is the chief agent responsible for its modest greenhouse warming, through the opacity it provides to thermal radiation emitted by the surface in the wavelength region around 15 μm. The apparently most cost-effective way to warm Mars is by enhancing the atmospheric greenhouse effect through the introduction of gases that absorb at thermal wavelengths outside the 15 μm region. There are two possible sources of such gases—exogenous to Mars and endogenous. We could imagine, for example, transporting ice from the outer moons or rings of Saturn to Mars, but pure water, which appears to be the composition of the rings of Saturn, while an important greenhouse gas on Earth, would be ineffective on Mars, because the temperatures there are so low that water freezes out. In general for both Earth and Mars, the water abundance is buffered by reservoirs at the surface, so it can only amplify the greenhouse effect of other gases once they produce significant warming. We require gases that not only are effective infrared absorbers, but that also remain in the gas phase at present Martian temperatures. Below we discuss three candidates, NH_3, CO_2 and halogen compounds.

A. Ammonia

It would be far more effective to utilize materials already on Mars. NH_3 absorbs strongly both near 10 μm and longwards of 20 μm. Small amounts can greatly enhance the greenhouse effect on Mars. Calculations suggest (Sagan and Mullen 1972; Pollack 1979) that a partial pressure of ammonia of $\sim 10^{-4}$ bar is needed to warm the Martian surface to above the freezing point of water.

The required amount of NH_3 might be generated in several ways. First, consider appropriately designed microorganisms that convert atmospheric N_2 to NH_3, an exothermic reaction, with the hydrogen derived from subsurface bound or frozen water. Nitrogen-fixing microorganisms are abundant on Earth and central to agriculture; they routinely convert N_2 to NH_3. But none is known to metabolize under simulated Martian conditions. Cryophilic nitrogen-fixers that extract their water requirements from bound or solid water would have to be engineered; and/or heated artificial habitats created that would vent NH_3 to the outside atmosphere. As the partial pressure of N_2 on Mars is $\sim 2 \times 10^{-4}$ bar, about 25% of the available atmospheric nitrogen would be required to generate 10^{-4} bar of NH_3. Properly engineered microorganisms might be placed on the permanent water ice north polar cap. Despite the summertime temperatures of about $-30°C$, the solar flux might be sufficient for such microbes to provide photosynthetically the energy needed to manufacture ammonia.

Alternatively, ammonia might be generated by chemical factories that use nuclear fusion as a power source to convert atmospheric N_2 to NH_3. The hydrogen needed both for fusion fuel and for the chemical conversion of N_2 to NH_3 could come from the residual north polar cap, or perhaps from the regolith. If as much as 500 calories of activation energy were needed to

convert each N_2 molecule to NH_3, $\sim 1.7 \times 10^{13}$ g of water for a terrestrial D/H ratio would be needed for fuel to produce 10^{-4} bar of NH_3 over the entire planet. (Even less water is needed if the observed Martian atmospheric D/H ratio is used). This corresponds to only about the top 40 μm of the north polar cap. However, about 30 cm of H_2O from the north polar cap would be consumed in supplying the H needed to convert 1/4 of the current atmospheric N_2 to 10^{-4} bar of NH_3.

As a third possibility, especially if the budget of atmospheric N_2 and surface nitrates is not quite adequate to generate $T > 0°C$ through an ammonia greenhouse, N_2 might be derived from the atmospheres of Venus, Earth, or Titan and transported to Mars. For example, to generate an NH_3 partial pressure of $\sim 1 \times 10^{-3}$ bar, amounts of N_2 equal to $\sim 5 \times 10^{-4}$, 1×10^{-3}, and 5×10^{-5} times the N_2 content of the atmospheres of Venus, Earth, or Titan, respectively, would be required. If liquified, the requisite amount of N_2, $\sim 6.8 \times 10^{17}$ g, would measure about 10 km across. If a ΔV of 10 km s^{-1} were needed to transport the liquified N_2 to Mars, $\sim 4.5 \times 10^{15}$ g of H_2O would be required to power the fusion engine. This is equivalent to the amount of water in a 1-km-sized comet. Alternatively, the N_2 might be derived from the N in organic compounds contained in C or D asteroids, whose orbits come close to that of Mars. Such bodies may contain ~ 0.5 wt% N. Thus, a much lower ΔV might suffice for transporting nitrogen to Mars. An asteroid of diameter about 50 km would be needed to supply the requisite amount of N. The cheapest, although least responsible, method of extracting the nitrogen would be by simply crashing the asteroid into Mars; this would generate a crater hundreds of km across, destroying the underlying terrain.

There is, though, an important problem with any scheme to use NH_3 as a Martian greenhouse gas: ammonia is readily dissociated by solar ultraviolet radiation at wavelengths, λ, shortward of ~ 0.23 μm (see, e.g., Atreya et al. 1978). In the Martian atmosphere, neither CO_2 nor H_2O would block ultraviolet radiation at $\lambda > 0.2$ μm. A net dissociation rate of only $\sim 50\%$, a reasonable lower bound, implies $\sim 2.4 \times 10^{12}$ molecules dissociated cm^{-2} s^{-1} at Mars (*ibid.*). Thus, 10^{-4} bar of NH_3 would be totally converted back to N_2 by solar ultraviolet radiation in about 30 yr.

However, 30 yr is much longer than the radiative time constant for the Martian atmosphere; so before NH_3 is appreciably photodissociated, its infrared opacity will have significantly warmed the surface. Thereafter, the gas phase dissociation products of NH_3, N_2 and H_2 would be recycled to NH_3 by essentially the same technology used to produce it initially. Another potential problem is the reaction of NH_3 with atmospheric CO_2 to form ammonium carbonates and other salts at the low temperatures of the winter polar regions; but the kinetic time scales are unknown for these compounds, and any that form in winter may be vaporized the following spring.

B. Carbon Dioxide

A second possible way to augment the greenhouse effect on Mars is to increase

the carbon dioxide content of its atmosphere. At sufficiently high pressures, pressure-induced transitions and hot bands of gas phase CO_2 close the thermal infrared windows. Clement average temperatures can be achieved by raising the surface pressure from its current value of 7 mbar (millibars) to $\sim$ 1 bar (Pollack et al. 1987).

One conceivable source of atmospheric CO_2 might be the dry ice contained in the polar caps. However, the large perennial cap in the northern hemisphere is composed, apparently exclusively, of water ice. The smaller perennial cap in the south had a surface layer of CO_2 ice during the first year of the Viking spacecraft mission (1976–77), but in other years this layer may be very thin or even disappear. An upper limit to the amount of CO_2 available in the Martian polar caps can be derived from the fact that under high enough hydrostatic pressures, solid CO_2 will liquify; this limit corresponds to a cap thickness of $\sim$2 km (Sagan 1973). Even if all the south cap were CO_2 to a depth of 2 km and even if all of this carbon dioxide were liberated in the gas phase, it would amount only to some 60 mbar pressure when uniformly distributed over the planet. Thus, the available CO_2 from the polar cap falls short of the required abundance by more than an order of magnitude.

If, nevertheless, it were desirable to vaporize the dry ice in the polar cap to begin increasing CO_2 partial pressures, we might consider lowering the cap's albedo by depositing very dark carbonaceous material of asteroidal origin, or genetically engineering plants with dark, broad leaves that can survive on the polar caps (Sagan 1973).

Another source of CO_2 is carbonate anions in the rocks. Spectroscopic observations of Mars provide evidence for significant amounts of carbonates (Pollack et al. 1990). Suppose that these rocks in toto contain more than a bar of CO_2 and that they are preferentially concentrated in certain basin areas. If nuclear fusion were used as a power source to liberate this CO_2, then $\sim 5.7 \times 10^{17}$ g of H_2O would be required to generate 1 bar of CO_2—equivalent to about the top meter of water in the north polar cap. This solution for planetary engineering of the Martian climate, while much more energy-intensive than the NH_3 solution, has the advantage of being much more permanent. CO_2 would not be converted to some other gas species by solar ultraviolet light and would only very slowly be transformed into carbonate rocks by water-abetted weathering of surface rocks. The latter process would have a time scale of about 10 Myr (Pollack et al. 1987).

If the spectroscopically derived carbonate abundance is approximately correct and representative of the top several km of crust, the entire surface of Mars would have to be plowed up and processed to a depth of 2 km in order to generate a 1-bar CO_2 atmosphere. Apart from the daunting obstacles in practical engineering that this represents, it would also constitute the irresponsible destruction of a unique scientific resource and database—the Martian surface. If carbonate deposits are concentrated preferentially (by, say, an order of magnitude) in basins, then the basins would have to be excavated to a considerable depth—better, at least, than bulldozing the entire planet.

We find that both environmental ethics and probable cost argue against large augmented greenhouse effects from indigenous CO_2.

C. Chlorofluorocarbons (CFCs)

A number of CFCs have vibrational fundamentals that lie within the 8 to 13 μm window region of the atmosphere, and have been suggested as a significant potential agent for global warming of Mars (Lovelock and Allaby 1985; McKay et al. 1991). At current Martian temperatures, though, these absorption bands would cause only a small amount of greenhouse warming, because they lie on the short-wavelength tail of the Planck blackbody function. But they would give rise to a more substantial greenhouse effect as the surface temperature approaches 273 K due to other greenhouse gases, and the Wien peak of the blackbody function shifts to shorter wavelengths.

In principle, CFCs could be much more potent in significantly elevating the present Martian surface temperatures if they had pure rotational spectra that began near 20 μm and extended to longer wavelengths, as does NH_3. To do so, they must have permanent dipole moments and large moments of inertia. We suspect that HCFCs (e.g., $CHCl_2F$) might meet both these requirements. Unfortunately, we are not aware of any relevant laboratory measurements at the long wavelengths of interest. We defer a discussion of the source of CFCs until the next subsection.

For all greenhouse gases, the temperature increase would occur essentially on the time scale of release of the corresponding gases: the thermal time constant of the present Martian atmosphere is around 2 days; for $\sim$1 bar CO_2, it would be around 300 days, still very short compared to the other time scales of the project. As temperatures approach the freezing point, water vapor partial pressures in the atmosphere would increase, water vapor would augment the greenhouse effect due to the other gases, and the final stages of reaching clement temperatures might become comparatively easy.

Direct fusion heating of the Martian surface from power plants distributed over its surface is briefly described in the section on Titan below.

D. Possible Completion Scenarios

So far we have focused exclusively on schemes for increasing the mean global surface temperature of Mars to above the freezing point of water, i.e, into a habitable temperature regime. We now consider scenarios by which Mars is made habitable in a more complete sense for microbes, plants, other animals, and humans, with Table I as our guide for what needs to be done to the atmospheric composition to achieve this ultimate goal. Here we draw heavily upon the discussion of McKay et al. (1991).

The first human settlements on Mars will result from much more modest planetary engineering. We imagine these settlements to be small, enclosed microenvironments, in which the conditions of Table I are met. From them, in time, flows the means for generating more and larger microenvironments, culminating in a full-scale remaking of the entire planet's climate into one

suitable first for microbes and plants, and later for widespread habitation by animals, including humans. How might this be brought about?

McKay et al. (1991) have suggested that the first milestone, that of habitation by macroscopic plants, could be achieved by priming the greenhouse pump and allowing for natural amplification mechanisms to greatly increase the resulting warming. In particular, they recommend introducing enough greenhouse gases to warm the surface by about 20 K above present ambient. Such warming, they suggest, will lead to a large outgassing of CO_2 from the regolith, where it is adsorbed onto grain surfaces, and from the small perennial south polar cap. In this way the CO_2 pressure may be increased to a value as high as 100 mbar. Additional release of CO_2 might either occur from regolith or polar CO_2 reservoirs, or be accomplished by releasing CO_2 from carbonate rocks, until a CO_2 pressure of a bar to a few bars was achieved, at which point surface temperatures would permit survival and growth.

However, we believe it is unrealistically optimistic to think that any significant fraction of one to a few bars of CO_2 could be released from the regolith or polar regions, based on what is known about the nature of the Martian surface (see above). Furthermore, even 100 mbar of CO_2 would not significantly enhance the surface temperature (Pollack et al. 1987). Therefore, this scenario must draw heavily upon the release of CO_2 from carbonate rocks. In this regard, fusion power may play an essential role because of the very large amount of energy released per gram of water, as illustrated in our earlier discussion in this section. However, as discussed above, it will still be a highly nontrivial task to generate the requisite amount of CO_2 from carbonate rocks. The introduction of much more modest amounts of NH_3 or hydrogenated CFCs appears to offer a much more viable approach to raising the surface temperature of Mars into the habitable regime. According to Table I, several chemical modifications of the Martian atmosphere would be needed to make the planet habitable for plants, apart from the greenhouse warming discussed above. Here too, fusion power may prove to be very useful by helping to generate modest but enhanced partial pressures of O_2 (e.g., by conversion of small quantities of CO_2) and N_2 (e.g., by conversion of nitrates).

Once this milestone is achieved, the second stage—making Mars habitable for animals and for humans—can be addressed (cf. Table I). McKay et al. (1991) suggest that plant photosynthesis could convert enough atmospheric CO_2 to O_2 to meet the oxygen requirements. However, this suggestion rests on a misunderstanding, as the O_2 released by green plants derives exclusively from H_2O and not from CO_2. Such a photosynthetic source of O_2 would require at least 260 mbar of H_2 (which need not all be in the atmosphere at the same time) and 130 mbar of CO_2. Alternatively, fusion power could extract the needed oxygen from the highly oxidized surface, although several meters of regolith would need to be converted on global average. In addition, an enormous amount of N_2 would need to be introduced into the atmosphere to buffer the O_2 and prevent ready combustion from occurring. It is not clear that Mars has the requisite amount of N in easily accessible reservoirs. The

importation of the needed amount of N_2 from elsewhere in the solar system would be a formidable task, as illustrated by our discussion of obtaining a much more modest amount of N for generating NH_3.

Finally, the relative abundances of greenhouse gases would need to be carefully adjusted. Partial pressures of CO_2 greater than $\sim$10 mbar are toxic for humans. Thus, the CO_2 pressure might need to be greatly reduced from its earlier value, and CO_2 could not a prominent role in the late stages of greenhouse warming for human habitability. Likewise (NIOSH/OSHA 1978; Sittig 1985), 0.05 mbar of NH_3 is the short-term inhalation limit for humans at 1 bar (although some microbes can survive much higher NH_3 mixing ratios), which is comparable to the NH_3 partial pressure needed for it to become the dominant greenhouse gas. Note that the required NH_3 partial pressure decreases from $\sim$0.1 mbar to a much lower value as the total pressure increases from 7 mbar to $\sim$1 bar, so it seems possible to engineer a significant NH_3 greenhouse effect without reaching toxic levels.

Alternatively, CFCs or related compounds which have very low toxicity and ultraviolet lability, might serve as the chief greenhouse gases in completion scenarios, but subject to the far-infrared absorption considerations cited above. Manufacture on Earth of enough CFCs to warm Mars can be readily foreseen, because in only a few decades with present technology we have managed to synthesize enough to contribute to global warming on our planet. Transportation to Mars would be expensive: using *Saturn V* or *Energia* class boosters would require at least a launch a day for a century. The manufacture of sufficient quantities of CFCs and allied substances on Mars from evaporites or igneous fluorite or fluorospar might be feasible (J. Lewis, personal communication, 1992; Fogg 1992). In either case, as the CFCs would be photolytically destroyed in about a century, judging from the situation on Earth, this delivery or manufacturing rate would have to be continued forever, although recycling procedures would become the method of choice. However, CFCs as a principal greenhouse gas on Mars have a different undesirable property: they inhibit the formation of a substantial ozonosphere. The temperatures may be made clement, but the solar ultraviolet flux would still pose a very serious hazard. Perhaps this in turn could be fixed with a high-altitude layer of ultraviolet-absorbing particles, but the difficulty of planetary engineering on Mars is thereby greatly compounded. Perhaps some nontoxic, long-infrared-absorbing, ozone-noninteractive, and preferably inexpensive gas phase molecule that would do much better than CO_2, NH_3 or the CFCs remains to be synthesized. A more promising and less hypothetical alternative is to employ bromine- and chlorine-free halocarbons or related compounds (e.g., SF_6); F is much less effective in destroying O_3 than are Cl and Br (see, e.g., Fogg 1992; Chamberlain and Hunten 1987). Thus, NH_3 along with F compounds appear to offer the most promising means of warming Mars for habitation by humans.

In summary, it appears possible, although difficult to bring Mars globally to clement mean surface temperatures, more difficult to create conditions

suitable for plant habitation, and much more difficult to create conditions necessary for unprotected humans to live there. Even with rather optimistic assumptions on the progress of technology over the next century, and with very optimistic assumptions about the cost we will be willing to pay, it is not entirely clear that extensive planetary engineering of Mars is within reach.

III. VENUS

Because of its proximity to the Sun and its massive atmosphere, a very effective greenhouse operates on Venus that raises its surface temperature from what it would be without an atmosphere (about 240 K with the present albedo) to about 730 K. CO_2 is the major constituent of the Venus atmosphere and its opacity is the chief source of the greenhouse effect (Sagan 1960; Pollack et al. 1980). As the CO_2 pressure at the surface is $\sim$90 bar, equivalent to $\sim 10^5$ g cm^{-2}, engineering Venus by eliminating almost all its atmospheric CO_2 would appear to be a formidable undertaking.[a] We say "almost," because some residual greenhouse effect might be needed to bring the temperature from about 240 K to about 300 K. Alternatively, such fine tuning might be provided by adjusting the surface albedo or the albedo of high-altitude aerosols.

A. Impact Erosion

As a calibration of this formidability, we consider bombarding the planet with large, asteroid-sized bodies (tens to hundreds of km in size)—recognizing that such bodies may have been effective in eroding the atmospheres of the terrestrial planets, especially during the first billion years or so of their history when impact fluxes were much higher (Walker 1986; Melosh and Vickery 1989). High-velocity collision with the solid surface of a planet releases the kinetic energy of the impactor on a very short time scale. Several tens of percent of this energy is released into the atmosphere near the impact site as hot rock vapor, generating a shock wave that propagates outwards through the atmosphere. When the energy in the shock and vapor plume exceeds the gravitational binding energy of the portion of the atmosphere it traverses, this portion of the atmosphere can escape to space. For all but perhaps the largest impactors, the spherical geometry of the planet limits the affected portion of the atmosphere to the local tangent plane, corresponding roughly to 3×10^{-4} of the total volume of the atmosphere. According to calculations by Vickery and Melosh (1990), silicate bodies having masses $\gtrsim 5 \times 10^{18}$ kg (corresponding to objects larger than 150 km in diameter), and velocities in excess of $\sim$20 km s^{-1} are able to blow away all of the Earth's atmosphere above the tangent plane. The atmospheric mass dissipated falls rapidly as the

[a] If we could wholly remove other greenhouse gases such as H_2O and SO_2, we could lower the surface temperature by hundreds of °C, but still not nearly enough to fall below the normal boiling point of water (cf. Pollack et al. 1980). However, if the CO_2 abundance were massively reduced, pressure-broadening of H_2O and SO_2 would be greatly diminished, and their infrared opacity would become comparatively small.

mass of the impacting object diminishes, with no loss occurring for objects smaller than ~2 km in diameter, regardless of impact velocity. Also, no loss occurs when the impacting velocity falls below 20 km s^{-1}, regardless of the impacting body's mass. Similar thresholds apply to cometary impactors, although limited blowoff can occur at velocities somewhat below 20 km s^{-1}.

Because the escape velocity for Venus is similar to that for the Earth (10.4 vs 11.2 km s^{-1}), very similar limits on the minimum velocity needed to cause blowoff apply there. However, the mass requirements are increased by about a factor of 100 because of the greater atmospheric pressure at Venus' surface. Thus, impactors larger than ~10 km are required to produce any blowoff and ones larger than ~700 km are needed to blow off all the atmosphere above the tangent plane. To remove most of Venus' atmosphere through impact erosion would therefore require about 3000 impactors each larger than 700 km and traveling at velocities in excess of 20 km s^{-1}. There is nothing like this number of large bodies in the solar system, and if there were, it would surely be irresponsible to destroy them all (to say nothing of the present surface of Venus). Even with fusion power, excessive amounts of energy would be needed to accomplish the atmospheric removal. We conclude that impact stripping of the Venus atmosphere is an ineffective approach to planetary engineering.

B. Microbes

A much less energy-intensive and gentler approach is the introduction of genetically engineered micro-organisms into the clouds of Venus (Sagan 1961), where moderate temperatures prevail and where there is some water available. They would fix CO_2 into involatile organics or graphite, undoing much of the greenhouse effect, and lowering surface temperatures so that the sparse atmospheric water vapor would condense out on the surface in pools and shallow seas. Very special micro-organisms would be required because we now know the clouds of Venus to be a 75% solution of concentrated sulfuric acid which, among other disabilities, is a powerful desiccant. Nevertheless, there are micro-organisms known that live in concentrated solutions of H_2SO_4 and die in pure water (see also Seckbach and Libby 1969). We assume an H_2SO_4-tolerant N_2-fixing green plant photosynthetic micro-organism that could survive and replicate in the clouds of Venus. Biological solutions to planetary engineering have the advantage that microorganisms can rapidly replicate in the right environment; only a comparatively small number of such microbes then need be introduced.

This approach has been criticized for ignoring the fact that for every mole of CO_2 consumed in photosynthesis, a mole of water vapor is also required. As there is 3×10^{-5} as much H_2O in the Venus atmosphere as CO_2, it has been suggested that the process would grind to a halt far short of the desired transformation. However, the original proposal addressed this

issue: photosynthesis goes according to the standard heuristic equation

$$6CO_2 + 12H_2O \longrightarrow C_6H_{12}O_6 + 6O_2 + 6H_2O \quad (1)$$

in which the O_2 derives exclusively from the water. The carbohydrates thus synthesized will be circulated to the deeper hotter layers of the Venus atmosphere and pyrolyzed:

$$C_6H_{12}O_6 \longrightarrow C_6 + 6H_2O. \quad (2)$$

The carbohydrate is reduced to elemental carbon, or something approaching it, and the water is released, circulated back to the upper atmosphere and available for the next round of chemical reactions. The net chemistry is then approximately

$$6CO_2 \longrightarrow C_6 + 6O_2. \quad (3)$$

There are two critical questions: (1) whether all (within 3×10^{-5}) of the carbohydrate hydrogen is released at depth and recirculated, and (2) whether the kinetic barrier to graphite oxidation will be maintained at the temperatures of the middle and lower atmosphere of Venus. These are matters accessible to laboratory experimentation and thermodynamic equilibrium-kinetic calculations. Preliminary evidence on (2) is unfavorable: everyday experience with self-cleaning ovens suggests that re-oxidation of carbohydrate pyrolysis products to CO_2 at 700 K and 0.2 bar O_2 occurs on a time scale $\sim$1 hr.

Even if a microbiological approach to engineering Venus were feasible, however, the surface would be buried by hundreds of meters of fine graphite or organic particles sedimenting out of the atmosphere; and $\sim$65 bars of O_2 would be generated—the first at least inconvenient for, and the second lethal for human settlers on an engineered Venus. Such microbiological schemes at best carry planetary engineering on Venus only part-way towards human habitability.

C. Anti-Greenhouse Effect

Greenhouses can be turned down or off by substantially reducing the amount of sunlight reaching the surface, as in nuclear winter (Turco et al. 1983; see also, Hoyle and Wickramasinghe 1978). In the case of Venus, this might be accomplished, despite the resulting lowered albedo, by introducing a planet-wide layer of highly absorbing particles into its upper atmosphere. Naturally, this layer will slowly sediment towards the lower atmosphere and ground, and so would need periodic replenishment.

Elemental C is a very good absorber of sunlight. Taking a typical value of 0.3 for the imaginary part of its refractive index (Turco et al. 1983), we find that particles of radius 0.1 μm or more absorb almost all of the visible light passing through them. Particles of such sizes have an interaction cross section comparable to their geometric cross section at mid-visible wavelengths. Small particles are desirable for this scenario because (1) the total mass required for

a given amount of solar attenuation varies as the mean particle size (as long as the interaction cross section is comparable to the geometric value), and (2) their fall velocity varies inversely (as the first or second power) as their size. If a particle radius of 0.1 μm is assumed and an elemental C particle absorption optical depth of 5 is needed to diminish adequately the amount of sunlight reaching the Venus surface, then $\sim 6 \times 10^{14}$ g of carbon are required. This is equivalent to a single l-km-sized asteroid made of carbonaceous chondritic material. Particles 0.1 μm in radius would take several hundred years to fall to the surface at their sedimentation velocity (see, e.g., Ryan 1964), but might actually have lifetimes of only a few years when allowance is made for vertical transport by winds at a variety of length scales.

Carbon-rich (especially C- and D-type) asteroids that cross the Venus orbit might serve as the source of the carbon needed to create an anti-greenhouse effect on Venus. Orbital velocity changes needed to transport the asteroidal material to Venus are ~ 10 km s^{-1}, corresponding to $\sim 5 \times 10^{11}$ erg g^{-1} elemental C transported. Note that this specific energy is much larger than that needed to produce C from organic compounds found in asteroids. If once again fusion power is used, 1 g of H_2O would be required to transport about 2×10^3 g of C or a total of 2×10^{12} g of H_2O would be needed to create the opaque particle screen. We neglect the energy cost of pulverization and distribution of fine particles. This is clearly a low-mass, low-energy option for transforming Venus.

There are, however, several disadvantages to the anti-greenhouse option. It is still expensive in asteroids—requiring the periodic pulverization of a 1-km-sized organic-rich, water-rich asteroid. The thermal response time of the Venus atmosphere is a few decades, so if the high altitude layer of sedimenting opaque particles is not replenished, the carbon dioxide greenhouse effect reasserts itself on such a time scale and the surface temperature rises again to ~ 730 K. In addition, by its very nature, this option plunges the surface of Venus into deep gloom, with ambient light levels in daytime perhaps only as bright as on a moonlit night on Earth, and the oppressive 90 bar atmosphere remains untouched. This might conceivably be acceptable for an exploratory mission with human crews, but seems rather stark for a self-sustaining community of people on Venus. Further, as for all schemes, settlers would have to adjust to the very unearthlike 243 day sidereal period and 125-day synodic period (sunrise to sunrise).

D. Sunshades

Another approach is to construct a system of sunshades, perhaps of adjustable opacity, in space around Venus (Dyson 1989; Birch 1991). Alternatively, we might place the sunshade at the first Lagrangian point, L_1. The most efficient such sunshield, with sunlight assumed incident in parallel rays, would have a shadow the size of the object casting it. Were the shield as thin as 1 μm, its mass would be $\sim 10^{14}$ g. But for Venus to be in the umbra rather than the penumbra of the shadow cast, the shield would have to be roughly an order

of magnitude larger and, therefore, have a mass $\sim 10^{16}$ g. We thus run into apparently daunting economic problems, as discussed below for the Earth. With sunshades the Venus surface temperature could be brought to very low values; Dyson (1989) and Birch (1991) even discuss the possibility of freezing out the atmospheric carbon dioxide into CO_2 oceans or glaciers.

E. Chemical Transformations

Birch (1991) further suggests that the condensed CO_2, when in the liquid phase, would flow to the lowlands, forming CO_2 oceans. He then advocates covering the CO_2 seas to prevent their re-evaporation, bringing a several hundred km ice-rich moon from the outer solar system to form water oceans on top of the covered CO_2 seas, and only partially blocking the sunlight to establish an Earth-like temperature at the surface. Finally, a soletta (a sunshade that opens and closes) could create a 24 hr day/night cycle, while the establishment of plants on land could generate O_2 from a residual amount of CO_2 in the atmosphere (perhaps leaked from the CO_2 oceans).

From a conceptual point of view, the Dyson/Birch proposal comes closest to making Venus into a habitat for plants, animals and people. One could imagine adjusting the amount of solar blockage during the early phases of the project so that CO_2, but not N_2 condenses out of the atmosphere, with a view of meeting the N_2 requirements of Table I. However, this proposal involves a massive engineering effort (e.g., to build the sunshade and to cover the CO_2 seas), far beyond anticipated human resources during the next century. Furthermore, preventing the leakage of CO_2 back into the atmosphere once an equable climate is established would be a formidable task: at room temperature, essentially all the CO_2 sea would become vapor. This sea would cover most of the Venus surface area. Could leaks be prevented from occurring?

A lower-energy method of removing the atmospheric CO_2 would be to convert it to carbonate rocks, because the net reaction is exothermic. However, this approach (whether microbiological or industrial) is complicated by considerations of the mass of reacting (and pulverized) materials needed. The converted CO_2 would form a layer 400 m deep over the entire surface area of the planet, and an equivalent depth of CaO, MgO, or other suitable materials for forming carbonates would have to be processed. Again, many surface features would have to be not just covered, but obliterated.

Nevertheless, let us examine the prospects of converting almost all the CO_2 to low vapor pressure materials. Two ways to do this have been suggested. Birch (1991) proposes importing massive amounts of Ca and Mg as metals from Mercury. These metals would first reduce CO_2 to C and their metal oxides. The oxides would then react with CO_2 to produce carbonates. To gather enough metals from Mercury, however, would require processing $\sim 0.5\%$ of the planet's mass. Birch imagines that this would be done by self-replicating robots.

Alternatively, J. Lewis (personal communication, 1992) has suggested

that hot liquid water circulating through the near-subsurface of Venus would chemically weather the crust (due to dissolved CO_2 in the ground water, forming carbonic acid), leading to the formation of alkali metal and alkaline earth carbonates (see also Gillett 1991). A similar process occurs at room temperature on Earth and is a major component of the long-term geochemical C cycle (cf. Pollack et al. 1987). One could imagine the following scenario by which Lewis' scheme is used to create habitats on Venus: After Birch (1991), a one to several hundred km ice-rich body (either a comet or a satellite of an outer planet) is brought to Venus. This is needed so that substantial quantities of water will condense at temperatures only somewhat below its critical point. Next, the amount of sunlight reaching Venus' surface is somewhat reduced—either by using a sunshade or by introducing a limited amount of dark absorbing particles into its high atmosphere. Hot liquid water now is present at the surface and greatly accelerates the weathering rate of atmospheric CO_2, leading eventually to the elimination of almost all the atmospheric CO_2. At this point, clement temperatures can be established and maintained by keeping the amount of sunlight reaching the surface somewhat below levels that would obtain if there were no attenuating screens. (Otherwise, a runaway greenhouse would occur, with the formation of a steam atmosphere [Pollack 1991].) Now, plants that grow at the comparatively low light levels can be introduced and the final steps taken to make Venus habitable for humans.

F. Summary

In summary, making Venus habitable for either plants or people is an extremely formidable task. There does not seem to be any elegant way to do it. The key is to eliminate almost all the CO_2 from the atmosphere. There simply are not enough big bodies in the solar system to strip away its atmosphere (and even if there were, we might not want to use them). Atmospheric CO_2 might be condensed at the surface or chemically transformed. The latter possibility appears to be more promising in that the CO_2 is permanently sequestered although hundreds of meters of carbonates would now cover the surface. In addition, there is a need to augment significantly the water content of Venus' atmosphere and surface through the importation of water from elsewhere in the solar system. Finally, a limited blockage of sunlight must be maintained to prevent a runaway greenhouse from occurring. None of these proposals seems feasible for the next century.

IV. TITAN

Titan, the largest moon of Saturn, is the only satellite in the solar system with a substantial atmosphere. Its surface pressure is $\sim$1.5 bar (the column mass density is ten times that of Earth). The atmosphere is made mostly of N_2, with significant amounts of CH_4 (5 to 10% at the surface) and H_2 (a few tenths of a percent), as well as trace amounts of gas-phase hydrocarbons and

nitriles. An optically thick organic smog layer (created by solar ultraviolet light and Saturn magnetospheric electrons) is present in its stratosphere. The surface may contain extensive reservoirs of organics, and water and ammonia ice deposits. Currently, the surface temperature is only about 95 K. A modest greenhouse warming of 10 K is produced by pressure-induced transitions of N_2, CH_4 and H_2 (McKay et al. 1989). A major infrared window region, that extends from ~17 to 35 μm wavelength prevents a more substantial warming from occurring.

In principle, Titan's greenhouse could be significantly augmented by introducing gases that absorb strongly in the window region. However, the low surface and tropospheric temperatures greatly limit the vapor pressures of essentially all cosmically abundant potential candidates (e.g., NH_3, CO_2 and H_2O), which makes this approach difficult to implement. Many less abundant organics do not absorb preferentially in this window region, because their vibrational fundamentals tend to lie at shorter wavelengths and their rotational transitions at longer wavelengths. However, there are molecules that absorb in this region, including some alkanes, alkenes, and amides, and it might be possible to manufacture substantial quantities of such molecules from indigenous or exogenous resources.

An alternative approach is to heat the surface directly. McKay et al. (1989) have estimated the greenhouse warming that may have occurred when the sunlight reaching the lower atmosphere and surface was augmented by accretional heating due to the planetesimals that helped to build this satellite. There exists a very strong feedback between total surface flux (sunlight and accretional heating) and surface temperature, because of the exponential dependence of the vapor pressures of ammonia and water on surface temperature and the potential of these molecules for closing the greenhouse windows. Surface temperatures comparable to terrestrial values can be achieved when the accretional heat flux reaches $\sim 10^5$ erg cm^{-2} s^{-1}, which is ~0.05 the solar constant at Earth.

Clement temperatures on Titan could also be achieved through heat released by nuclear fusion, with surface volatiles supplying the D fuel. (We assume that there are substantial deposits of ammonia and water ice in contact with the atmosphere so these greenhouse gases may be mobilized. There is no shortage of sources of H.) To generate a mean heat flux of $\sim 10^5$ erg cm^{-2} s^{-1} across Titan's surface (with widely dispersed power plants), or equivalently 8×10^{22} erg s^{-1}, 2×10^4 g of deuterium would need to be consumed per second. An equivalent depth across Titan of some 400 μm of H_2O (or NH_3 or CH_4 or C_2H_6) would be needed for fuel each year (assuming a terrestrial D/H ratio). Such a requirement might easily be met and could be sustained for an extended period of time without exhausting the supply of H (which is easy) or without doing extensive environmental damage (which is more difficult). We also recognize that just the act of warming Titan to room temperature might cause extensive modification of its surface, as well as its atmosphere. Indeed, the nature of Titan's surface after such a warming would be an important

element in determining the desirability of altering its climate. If Titan were not habitable or could not be made habitable after this change, there clearly would be little motivation to do it.

In certain ways, Titan might be the easiest extraterrestrial object to make habitable. Unlike Mars, there is enough N_2 in its atmosphere to meet the requirements of Table I. Very likely abundant water resides at or close to Titan's surface. There is also enough oxygen in the form of H_2O, CO and CO_2 to serve as source material for generating O_2. However, the greatly reduced solar flux at Titan's distance from the Sun (down a factor of 100 from the Earth) would limit the rate of photosynthesis (and perhaps prevent it from occurring for some plants). Also, there is the question of how much dry land there would be once its surface temperature was raised above the melting point of water ice. Finally, the abundance of some atmospheric gases might need to be limited to prevent toxic levels from being reached (e.g., NH_3).

A similar scheme might be considered for Mars, but Titan has several advantages in this respect: the atmospheric pressure is much greater, and NH_3 and other greenhouse molecules (besides H_2O) are already present in the condensed phase. There is thus a much stronger greenhouse positive feedback to surface warming on Titan than on Mars. We estimate about an order of magnitude more technological power dissipation is required to bring Mars to 0°C than to bring Titan, despite the much lower present temperatures on Titan. A similar scheme might apply to Neptune's moon Triton if it has near-surface NH_3 ice.

V. EARTH

Our own planet has a uniquely suitable climate for our kind of life—no coincidence because the environment and the biology have co-evolved. However, as mentioned in the introduction, human technology is swiftly and dangerously altering that environment. We here briefly consider ameliorating one such potential danger, i.e., countering global warming with some form of planetary engineering. We note immediately that the scale of global warming over the next century is predicted to be several degrees C, so any solution to the increasing greenhouse effect must have high precision. Beyond that, even if the global mean temperature increase can be slowed or stopped, we must be careful not to cause local and regional agricultural and economic disasters in the process. The cure must not be worse than the disease. The debate about incompletely understood and possibly highly nonlinear feedback effects in the global climate, aired in the greenhouse context (see, e.g., DOE Multi-Laboratory Climate Change Committee 1990), is a particular reason for caution. Such considerations already suggest that planetary engineering of the sorts discussed above may be inappropriate for an already inhabited planet.

A. Anti-Greenhouse Effect

In countering greenhouse warming, we might try to reduce slightly (by a few percent) the amount of sunlight reaching the lower atmosphere by creating a carefully titrated optically thin particle layer at high altitudes. A 1% change in the albedo of the Earth then buys us about a 1°C change in temperature directly, and about a 2°C change when we take account of the water vapor greenhouse feedback. The mass of fine particles required is some 10^{12} to 10^{13} g; they would not seriously diminish the light levels at the Earth's surface. If they were emplaced in orbit, the required mass would be the equivalent of some tens of thousands of shuttle launches. Even with a substantial improvement in the human species' heavy lift capability, this seems out of reach for some time. Moreover, the pollution generated by the required launcher traffic might cause significant damage to the ozone layer. Alternatively, the fine particles could be placed at stratospheric altitudes. As in the case of Venus, the layer would then have to be replenished about once a year.

One of the first suggestions on how to mitigate greenhouse warming was to carry sufficient elemental sulfur to the lower stratosphere, and to burn it to sulfur dioxide which is then converted into sulfuric acid droplets which scatter sunlight back to space (Budyko 1974,1977, and references therein; see also Broecker 1985; for a more general suggestion on cloud condensation nuclei see Latham [1990]). Alternatively, SO_2 could be carried up directly. A modern version of Budyko's calculation might go as follows: 1 to 10 Tg (1 Tg = 10^{12} g) of H_2SO_4 in small droplets would probably suffice to nullify a few degree greenhouse warming. This would require the transport of some 1 to 10 Tg of S or SO_2 to the lower stratosphere every year (assuming one year for the aerosols to fall out), or equivalently 10^6 to 10^7 metric tons per year. The military transport aircraft with the largest payloads in the world are the Russian AN 225, which can lift 250 metric tons, and the U. S. C-5B which can lift 125 metric tons (Cheney 1989). Allowing for the apparatus to oxidize the elemental sulfur at altitude (if SO_2 itself is not carried), and the desire to get to higher altitudes than just above the tropopause, let us assume aircraft with 100 ton payloads. This then requires between 30 and 300 flights per day into the indefinite future. Budyko points out that the resulting sulfur precipitation rate would be only a small fraction of that which occurs naturally (e.g., from volcanoes) without such intervention. However, the impact of aircraft effluents (via NO_x) on ozone, and on the greenhouse effect itself need to be examined.

An alternative is to pulverize and chemically process in orbital factories, one small Earth-crossing (Apollo object) or Earth-approaching (Amor object) asteroid a year. But this is a voracious waste of the limited number of Apollo or Amor objects, and much more expensive than delivering an annual layer of transparent fine particles to stratospheric altitudes from the surface.

The fallout of fine particles from the atmosphere, necessitating a replacement of the light scattering layer every year or two, can be avoided if

a continuous solid shield were in place at much higher altitudes (see, e.g., Fogg 1987). One proposal to mitigate greenhouse warming (discussed in Broad 1988) is to deploy a vast array of orbiting satellites constructed of thin (mylar-like) films that would attenuate sunlight. However, to compensate for a doubling in the CO_2 abundance, such a satellite array would require an area equivalent to a few percent of the Earth's surface, if the films were good reflectors. Such films may be expensive. At contemporary prices, 10^7 km^2 of 6 μm thick aluminized mylar might cost nearly $\$10^{13}$; kapton would be perhaps 100 times more expensive (Friedman 1989, personal communication). The progressive darkening of such films on continued ultraviolet and charged particle irradiation (the transparency lifetime of mylar films might be about a month) would be an additional complication. Conceivably, much cheaper radiation-nondegradable films might be developed in the future, but they do not appear to be competitive with some of the alternative possibilities for ameliorating global warming.

Another proposal (Early 1989) calls for a "glass" shield 2000 km in diameter and perhaps 10 μm thick, made from lunar materials and suitably coated, positioned between the Earth and the Sun near the first Lagrangian point, about 1.5×10^6 km from Earth. The cost is estimated between $\$10^{12}$ and $\$10^{13}$ and the shield's long-term radiation and dynamical stability remains to be demonstrated. Dynamical instabilities may be addressed by equipping the shield with thrusters. In addition—unlike fine stratospheric particles, which take about a year to fall out and an orbital particle layer, which would be extremely difficult to remove—a sunscreen with an attached propulsion system can quickly be moved if unanticipated climatic side effects are found to be emerging on Earth. Better still for this purpose would be a sunscreen with louvers to fine-tune its effective opacity. We note also that reflecting shields need not be at high altitude or in space; massive reflectors floating in the oceans can also be envisioned, although at unknown environmental cost.

At best, however, the foregoing class of interventions in the Earth's climate might be considered as stopgap measures while the human species improves fossil fuel efficiency, terminates CFC manufacture, switches to alternative energy sources and takes other obvious groundbased measures. But why are any such heroic schemes needed, as we have already postulated abundant fusion power sometime in the next century? Why not simply assume that all power generation will derive from greenhouse-neutral fusion and consider the problem solved? The trouble is that even if cheap and safe fusion power were discovered in the laboratory tomorrow, its development to commercial scale and its deployment all over the world—including developing nations that have ready access to fossil fuels (China, for example, has the second largest coal reserves on Earth)—will take time. And in that time, the amount of CO_2 continues to build. Moreover, there are certain troublesome potential feedbacks—e.g., that a few degrees of global warming will release substantial quantities of now-sequestered bog methane, further augmenting the greenhouse effect. Even if much more serious efforts were made to deal

with global warming than now seem (politically) feasible worldwide, it is very difficult to imagine stopping the buildup of greenhouse gases altogether, at least in the next century: the world economy is far too dependent on fossil fuel energy sources. Many sources of greenhouse gases (some 30% of U. S. CO_2 emissions derive from automobiles) are not likely to be replaced by massive fusion power plants (although electric or hydrogen-fueled autos recharged from fusion power plants might solve this problem in the long term). So the consequences of continued greenhouse warming, even if somewhat delayed, might be considered sufficiently perilous to justify major, and costly, efforts at amelioration.

However, planetary engineering schemes utilizing extraterrestrial materials in Earth orbit or at L_1, or even terrestrial materials lofted to stratospheric altitudes, seem to us indeed to be cures that might be more dangerous than the disease—given both our present state of ignorance about climatic feedbacks and atmospheric chemistry, as well as our history of discovering unpleasant inadvertent side-effects of global scale technology. An additional deficit of many such schemes for planetary engineering of the Earth is that groundbased optical frequency astronomy would essentially cease. Although most (including us) would consider this an acceptable sacrifice to preserve the habitability of the Earth, it would be a tragic loss. Lagrangian point sunscreens would inhibit mainly solar astronomy, leaving the rest of the sky no more obscure than usual.

B. Trees

We advocate instead of particle shields or sunscreens, a well-tried biological solution, neither endangering astronomy nor posing unprecedented technological or climatic problems. We propose reforesting the world, especially in the tropics, in accordance with the ancient oriental wisdom, "He who causes trees to be planted lives long" (Polo 1300).

The added biomass would fix atmospheric CO_2, offsetting its continued buildup. It would buy time while massive conversion to nonfossil fuel energy sources was underway; it would also have positive ecological benefits (including relief for many endangered species); and it would be less likely to produce potentially disastrous regional shifts in climate. But is reforestation an allowable solution from the point of view of mass? The analysis of this issue was pioneered by Dyson (1977) and Dyson and Marland (1979). A restatement of the calculation goes as follows: doubling the CO_2 content of the atmosphere from its pre-industrial level would augment the CO_2 abundance by 0.55 g cm^{-2}, or equivalently by 0.15 g cm^{-2} of C. The dry biomass in a typical rain forest is about 6 g cm^{-2} (Lieth and Whittaker 1975), of which about half is C. Thus, about 5% of the Earth's surface would need to be reforested with rain forests (or a higher percentage with other types of forests) to fix the added CO_2 (while leaving stored C in storage). This is equivalent to an area of about 2.5×10^7 km^2, or about 15% of the landmass of the Earth. By comparison, tropical rain forests currently cover somewhat less area, while

tropical rain forests, rain-green forests, and summer-green forests collectively cover a somewhat larger area than the required 2.5×10^7 km^2. A massive replanting would be needed, necessarily including tropical latitudes where forests are now being destroyed on a massive scale for purposes of short-term commercial profit and "development." Rain forests serve to stabilize thin layers of topsoil, which, unfortunately, promptly (after a few growing seasons) erode once the forest is cut down. This trend towards irreversibility in the destruction of rain forests makes the practice especially short-sighted.

Mature forests, on yearly average, fix very little CO_2. The desired steady-state CO_2 fixation rate can be achieved with a smaller fraction of the Earth's surface forested by felling mature forests and re-planting them with fast-growing timber. This must be done with care to minimize the penalty to key ecosystems. As much of this wood as possible should be used for construction, furniture and other products of domestic economy. The remainder should be sequestered so that it is gradually converted to involatile kerogens and, eventually, elemental C. Obviously, burning these trees or the lumber milled from them defeats the purpose of growing them. When the world energy economy moves significantly away from fossil fuels, these strictures could be relaxed.

Net primary production in a tropical rain forest occurs at a rate of $\sim$0.2 g cm^{-2} yr^{-1} (Lieth and Whittaker 1975). Thus, $\sim$30 yr is needed between the initial replanting and the achievement of a steady-state biomass. This time scale is only somewhat less than current estimates of the doubling time of atmospheric CO_2. It pays to plant soon.

At their current rate of increase, gases other than CO_2 (most notably CH_4, CFCs, and N_2O) are generating an additional greenhouse warming of the atmosphere (Lacis et al. 1981; DOE Multi-Laboratory Climate Change Committee 1990), which collectively constitute about 50 to 100% of the warming due to added CO_2. Thus, an adequate strategy for avoiding significant global warming over the next century must involve controls on these gases as well. Controls are already being exercised on CFCs for other environmental reasons. As cattle and rice paddies are among the primary sources of CH_4, controlling the increase in methane could be integrally tied to such sensitive issues as the need for animal protein (and other resources provided by ungulates) and population control in the developing world, which in any case is a key component of any solution, because the developing nations together already constitute (after the U. S. and Russia) the third largest source of CO_2 emission on the planet (Brown 1989).

It has been proposed that the phytoplankton productivity in the southern ocean is limited by the availability of iron as a metabolite, that pre-industrial global CO_2 abundances were much higher than would have been the case without this iron deficiency, and that the abundance of iron-rich atmospheric dust is anticorrelated with atmospheric CO_2 levels (Martin 1990). This has led to the suggestion (Martin et al. 1990; Davies 1990) that fertilizing the southern ocean with 10^5 to 10^6 tons of finely pulverized soluble iron may lead

to a phytoplankton bloom that would significantly mitigate CO_2 greenhouse warming. This is the oceanic equivalent of massive re-forestation and superficially is far easier—fiscally and politically. However, unlike re-forestation, this is an intervention where we have no large-scale experience. More recent evidence suggests that iron may be much more available to oceanic phytoplankton than had previously been thought (Wells et al. 1991), and that ocean dynamic considerations vitiate the idea (Peng and Broecker 1991). Even if the iron-deficiency hypothesis were confirmed (cf. Davies 1990; Joos et al. 1991), it seems wise to proceed with great caution. Propagating negative ecological consequences have been suggested; this is also true for alternative interventions such as massive, periodically harvesting towed seaweed farms or piping liquified CO_2 to the abyssal depths where, perhaps, it might solidify (Blakeslee 1990).

Likewise, a scheme for mitigating stratospheric ozone loss by massive (5×10^4 ton annually) injection of ethane or propane into the Antarctic stratosphere (Cicerone et al. 1991) is partially disavowed by the scientists who suggested it on the grounds that it could increase, not decrease, ozone depletion; one of them is quoted (Dye 1991) as saying that "the proposal was meant to illustrate that such a plan is not feasible." (See also Cicerone et al. 1992.)

Since this chapter was prepared, an excellent survey of possible planetary engineering schemes to mitigate global warming on Earth has appeared (Committee on Science, Engineering, and Public Policy 1992). Favored options include re-forestation, delivery of dust to the stratosphere (using naval gunnery) and increasing oceanic cloudiness by supplying tropospheric cloud condensation nuclei (via SO_2). Possible side effects—such as enhanced ozone depletion and acid rain for the latter two schemes—are stressed. Clearly more work is needed, but comparatively inexpensive and environmentally prudent methods of mitigating greenhouse warming on Earth may be within reach in the next few decades.

VI. SUMMARY AND RECONSIDERATION OF SOLAR SYSTEM ENVIRONMENTAL ETHICS

We find that with optimistic estimates of late 21st century technology and physical principles at least moderately well-understood today, it may be possible to effect massive changes in the environments and climates of Mars, Venus, Titan and perhaps other worlds in the solar system. Some schemes—such as impact stripping the atmosphere of Venus or mining much of Mars or Venus down to depths of hundreds of meters or more—represent a wholly irresponsible waste of solar system resources and the loss of irreplaceable scientific knowledge. They are extreme examples of approaches that must be avoided. Fortunately, they are also prohibitively expensive. Other possible schemes (including those that utilize the self-replication of genetically engineered micro-organisms) are less destructive in their immediate impact, but

uncertain in their efficiency and long-term consequences. In particular, the introduction for a given purpose of microorganisms into a previously uninhabited planetary environment may, through evolutionary "adaptive radiation" into untenanted ecological niches, lead later to a set of entirely unexpected consequences.

Schemes that moderate atmospheric greenhouses by emplacing spherical shells of fine particles at high altitudes or sunscreens between the planet and the Sun to some extent avoid both of the preceding pitfalls: they are not wholly voracious in their appetite for asteroidal (or cometary) resources, and, because they naturally decay on a time scale of about a year or can be moved, do not seem to threaten dangerous long-term consequences. Because of the high specific energy of thermonuclear processes, some planetary engineering schemes that employ fusion power plants in a central role (e.g., by heating the surfaces of cold worlds) display both high efficiency and relative environmental responsibility.

It is clear that human technology is now able—even inadvertently, and certainly were there a concerted and purposeful effort—to alter entire planetary environments. Doubtless there will be schemes invented in coming years that are much less energy- and mass-intensive, much more precise, much freer of inadvertent side-effects, and much more stable than any proposed here.

We recognize that many people feel a powerful attraction to the idea of making other worlds in the solar system suitable for human habitation and then establishing observatories, exploratory bases, communities, or homesteads there. Because of its pioneering history, this may be a particularly natural and attractive idea in the United States. However, we believe that the motives for altering other worlds need much more serious consideration than they have received heretofore. Because some 240,000 more people are born than die every day on Earth, emigration to newly engineered worlds cannot be a feasible means of dealing with the world population crisis. If we are concerned that the human species might self- destruct, certainly it makes more sense to devote limited resources to preventing self-destruction than to preparing an escape hatch for a fortunate few. In any case, massive alteration of the environments of other worlds can be done competently and responsibly only when we have a much better understanding of those worlds than is available today. Advocates of planetary engineering should first be advocates of the long-term and exhaustive scientific exploration of other worlds.

A short-term imperative for planetary engineering exists only for one world in the solar system, our own. Careless or reckless applications of human technological genius have put the global environment at risk in several different ways. The Earth is not a disposable planet. It is just conceivable, as we have discussed, that some of the techniques that in the long term might be applied to engineering other worlds might also be utilized to ameliorate the damage being done to this one. Perhaps a safe way to test our protocols is to implement them in carefully circumscribed ways on other worlds. But considering the relative urgencies, a useful indication of when the human

species is ready to consider planetary engineering seriously is when we have put our own world right. We can consider it a test of the depth of our understanding and our commitment. The first step in engineering the solar system is to guarantee the habitability of the Earth.

Acknowledgments. This work was supported in part by two NASA grants. We thank C. Chyba, F. Dyson, M. Fogg, L. Friedman, J. Kasting, A. Lacis, C. McKay, G. Marland, J. Oberg, J. Pike, W. Rossow, R. Turco and, especially, J. Lewis for helpful comments.

REFERENCES

A. C. D. A. 1982. *Arms Control and Disarmament Agreements* (Washington, D. C.: U. S. Arms Control and Disarmament Agency).

Atreya, S. K., Donahue, T. M., and Kuhn, W. R. 1978. Evolution of a nitrogen atmosphere on Titan. *Science* 201:611–613.

Averner, M. M., and MacElroy, R. D., eds. 1976. *On the Habitability of Mars*, NASA SP-414.

Birch, P. 1991. Terraforming Venus quickly. *J. British Interplanet. Soc.* 44:157–167.

Blakeslee, S. 1990. *New York Times*, Nov. 20, C4.

Broad, W. J. 1988. *New York Times*, Aug. 16, 1986, C1, C9.

Broecker, W. S. 1985. *How to Build a Habitable Planet* (Palisades, N. Y.: Eldigio Press).

Brown, L. 1989. *The State of the World, 1989* (New York: W. W. Norton).

Budyko, M. I. 1974. *Meteorology and Hydrology* 2:91–97 (in Russian).

Budyko, M. I. 1977. *Climatic Changes* (Washington, D. C.: American Geophysical Union).

Burns, J. A., and Harwit, M. 1973. Towards a more habitable Mars—or—the coming Martian spring. *Icarus* 19:126–130.

Chamberlain, J. W., and Hunten, D. M. 1987. *Theory of Planetary Atmospheres* (San Diego: Academic Press), esp. pp. 136–140.

Chaun, R. L., and Woods, D. C. 1984. Comets, ice ages, and ecological catastrophes. *Geophys. Res. Lett.* 11:553–556.

Cheney, R. B. 1989. *Soviet Military Power: Prospects for Change* (Washington, D. C.: U. S. Dept. of Defense).

Cicerone, R. J., Elliott, S., and Turco, R. P. 1991. Reduced antarctic ozone depletions in a model with hydrocarbon injections. *Science* 254:1191–1194.

Cicerone, R. J., Elliot, S., and Turco, R. P. 1992. Global environmental engineering. *Nature* 356:472.

Committee on Science, Engineering, and Public Policy. Panel on Policy Implications of Greenhouse Warming. 1992. *Policy Implications of Greenhouse Warming: Mitigation, Adaptation, and the Science Base* (Washington: National Academy Press).

Cordey, J. G., Goldston, R. J., and Parker, R. R. 1992. Progress toward Tokamak fusion. *Physics Today* 45:22–30.

Davies, A. G. 1990. Taking a cool look at iron. *Nature* 345:114–115.

DOE Multi-Laboratory Climate Change Committee. 1990. *Energy and Climate Change* (Chelsea, Mich.: Lewis Publishers).

Dye, L. 1991. *Los Angeles Times*, Dec. 10.

Dyson, F. J. 1977. *Energy* 2:287–291.

Dyson, F. J. 1989. Terraforming Venus [letter/response to M. J. Fogg, The terraforming of Venus, 40:551–564]. *J. British Interplanet. Soc.* 42:593.

Dyson, F. J., and Marland, G. 1979. In *Carbon Dioxide Effects Research and Assessment Program: Workshop on the Global Effects of Carbon Dioxide from Fossil Fuels*, eds. W. P. Elliott and L. Machta, March 7–11, Miami Beach, Fl. U. S. Dept. of Energy, Conf. Rept. 770385, pp. 111–118.

Early, J. T. 1989. Space-based solar shield to offset greenhouse effect. *J. British Interplanet. Soc.* 42:567–569.

Fogg, M. J. 1987. Space-based solar shield to offset greenhouse effect. *J. British Interplanet. Soc.* 40:551–564.

Fogg, M. J., ed. 1989. Terraforming. *J. British Interplanet. Soc.* Special Issue 42:553.

Fogg, M. J. 1992. A synergic approach to terraforming Mars. *J. British Interplanet. Soc.* 45:315–329.

Gillett, S. L. 1991. Establishment and stabilization of Earthlike conditions on Venus. *J. British Interplanet. Soc.* 44:151–156.

Hansen, J., Fung, I., Lacis, A., Rind, D., Lebedeff, S., Ruedy, R., Russell, G., and Stone, P. 1988. Global climate changes as forecast by Goddard Institute for Space Studies Three-Dimensional Model. *J. Geophys. Res.* 93:9341–9364.

Harwell, M. A., and Hutchinson, T. C. 1985. *Environmental Consequences of Nuclear War: Volume II. Ecological and Agricultural Effects* (New York: J. Wiley).

Hoyle, F., and Wickramasinghe, N. C. 1978. Comets, ice ages, and ecological catastrophes. *Astrophys. Space Sci.* 53:523.

Joos, F., Sarmiento, J. L., and Siegenthaler, U. 1991. Estimates of the effects of South Ocean iron fertilization on atmospheric CO_2 concentrations. *Nature* 349:772–775.

Lacis, A., Hansen, J., Lee, P., Mitchell, T., and Lebedeff, S. 1981. Greenhouse effect of trace gases, 1970–1980. *Geophys. Res. Lett.* 8:1035–1038.

Latham, J. 1990. Control of global warming. *Nature* 347:339–340.

Lieth, H., and Whittaker, R. H. 1975. *Primary Productivity in the Biosphere* (New York: Springer-Verlag).

Lovelock, J., and Allaby, M. 1985. *The Greening of Mars* (New York: St. Martins Press).

Martin, J. H. 1990. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleocean.* 5(1):1–13.

Martin, J. H., Gordon, R. M., and Fitzwater, S. E. 1990. Iron in antarctic wars. *Nature* 345:156–158.

McKay, C. P., Pollack, J. B., and Courtin, R. 1989. The thermal structure of Titan's atmosphere. *Icarus* 80:23–53.

McKay, C. P., Toon, O. B., and Kasting, J. F. 1991. Making Mars habitable. *Nature* 352:489–496.

Melosh, H. J., and Vickery, A. M. 1989. Impact erosion of the primordial Martian atmosphere. *Nature* 338:487–489.

Morrison, D. 1992. *The Spaceguard Survey: Report of the NASA International Near-Earth-Object Detection Workshop*, ed. D. Morrison (Pasadena: Jet Propulsion Laboratory).

NASA. 1988. Present State of Knowledge of the Upper Atmosphere: An Assessment Report (Washington, D. C.: U. S. Government Printing Office).

NIOSH/OSHA. 1978. *Pocket Guide to Chemical Hazards* (Washington, D. C.: U. S. Depts. of HEW and Labor).

Oberg, J. 1981. *New Earths* (Harrisburg, Penn.: Stackpole Books).
Peng, T.-H., and Broecker, W. S. 1991. Dynamical limitations on the antarctic iron fertilization strategy. *Nature* 349:227–229.
Pittock, A. B., Ackerman, T. P., Crutzen, P. J., MacCracken, M. C., Shapiro, C. S., and Turco, R. P. 1986. *Environmental Consequences of Nuclear War: Volume I. Physical and Atmospheric Effects* (New York: J. Wiley).
Pollack, J. B. 1979. Climate change on the terrestrial planets. *Icarus* 37:479–553.
Pollack, J. B. 1991. Kuiper Prize Lecture: Present and past climates of the terrestrial planets. *Icarus* 91:173–198.
Pollack, J. B., Toon, O. B., and Boese, R. W. 1980. Greenhouse models of Venus' high surface temperature, as constrained by Pioneer Venus measurement. *J. Geophys. Res.* 85:8223–8231.
Pollack, J. B., Kasting, J. F., Richardson, S. M., and Poliakoff, K. 1987. The case for a wet, warm climate on early Mars. *Icarus* 71:203–224.
Pollack, J. B., Roush, T., Whitteborn, F., Bregman, J., Wooden, D., Stoker, C., Toon, O. B., Rank, D., Dalton, B., and Freedman, R. 1990. Thermal emission spectra of Mars (5.4–10.5 μm): Evidence for sulfates, carbonates, and hydrates. *J. Geophys. Res.* 95:14595–14628.
Polo, M. C. 1300. *Il Milione.* English trans. *The Travels of Marco Polo*, ed. R. E. Latham (Penguin, 1958), p. 156.
Ryan, J. A. 1964. Notes on the Martian yellow clouds. *J. Geophys. Res.* 69:3759–3770.
Sagan, C. 1960. The Radiation Balance of Venus. JPL Tech. Rept. 32-34 (Pasadena: Jet Propulsion Lab).
Sagan, C. 1961. The planet Venus. *Science* 133:849–858.
Sagan, C. 1973. Liquid carbon dioxide and Martian polar laminas. *J. Geophys. Res.* 78:4250–4251.
Sagan, C. 1986. Honda Prize Address. Transcript available from The Honda Foundation, Tokyo.
Sagan, C. 1992. Between enemies. *Bull. Atomic Scientists* 48:24–26.
Sagan, C. A., and Mullen, G. 1972. Earth and Mars: Evolution of atmospheres and surface temperatures. *Science* 177:52–56.
Sagan, C., and Turco, R. 1990. *A Path Where No Man Thought: Nuclear Winter and the End of the Arms Race* (New York: Random House).
Seckbach, J., and Libby, W. F. 1969. Vegetative life on Venus? Or investigations with algae which grow under pure CO_2 in hot acid media and at elevated pressures. In *Planetary Atmospheres*, eds. C. Sagan, T. C. Owen and H. J. Smith (Dordrecht: D. Reidel), pp. 62–83.
Sittig, M. 1985. *Handbook of Toxic and Hazardous Chemicals*, 2nd ed. (New York: Noyes).
Stapledon, O. 1948. Interplanetary man? *J. Brit. Interplanet. Soc.* 7:213–233.
Strong, A. E. 1989. Greater global warming revealed by satellite-derived sea-surface-temperature trends. *Nature* 338:642–645.
Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B., and Sagan, C. 1983. Nuclear winter: Global consequences of multiple nuclear explosions. *Science* 222:1283–1292.
Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B., and Sagan, C. 1990. Climate and smoke: An appraisal of nuclear winter. *Science* 247:166–176.
Vickery, A. M., and Melosh, H. J. 1990. Atmospheric erosion and impactor retention in large impacts, with application to mass extinction. In *Global Catastrophes in Earth History,* eds. V. L. Sharpton and P. D. Ward, Geological Soc. of America SP-247, pp. 289–300.
Walker, J. C. G. 1986. Impact erosion of planetary atmospheres. *Icarus* 68:87–98.

Wells, M. L., Mayer, L. M., Donard, O. F. X., de Souza Sierra, M. M., and Ackelson, S. G. 1991. The photolysis of colloidal iron in the oceans. *Nature* 353:248–250.
Wittenberg, L. J., Santarius, J. F., and Kulcinski, G. L. 1986. *Fusion Tech.* 10:167–178.
WMO. 1985. *Report 16* (Geneva: World Meteorological Organization).

Glossary

GLOSSARY*

absolute magnitude (M)	apparent magnitude that a star would have at a standard distance of 10 pc without absorption. The absolute magnitude (g) of a solar-system body such as an asteroid is defined as the brightness at zero phase angle when the object is 1 AU from the Sun and 1 AU from the observer.
A class	a rare asteroid taxonomic classification denoted by moderately high albedos and extremely reddish spectra shortward of 0.7 μm. A very strong near-infrared absorption feature centered around 1.05 μm is interpreted as being due to olivine.
albedo	ratio of the total flux reflected in all directions to the total incident flux.
Amor asteroids	asteroids having perihelion distance 1.017 AU$<q\leq$1.3 AU.
anorthite	Ca-rich end member of feldspar solid solution.
anorthosite	a magmatic rock rich in anorthite; common in the lunar highlands.
aphelion	Q; in the orbit of a solar system body, the most distant point from the Sun.
Apollo asteroids	asteroids having semimajor axis $a\geq$1.0 AU and perihelion distance $Q\leq$1.017 AU.
Aten asteroids	asteroids having semimajor axis $a<$1.0 AU and aphelion distance $Q>$0.983 AU.

* We have used some definitions from *Glossary of Astronomy and Astrophysics* by J. Hopkins (by permission of the University of Chicago Press, copyright 1980 by the University of Chicago), from *Astrophysical Quantities* by C. W. Allen (London: Athlone Press, 1973), and from *The Planetary System* by David Morrison and Tobias Owen (Reading, Mass.: Addison-Wesley Publishing Co., 1988). We also acknowledge definitions and helpful comments from various chapter authors.

augite	Ca-rich pyroxene variety.
basalt	a dark, fine-grained, mafic igneous rock composed primarily of plagioclase and pyroxene.
B class	a subclass of the C asteroids, distinguished by higher albedos than the average C type.
Beer's Law	a law describing the attenuation of light by an absorbing medium; intensity drops off exponentially with path light.
beneficiation	physical processing of a mineral mixture to enrich the concentration of useful ore.
Biosphere 2	a large closed chamber built in Oracle, Arizona, designed to support 8 humans by the use of biological recycling.
buffer gas	an inert non-toxic gas that when mixed with O_2 comprises breathable air. Examples include N_2, Ar, and He.
bytownite	Ca-rich feldspar variety, Na-bearing.
carbonaceous chondrite	a meteorite characterized by unmelted, undifferentiated textures and abundant organic matter and bound water.
C class	a very common asteroid type in the outer part of the main belt; they typically have flat spectra longward of 0.4 μm and are presumably similar in surface composition to some carbonaceous chondrites. The relative strength of a ultraviolet absorption feature may be correlated with the presence of water of hydration. B, F and G are subclasses of the C class.
column amount	total amount of a substance in the entire column of the atmosphere over a unit area.
coma	the usually spherical region of diffuse gas, typically 150,000 km in diameter, which surrounds the nucleus of a comet. Together, the coma and the nucleus form the comet's head.

comet — a comet consists of three parts: (1) a small icy-conglomerate nucleus; (2) an extended outflowing atmosphere or "coma" of volatile gases evolved from the nucleus plus entrained dust particles, and (3) tails of dust and ionized gas molecules. A comet's orbit is usually highly elliptical or even parabolic, and can cross the orbits of many planets.

CRAF mission — Comet Rendezvous Asteroid Flyby; a proposed NASA mission to orbit a comet nucleus with at least one asteroid flyby en route.

D class — an asteroid type that is rare in the main belt, but becomes increasingly dominant beyond the 2:1 Jovian resonance. Their spectra are neutral to slightly reddish shortward of 0.5 μm, and for some objects the spectrum tends to flatten longward of 0.95 μm. Coloring may be due to kerogen-like materials.

ΔV — velocity change.

Deimos — the outer satellite of Mars, $15\times12\times11$ km p=30$^{\rm h}$18$^{\rm m}$, e=0.003; inclination of orbit to planetary equator 1°.6. Visual geometric albedo 0.06. Deimos is 23,500 km from Mars. Its surface is similar to carbonaceous chondrite. Both Phobos and Deimos are locked in synchronous rotation with Mars. Discovered by A. Hall in 1877.

diffusion — the net movement of atoms within a material. The rate at which atoms move is controlled by the temperature.

drift scanning — a method for exposing a CCD to a moving image whereby the rate of charge transfer along the parallel columns of the CCD matches the rate at which the image moves across the field of view.

E class — a rare asteroid type with featureless 0.3 to 1.1 μm spectra (identical to M and P classes) but distinguished by high albedos. Surface composition may be similar to enstatite achondrites.

F class — a subclass of the C asteroids, distinguished by a weak to nonexistent ultraviolet absorption feature.

feldspar	Al-bearing framework silicate.
fracture toughness	a material parameter which measures a material's resistance material to failure when a flaw is present.
frost point temperature	the temperature at which frost first forms as the air is cooled. The frost point temperature depends only on the amount of water in the atmosphere.
G class	a subclass of the C asteroids, distinguished by a strong ultraviolet absorption feature.
GEO	geosynchronous orbit (orbital period of exactly one day).
halogens	the elements fluorine, chlorine, bromine and iodine.
HF	hydrogen fluoride.
highland	heavily cratered upland with anorthositic composition.
ilmenite	Fe-Ti-oxide.
Joule-Thompson expansion	free expansion of a compressed gas into a near vacuum.
KREEP	lunar basaltic material rich in radioactive elements (K for potassium, REE for rare Earth elements, P for phosphorus).
KRFM	a Soviet instrument on the Phobos mission; KRFM measures radiation in the visible and infrared portions of the spectrum.
Kuiper belt	a hypothetical belt of 10^8 to 10^{10} unaccreted icy planetesimals in low eccentricity orbits beyond the orbit of Neptune, left over from the formation of the planetary system, and the likely source of most short-period comets. The first Kuiper belt object, 1992 QB_1 was recently discovered.
LEV	lunar excursion vehicle.
lime	Ca-oxide or -carbonate.

liquid phase sintering	a densification mechanism which involves the presence of a viscous liquid in addition to solid particles at the sintering temperature.
long period comet	a comet with an orbital period longer than 100 yr, but typically on the order of 1 Myr.
LLOX	liquid oxygen made on the Moon: lunar LOX.
mafic	composed of silicates of magnesium and iron.
main belt asteroids	asteroids that occupy the main asteroid belt between Mars and Jupiter, sometimes limited specifically to the most populous parts of the belt, from 2.2 to 3.3 AU from the Sun.
mare	*pl.*, maria; an area on the Moon or Mars that appears darker and smoother than its surroundings. Lunar maria are basins flooded by basaltic lavas.
M class	a fairly common asteroid type in the main belt, with featureless 0.3 to 1.1 μm spectra (identical to E and P classes) that are distinguishable by moderate albedos; presumed to have metallic (Ni-Fe) compositions, but with varying metal contents.
MLS	Minnesota Lunar Simulant; a simulant of the physical and gross chemical properties of lunar regolith made from terrestrial rock.
Mohr-Coulomb strength theory	a theory relating the shear strength of granular material to its function angle and cohesion.
near-Earth asteroids (NEAs)	a heliocentric orbiting population of small bodies having perihelion distances $\leq$1.3 AU.
noble gases	the gases He, Ar, Kr, Ne, Xe, Rn, which rarely undergo chemical reactions; also known as inert gases and rare gases.
nucleus (of a comet)	a small body composed of volatile ices (primarily water), silicate dust, and hydrocarbons, orbiting the Sun. When it approaches the Sun, ices on the nucleus surface sublimate and the evolving gases carry entrained dust particles with them, producing an extended atmosphere, or "coma."

olivine — the most abundant mineral in chondritic meteorites, $(Mg,Fe)_2AiO_4$.

Oort cloud — a spherical cloud of 10^{12} to 10^{13} comets surrounding the solar system with orbits extending to interstellar distances. Comets in the Oort cloud are perturbed into the planetary system by perturbations from random passing stars, molecular clouds, and the galactic tide.

partial pressure — the fractional pressure exerted by a mixture within a gas, directly proportional to the fraction of the mixture composed of that gas.

P class — a fairly common asteroid type in the outer main belt with a heliocentric distribution that peaks near the 3:2 Jovian resonance. Their spectra are featureless from 0.3 to 1.1 μm (identical to E and M classes) but the class is distinguishable by low albedos.

perihelion — the point in the orbit of an object orbiting the Sun where it is closest to the Sun's center of mass, q. Earth's perihelion occurs early in January.

Phobos — the potato-shaped inner satellite of Mars (about $27 \times 21 \times 19$ km), discovered by A. Hall in 1877. Mass about 1.1×10^{10} g.

PIA experiment — flew onboard the Giotto spacecraft.

Planck blackbody formula — a formula that determines the distribution of the intensity of radiation that prevails under conditions of thermal equilibrium at temperature T:

$$B_v = (2hv^3/c^2)[\exp(hv/kT) - 1]^{-1}$$

where h is Planck's constant and v is the frequency.

polar hood — a broad expanse of water-ice or CO_2-ice clouds occurring over the polar regions on Mars during the fall or winter seasons.

PUMA experiment — flew onboard the Vega spacecraft.

pyroxene — Fe/Mg/Ca-bearing chain silicate, $(Fe,Mg,Ca)SiO_3$.

radio isolation	the condition of being on the far side of shielding, as provided by the bulk of the Moon itself, sufficient to block the reception of radio signals from Earth, its magnetosphere, and its artificial satellites.
R class	a rare asteroid classification exemplified by 349 Dembowska and denoted by moderately high albedos and spectra with a strong absorption feature shortward of 0.7 μm and a fairly strong absorption feature centered near 1 μm.
S class	a very common asteroid class in the inner main belt with moderate albedos and reddish spectra shortward of 0.7 μm and moderate to nonexistent absorption features in the near-infrared; may be similar to stony-iron meteorites, but their meteoritical interpretation is uncertain.
selenology	lunar geology, lunar geophysics and other scientific disciplines concerning the Moon and its environs.
sintering	a high temperature heat treatment used to bond the surface of adjacent particles. The bonds are usually formed by the diffusion of viscous flow of material to points of contact between the particles with a concurrent reduction in the volume of the pore space.
soil	surface fines overlying the local bedrock.
Spacewatch	an ongoing CCD search for near-Earth asteroids and a radial velocity search for planets around other stars conducted at the 0.9 m newtonian telescope of the University of Arizona and Kitt Peak.
short period comet (SP)	a comet with an orbital period of less than 100 yr.
specific impulse (I_{sp})	the thrust produced by burning a rocket propellant at the rate of one pound per second, in seconds.
T class	a class of low albedo asteroids having a spectra with a moderate absorption feature shortward of 0.85 μm and generally flat in the near-infrared.

telepresence — techniques permitting a person to feel, to a limited extent, through vision, sound, touch and/or other senses, as if in a remote location, thereby controlling the operation of electromechanical equipment accurately via a telecommunications link to perform a physical function at that remote location.

Tisserand criterion — a pseudo-constant of the motion in the restricted three-body problem based on the Jacobi integral, used to identify returning comets even though their orbits may have been perturbed by a close Jupiter encounter: $T = a_J/a + 2\sqrt{[(a/a_J)(1-e^2)]}\cos i$ where a, e, and i are the comet's semimajor axis, eccentricity, and inclination (to the plane of Jupiter's orbit), respectively, and a_J is the semimajor axis of Jupiter's orbit.

trans-Earth injection (TEI) — departure from the vicinity of another planetary body to begin the return to Earth.

UBV system — a system of stellar magnitudes devised by Johnson and Morgan at Yerkes Observatory. The system measures an object's apparent magnitude through three color filters: the ultraviolet (U) at 3600 Å; the blue (B) at 4200 Å; and the "visual" (V) in the green-yellow spectral region at 5400 Å. It is defined so that for AO stars, B−V≡U−B≡0; the color indices are negative for hotter stars and positive for cooler stars. The UBV system has become the most widely used broadband photometric system.

UHF — ultra high-frequency radio waves.

V class — a rare asteroid classification exemplified by 4 Vesta. Spectra are very red shortward of 0.5 μm, moderately red from 0.5 to 0.7 μm, and show a strong near-infrared absorption feature centered around 0.95 μm. Surface composition may be similar to basaltic achondrites.

viscous flow — permanent deformation of a glassy material at high temperatures.

vulcanoid — a hypothetical object orbiting near the Sun interior to Mercury, or at the Langrangian points of Mercury.

Acknowledgments

ACKNOWLEDGMENTS

The editors ackowledge NASA Grant NAGW-2160 and The University of Arizona, especially Dr. Terry Triffet and the UA/NASA Space Engineering Research Center, for support in the preparation of this book. They wish to thank J. E. Frecker, who volunteered as one of the proofreaders of this book. The following authors wish to acknowledge specific funds involved in supporting the preparation of their chapters.

Baker, V. R.: NASA Planetary Geology and Geophysics Program Grant NAGW-285
Bell, J. F.: NASA Grant NAGW-1383
Fegley, B., Jr.: NASA Grant NAGW-2867
Friedlander, A. L.: NASA Technical and Administrative Support, Contract NASW-4543
Greenberg, R.: NASA Planetary Geology and Geophysics Grant NAGW-1029
Gulick, V. C.: NASA Graduate Student Researchers Program Grant NGT-50662
Haberle, R. M.: The NASA Planetary Atmosphere Program through RTOP 154-95-80-05 at Ames Research Center
Jakosky, B. M.: The NASA Planetary Atmospheres Program through Grant NAGW-552 to the University of Colorado
Kubiak, C. P.: NASA Grant NAG3-1336
McSween, H. Y.: NASA Grant NAG9-58
Pletka, B. J.: NASA Grant NASW-4486
Pollack, J. B.: NASA Grants NAGW-1870 *and* NGL 33-010-082
Sagan, C.: NASA Grants NAGW-1870 *and* NGL 33-010-082
Senior, C. L.: NASA Contract NAS9-18356
Singer, R. B.: NASA Planetary Geology and Geophysics Grant NAGW-1059 and a University of Arizona/NASA Space Engineering Grant
Staehle, R. L.: World Space Foundation, Space/Media, NASA and Jet Propulsion Laboratory

Index

INDEX